UNIVERSITY OF SOUTHAMPTON
FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Prediction of Course Completion based on Participants' Social Engagement on a Social-Constructivist MOOC Platform

by

Ayşe Saliha Sunar
BSc, MSc

A thesis for the degree of
Doctor of Philosophy
at the University of Southampton

September 2017

Supervisors: *Dr. Su White*
BSc(Econ), PGCE, PGDip Computer Science, PhD
and
Professor Hugh C Davis
BSc, MSc, PhD, MBCS, CITP, FHEA

Southampton Web and Internet Science Group
School of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ
United Kingdom

Dedicated to my family.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
School of Electronics and Computer Science

Doctor of Philosophy

PREDICTION OF COURSE COMPLETION BASED ON PARTICIPANTS' SOCIAL ENGAGEMENT ON A SOCIAL-CONSTRUCTIVIST MOOC PLATFORM

by Ayşe Saliha Sunar

MOOCs offer world-widely accessible online content typically including videos, readings, quizzes along with social communication tools on a platform that enables participants to learn at their own pace. In 2016, over 58 million people join MOOCs.

Far fewer people actually participate in MOOCs than originally sign up and then there is a steady attrition as courses progress. The observation of high attrition has prompted concerns among MOOC providers to mitigate their high attrition rates.

Recent studies have been able to correlate social engagement of learners to course completion. Researchers use participants' digital traces to make sense of their engagement in a course and identify their needs to predict future patterns and to make interventions based on these patterns.

The research reported here was conducted to further understand learners social engagement on a social-constructivist MOOC platform, the impact of engagement on course completion, and to predict learners' course completion.

The findings of this research show that a commonly known social feature, *follow*, which is integrated into the Futurelearn MOOC platform has potential value in allowing tracking and analysing the behaviours of participants. The patterns of learners social engagement were modelled and a completion prediction model was developed. This model was successful at predicting those who might complete the course at a high or low success rate.

The contributions of this research are that the *behaviour chains* could be the basis of a personalised recommender system, and the completion model based on social behaviour could contribute to wider prediction model based on a wider range of factors.

Declaration of Authorship

I, **Ayşe Saliha Sunar**, declare that the thesis entitled

Prediction of Course Completion based on Participants Social Engagement on a Social-Constructivist MOOC Platform

and the work presented in it are my own and have been generated by me as the result of my own original research. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published, as seen in the list of publications.

Signed: Ayşe Saliha Sunar

Date: 1 September 2017

Acknowledgements

I wish to express my heartfelt gratitude to my supervisors Dr. Su White and Professor Hugh Davis for their supervision and support throughout my research. I have greatly benefited from their excellent guidance, support and friendly discussions. Their patience, enthusiasm, encouragement and wisdom have been highly supportive to me in my work.

I would like to acknowledge my appreciation to Professor Toyohide Watanabe who was my supervisor during my master education in Nagoya University. I would not courage to start my PhD in the United Kingdom without his encouragement. I am also deeply grateful for his encouraging spiritual support throughout my PhD.

I would also like to thank Dr. Nor Aniza Abdullah for her guidance, support and great patience throughout my research. Her knowledge and attention inspired me to always work harder. Our discussions played an important role in the direction of my research.

I am deeply grateful to Manuel León Urrutia, Adriana Wilde, and Kate Dickens. They have always had time for my questions and have helped me endure the ups and downs of being a researcher.

I would like to thank Chris Lowis from the FutureLearn MOOC platform for providing additional data whenever I requested for my research.

During the journey of my PhD, I also had chance to collaborate with other fellow researchers from different institutes. I would like to thank Dr. Gülüstan Doğan and İsmail Duru from Yildiz Technical University, Turkey. I also would like to thank Dr. Ahmed Mohamed Fahmy Yousef from Fayoum University, Egypt, Dr. Rabeeh Ayaz Abbasi and Dr. Naif R. Aljohani from King Abdulaziz University, Saudi Arabia.

The support and kindness of the many friends and colleagues whom I have made throughout these past few years have been invaluable to me. Although there are too many names to mention, I would like to express my gratitude to all those, who

have been at least once with the University of Southampton, both past and present. I want to particularly thank Dr. Long Tran-Thanh, Elisabeth Coskun, Olja Rastic-Dulborough, Tania Aria Edries, Fadiyah Almutairi, Rehab Albeladi, and all other colleagues and staff, too numerous to mention here explicitly.

Without the support, patience and guidance of the following people, this study would not have been completed. It is to them I owe my deepest gratitude.

- I would like thank my dear friend Ghaithaa Manla for always being by my side to support me.
- Our *Quran Friends* circle: Nada Albunni, Susan Nazirizadeh, Stephanie Bispo, Miriam Animashaun, Uğur Mutlu, Raphael Sikorski, Duygu Cihan, Thomas Woollett, Dr. Aiman Alzetani, Sylvain Grosse, Sana Rashid, Dimitris Kostovasilis, Darbaz Muhamed, Shumaila Mahmood, Raid Hussein, and Anisa Ather. Thank you for mind-blowing discussions and amazing social events.
- I would like to thank everyone who participated in *Joint Summer School on Technology Enhanced Learning* in 2015 and 2016. I have made amazing friends and shared unforgettable memories together.
- I would like to thank my *social* friends on Twitter who have supported me from miles away.
- I always felt my cousins' support in my heart. Thank you Beyza, Ayşe Nur, and Tuba for your friendship and support.
- I would like to thank my friends Fatma Zehra, Esra, and Ayşe for always listening to me whenever I needed.
- Most of all, I am deeply grateful to Zeynep Aydın Asarkaya, my best friend for 15 years, who has always encouraged me throughout my education.

The financial support of the Republic of Turkey Ministry of National Education is also gratefully acknowledged.

I would also like to express my appreciation to the most caring person in my life, my mother Özden Kökcan Sunar, to my father Sezayi Sunar, to my lovely sister Bürde Süheyla Bayraktar, to my brother Esat Mustafa Sunar, and to my grandparents Öznur and Mustafa Kökcan as well as to my beloved husband, Halil Yetgin and his family, for their love, prayers, support and care for me.

List of Publications

Book Chapters

1. **A. S. Sunar**, N. A. Abdullah, S. White and H. C. Davis, "Personalisation in MOOCs: A critical literature review", *Communications in Computer and Information Science*, Springer International Publishing, vol. 558, pp. 152-168, February 2016.

Journal Papers

1. **A. S. Sunar**, S. White, N. A. Abdullah and H. C. Davis, "How learners' interactions sustain engagement: a MOOC case study", *Special Issue IEEE Transactions on Learning Technology*, November 2016. access on: http://ieeexplore.ieee.org/document/7762189

Conference Papers

1. I. Duru, **A. S. Sunar**, G. Dogan, and S. White, "Challenges of identifying second language English speakers in MOOCs", *5th European MOOCs Stakeholders Summit (EMOOCs 2017)*, Madrid, Spain, May 2017.
2. G. Dogan, **A. S. Sunar**, I. Duru, and S. White, "Who is the English as a second language speaker in this MOOC?", *2nd International Conference on Information and Network Technologies (ICINT 2017)*, Jakarta, Indonesia, May 2017.
3. **A. S. Sunar**, N. A. Abdullah, S. White and H. C. Davis, "Analysing and predicting recurrent interactions among learners during online discussions in a MOOC.", *11th International Conference on Knowledge Management (ICKM 2015)*, Osaka, Japan, November 2015.
4. **A. S. Sunar**, N. A. Abdullah, S. White and H. C. Davis, "Personalisation of MOOCs: the state of the art", *7th International Conference on Computer Supported Education (CSEDU2015)*, Lisbon, Portugal, May 2015.

Workshops and Posters

1. A. M. F. Yousef and **A. S. Sunar**, "Opportunities and challenges in Personalized MOOC Experience", *Workshop on Web Science Education*, ACM, Oxford, UK, June 2015.

2. **A. S. Sunar**, N. A. Abdullah, S. White and H. C. Davis,, "Analysis of social learning networks on Twitter for supporting MOOCs education", *ACM-W Europe womENcourage Celebration of Women in Computing*, Uppsala, Sweden, September 2015.

Table of Contents

Abstract v

Declaration of Authorship vii

Acknowledgements ix

Table of Contents xiii

1 Introduction 1
- 1.1 Introduction . . . 1
  - 1.1.1 Participation in Online Discussions and High Attrition Rates in MOOCs . . . 4
  - 1.1.2 Learning Analytics in MOOCs . . . 6
  - 1.1.3 Prediction Models and Educational Interventions in MOOCs . . . 7
- 1.2 Research Hypothesis . . . 8
- 1.3 Research Aims and Research Questions . . . 8
- 1.4 Methodological Approach . . . 9
- 1.5 Contribution of This Research . . . 9
- 1.6 Outline of the Thesis . . . 12

2 Design and Social Affordances of the FutureLearn MOOC Platform 15
- 2.1 Research Context . . . 15
- 2.2 Underlying Pedagogy of FutureLearn . . . 16
  - 2.2.1 Course Design in FutureLearn . . . 17
  - 2.2.2 Social Affordances in FutureLearn: post, scroll down, like, reply, follow, filter, notify . . . 18
  - 2.2.3 Social Actions and Roles of Learners in FutureLearn . . . 19
- 2.3 Provided Datasets by FutureLearn . . . 20
- 2.4 Summary . . . 23

3 Analysis and Classification of Learners' Behaviours in MOOCs 25
- 3.1 Introduction . . . 25
- 3.2 Critical Analysis on Classification of MOOC Learners' Behaviours . . . 26
- 3.3 Summary . . . 28

4 Social Participation in a FutureLearn MOOC 31
- 4.1 Introduction . . . 31

4.2 Analysis and Results . . . 32
4.2.1 General Statistics on Social Participation . . . 32
4.2.2 Involvement of Followers in Discussions . . . 37
4.2.3 Completion Success of Participants . . . 38
4.2.4 Mentors in the Data . . . 42
4.3 Summary . . . 44

5 Behaviour Chains of Learners and Correlations to Course Completion 47
5.1 Introduction . . . 47
5.2 Characterising Social Behaviours . . . 48
5.3 Observed Behaviour Chains of Participants . . . 51
5.3.1 Correlation between Course Completion and Chain Types . . . 54
5.3.2 Correlation between Course Completion and Frequency of Social Actions . . . 56
5.3.3 Correlation between Course Completion and Continuity to Contribution (Fullness of Chain) . . . 57
5.4 Summary . . . 62

6 Use of Prediction Models in MOOCs 65
6.1 Introduction . . . 65
6.2 Critical Analysis on Prediction Models . . . 66
6.3 Summary . . . 67

7 A Novel Approach for Predicting Learners' Future Participation 75
7.1 Introduction . . . 75
7.2 Feature Set Selection . . . 76
7.3 Machine Learning for Classification . . . 77
7.3.1 Random Forest Model . . . 77
7.3.2 Support Vector Machines . . . 78
7.4 Implementations of Prediction Models . . . 78
7.4.1 Imbalanced Data Problem . . . 78
7.4.2 Training Data: k-fold cross-validation . . . 78
7.4.3 The Workflow of Implementation . . . 79
7.4.4 Results . . . 79
7.4.4.1 Testing with Random Forest Model . . . 79
7.4.4.2 Performance of the raw variables in the classification with Random Forest Model . . . 83
7.4.4.3 Testing with Support Vector Machine . . . 84
7.4.4.4 Comparison of the Results by Models . . . 84
7.5 Discussion of the Results . . . 86
7.5.1 Testing on Different MOOCs . . . 86
7.5.2 Timely prediction for timely intervention . . . 87
7.5.3 Additional Feature Extraction from Participants' Behaviour . . . 88
7.6 Summary . . . 89

8 Conclusions and Future Research **91**
8.1 The Results of this Research . . . 92
8.1.1 Participants' Engagement with Social Affordances on FutureLearn 92
8.1.2 Identifying and Modelling Different Social Behaviours . . . 93
8.1.3 Prediction of Course Completion based on Learners' Social Behaviours . . . 94
8.2 Limitations of this Work . . . 95
8.3 Contribution of this Research . . . 97
8.4 Ideas for Future Work . . . 98
8.4.1 Improving the model on different courses . . . 98
8.4.2 Improving the data structure and the range of social affordances on the platform . . . 98
8.4.3 Personalised recommenders in MOOCs . . . 99
8.4.4 Gamification . . . 101
8.4.5 An Improved Model For Predicting Completion . . . 101
8.5 Final Remarks . . . 102

A Documents for Ethics Approval **i**

B Implementation of Machine Learning Algorithms with *R* **ix**

List of Figures **xi**

List of Tables **xiii**

References **xv**

Chapter 1

Introduction

1.1 Introduction

As the number of worldwide Internet users grows and use of the Web and its features evolve, how people communicate is also changing. The early version of the Web (known as Web 1.0) was only used to access information over static webpages. The role of users was not participatory; rather there was a large number of readers who accessed content created by a much smaller number of content creators (Nath et al., 2014).

As an educational tool, Web 1.0 was predominantly used to "publish" information to give students access to knowledge and information. The implementations of Web 1.0 in education followed the underlying idea of instructivist models (Gerstein, 2014). According to instructivist theory, knowledge is structured independently of the learner. According to this model of learning, learners passively accept knowledge as presented by instructors. Learners were still in the traditional role of *receiver*, and with the aid of technology, information was delivered to them via the Web (Reeves, 1993). Consequently, the level of engagement of learners with the content presented was quite low.

Subsequently, additional functionality such as wikis and web blogs has enabled the role of web users to evolve to a read-write-share role. This change has often been referred to as the *social web* or Web 2.0 (Musser and Oreilly, 2006). Users now expect to create, modify, and update online content and communicate among each other. The supportive nature of social web for individual production promotes some social web applications for communication (Nath et al., 2014). The social tools enable users

to create and share the knowledge. The social web provided a more collective and interactive web experience for users.

For example, a wiki is a collaborative website where numbers of users can co-create and co-evolve the content. Parker and Chao (2007) analyse the implementation of wikis in education as a teaching tool. The authors discuss that the rich and flexible collaboration environment of a wiki enhances peer interaction and group work, and facilitates sharing and distributing knowledge and expertise among a community of learners. Consequently, learners who use wiki or other similar social media tools (such as blogs, Facebook or Twitter) are familiar with the concept of being and working as a part of an online community to create, edit, and share content.

The number of people creating and sharing the knowledge as well as the number of people accessing, modifying and re-using that knowledge has rapidly increased with wider establishment of Internet infrastructure and the spread of its usage across the globe. Figure 1.1 shows the percentage of Internet users over time. Even though there is a big gap between the use of internet in the developed and the developing world, internet population is growing globally.

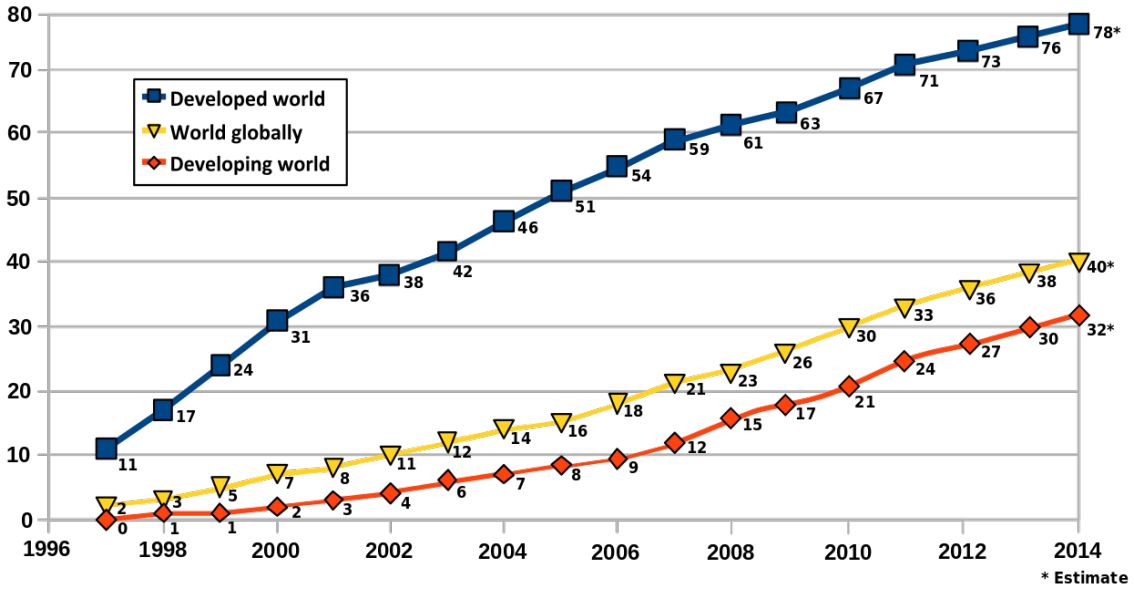


FIGURE 1.1: Internet users per 100 people over time (source: International Telecommunication Union).

The increase in Internet users has triggered discussions around open access to educational resources since the beginning of the 2000s. In 2001, Massachusetts Institute of Technology launched the OpenCourseWare project to publish all MIT course materials online. In September 2007, The Cape Town Open Education Declaration was signed

by hundreds of people in the educational fields[1]. This declaration points out the importance of free information sharing at a global level for educational purposes. In order to promote open access to digital resources, the Open Educational Resources (OER) movement has emerged and many universities have made their courses available online. They have publicly published course content and materials such as presentation slides and filmed classroom lectures (Yuan et al., 2008).

Evolution of the Web and promotion of the OER movement have also influenced the operation of distance online education. Widespread use of the Web in teaching and learning has resulted in the emergence of new pedagogies and learning paradigms in recent years. Stimulating new discussions around how people learn in a digitalised environment, Massive Open Online Courses (MOOCs) emerged in 2008 as an implementation of a so called "connectivist" learning theory, which was proposed by Siemens (2005); this put forward a new model on how we learn in a connected digital environment.

Siemens (2005) proposes that learning is a process of patterning between nodes (e.g. knowledge, information, sources and humans) to create networks; people learn by making meaningful connections amongst knowledge, information resources and ideas during the learning process. Web and social technologies facilitate the acquisition of useful knowledge and establish cognitive connections. Downes, who is the co-creator of the original MOOC, points out that learning is a result of personal experience of patterning in appropriately designed networks (Downes, 2008).

Hailed as the first MOOC (Fini, 2009), Connectivism and Connective Knowledge'08 (CCK'08) was based on Siemens' learning theory. It was launched in Autumn 2008 with 2000 people showing initial interest. In order to stimulate communication during the course, video lectures and tasks were regularly released, and group discussions were encouraged amongst participants. Siemens and Downes, therefore, provided social media tools, promoted their use and the creation of new ones to ensure that learners have enough opportunity to create their own pattern of building knowledge. These kind of massive open online courses were later named as cMOOCs.

However, not all implementations of MOOCs are following the same pedagogy or the underlying idea of connectivism and networked learning. In 2012, more loosely based instructivist MOOC platforms emerged. They are known as xMOOCs and appear to have a stronger tie with older established educational technology approaches grounded in instructivist approaches.

[1] http://www.capetowndeclaration.org

Two celebrity professors from Stanford University, Sebastian Thrun and Peter Norvig, launched their CS221 MOOC: Introduction to Artificial Intelligence (AI-Stanford) based on their classroom teaching. It was taught online alongside face-to-face delivery during 2011. Shortly after, Thrun and Norvig launched the MOOC platform, Udacity, that now offers many other free online courses mainly on technical subjects[2].

Courses developed using Udacity, and other similarly featured MOOC platforms such as EdX and Coursera are based on cognitive-behaviourist pedagogy with some small components from social constructivism. These are more centralised than the distributed connectivist model pioneered by Siemens (Kennedy, 2014). In these later frameworks, the role of instructors is similar to face-to-face teaching and the lectures are structured so that each week has a defined set of learning objectives (Rodriguez, 2012). Consequently, interactivity amongst learners and instructors is limited. In instructivist MOOCs, social tools are integrated for communication via forums and linked Facebook groups. In this approach they are typically used for asking questions about course content and assessments rather than the co-creation and co-evolution of learning content typical of connectivist MOOCs (Rodriguez, 2012).

In addition, Rodriguez (2012) found out that not all MOOCs are crafted around connectivism but social constructivist approaches have also been adopted by MOOCs. For example, the UK-based FutureLearn MOOC platform takes a social constructivist approach which argues that learning takes places through conversations.

1.1.1 Participation in Online Discussions and High Attrition Rates in MOOCs

Irrespective of differences in their pedagogies, all MOOCs appear to face the common problem of high attrition rates. Large proportions of learners who enrol on courses never participate and many others leave courses after their first visit. The 2012 study by Rodriguez (2012) identified a dropout rate of 85% in Stanford-AI courses and 40% in connectivist MOOCs. Since there is no standard metric for measuring the completion and participation in MOOCs, it can be challenging to compare different MOOCs. FutureLearn CEO, Simon Nelson, in a blog post[3], used data from FutureLearn alongside data shared by other MOOC providers. He reported a completion rate of 8% at Harvard and MIT compared to 12% in FutureLearn.

[2] https://www.udacity.com/us
[3] https://about.futurelearn.com/blog/completion-rates

This trend of decreasing participation rates, associated with low retention in MOOCs, is described by Clow (2013) as *the funnel of participation*, and appears to show similarities with the previously observed participation ratio in online discussion forums. Studies of discussions have demonstrated that no matter how high their volume, interactions are usually dominated by a small number of people, who post the largest amount of comments (Yeager et al., 2013; Jiang et al., 2014a). Even in connectivist MOOCs, it has been reported that 78% of the collaborative content was created by only 21% of its participants (Clow, 2013).

This is akin to the 90:9:1 Principle (1% Rule) proposed by Nielsen, which observed inequity in participation in online systems supporting behaviour change in that 90% of participants are passive *Lurkers*, 9% *Contributors* who contribute sparingly, and 1% *Superusers* who create the vast majority of the content[4].

Although some MOOCs have slightly better rates, completion of the majority of the course content remains low, prompting discussions around the possible reasons, impacts, and interpretations of this high attrition rate in MOOCs (Khalil and Ebner, 2014; Koller et al., 2013). There is evidence that some MOOC learners join a course only to follow one specific lecture or simply to have a MOOC experience. Koller et al. (2013) suggest, if learners leave the course before it is finished, their leaving early should not be considered as a failure or a loss to the learner, as long as their expectations have been met. On the other hand, there is some evidence that many learners leave courses even though they initially had an intention of completing. In their study, Khalil and Ebner (2014) investigate the reasons behind these high attrition rates, identifying some of the factors as follows:

- lack of time;
- loss of motivation;
- feelings of isolation;
- lack of interactivity in MOOCs;
- insufficient background knowledge and skills to cope with what is being taught in MOOCs.

Studies have shown that i) course completers are more interested in engaging with the course content and ii) learners who engage in social discussion forums are less likely to leave the course (Wang and Baker, 2015; Joksimović et al., 2015).

FutureLearn takes a social constructivist approach in order to provide an environment

[4] https://www.nngroup.com/articles/participation-inequality

that enables participants to easily reflect their opinion and interact with others for better social engagement. To achieve this, the platform inserts features facilitating social communication throughout the course adopting a Twitter-like *follow* system to help track and sustain interactive communication. Chapters 4 and 5 in this thesis present a study which analyses the use of social discussion threads and the *follow* feature and its relation to the completion status of the people who use it. This thesis proposes that identifying and encouraging *lurkers* in MOOCs to actively participate in their learning process might be useful to boost learners' engagement, and thus course completion.

1.1.2 Learning Analytics in MOOCs

In the meantime, the evolution of the Web consistently continues. Semantic web is introduced to the world as Web 3.0. The aim of Web 3.0 is to structure and link data in a way that computers can understand and process (Berners-Lee et al., 2001). While social web (Web 2.0) provides many social applications that people can use to produce knowledge and share with each other, Aghaei et al. (2012) identify the aim of the semantic web as structuring and linking data allowing the end users to discover, analyse, integrate and obtain new information. The authors describe Web 2.0 as a web of people connections and Web 3.0 as a web of knowledge.

Data mining concerns deriving high-level insights from data which is now available as semantically linked over the Web (Ristoski and Paulheim, 2016). This could be adopted to educational data mining so that an educational system could be more intelligent, more efficient and more adaptive to the needs of learners (Bittencourt et al., 2008). Subsequently, learning analytics has emerged as a new field in which sophisticated analytics tools are used to improve learning and education[5].

The Society for Learning Analytics Research (SoLAR) defines learning analytics as follows: "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*".

Among the strengths associated with learning analytics, Papamitsiou and Economides (2014) identified the ability to reveal critical moments and patterns of learning and to gain insights into learning strategies and behaviours. Hernández-García et al.

[5] http://www.learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

(2015) have completed an investigation on social learning analytics and visualisation of data to see visible and invisible interactions in online distance learning. The authors suggest that learning analytics is a link between educational data and learning to overcome the lack of physical contact in online learning and to facilitate communication through built-in synchronous and asynchronous capabilities in order to construct social learning.

1.1.3 Prediction Models and Educational Interventions in MOOCs

Khalil and Ebner (2016) discussed the potential of learning analytics to examine the rich repositories of data that MOOCs generate. The authors identified that MOOC researchers mainly use data mining techniques and statistics, as well as other learning analytics methods including text mining and linguistic analysis, visualisation, social network analysis (SNA), qualitative analysis and gamification. The researchers expect benefits from these methods to be of use in prediction, intervention, recommendation, personalisation, evaluation, reflection, monitoring and assessment improvement.

Understanding, explaining, and improving learning processes in MOOCs could be possible by applying social network analysis and learning analytics techniques to the massive scale of MOOCs data. Given the typically large numbers of MOOC participants, diagnostic analytic tools can be a particularly valuable means to inform educators about their learners progress.

Chapters 3 and 6 provide the reader with comprehensive critical analyses on the available literature in behavioural analysis, which has been achieved by using learning analytics, predictive models and their potential for the purpose of personalisation of MOOC education. Even though there are studies aimed at providing the best possible personalised MOOC experience, there is no high level personalised intervention service that is implemented by relatively big and known MOOC providers.

I am aware that this is a challenging task and raises concerns relating to privacy and ethical issues, especially in terms of accessing participants' personal data via multiple social media systems. On the other hand, people have already started dreaming of Web 4.0 as a symbiotic web which is a web of intelligence connections that requires smarter tools for interaction between humans and machines (Aghaei et al., 2012). It is believed that people will demand the same technology to be adopted in the context of teaching and learning, including MOOC education.

1.2 Research Hypothesis

This research aims to test the following hypothesis: The data extracted from participants' engagement in a MOOC can be used to identify social behaviour patterns of participants and this information can contribute to a model of course completion.

1.3 Research Aims and Research Questions

In order to test the hypothesis, there is a need to investigate participants' social and course completion behaviours, modelling them, and testing them in a prediction model.

The research questions and aims are defined in stages so that it could be helpful for navigating in the thesis.

- *Investigating participants' course completion and their engagement with the social affordances that are provided by a MOOC platform which takes a social-constructivist approach*
  RQ1: How is showing social presence in a MOOC associated to the participant's performance in course completion?
  RQ2: How do participants interact with social affordances that are provided by the FutureLearn MOOC platform?
  RQ3: How can we characterise the differences between completion rates comparing follow and discussion contribution behaviours?
- *Modelling the patterns of participants' social engagement*
  RQ4: How can we typify the different patterns of participants' social behaviours during a course?
- *Differentiating the impact of different social behaviour patterns on course completion*
  RQ5: What are the social behaviours most correlated to course completion in a MOOC?
- *Predicting participants' likely course completion based on their social engagement in the course*
  RQ6: Can we use these correlated behaviours in order to predict participants' course completion?

1.4 Methodological Approach

In order to answer the research questions, this research has been organised in three methodological stages. Figure 1.2 shows the organisation of the three methodological stages, their aims, the addressed research questions and the datasets that are used.

Methodological Approach	Aim	Addressed Research Questions	Used Datasets
Descriptive Statistical Analysis	Investigating participants' course completion and their engagement with the social affordances by using learning analytics tools	RQ1, RQ2 and RQ3	Comment, Step activity, and Enrolment datasets of **DYRP MOOC 2014**
Inferential Statistical Analysis	Modelling the patterns of participants' social engagement by using learning analytics tools	RQ4 and RQ5	Comment, Step activity, and Enrolment datasets of **DYRP MOOC 2014**
Building Prediction Model	Building a prediction model by using machine learning techniques	RQ6	Comment, Step activity, and Enrolment datasets of **DYRP MOOC 2014** and **DYRP MOOC 2016**

FIGURE 1.2: The methodological approaches in this research.

1.5 Contribution of This Research

The findings of this research demonstrated that MOOC learners who participated socially are more likely to complete the course than others. It was also demonstrated that some social features that are provided by the platform as an implementation of the social-constructivist approach are actually a good indicator for learners' patterns of engagement.

This research proposed a novel *behaviour chain* model which models MOOC participants' patterns of social engagement with the social affordances. This social behaviour modelling is valuable since it identifies which components of social behaviours might be the best discriminators for predicting their future behaviours.

By using the analysis of the *behaviour chain* model, a prediction model has been developed to predict which level of completion (*low, satisfactory*, and *high* completion) that a learner might achieve based on their previous social activities. The prediction model was tested on two different iterations of the same MOOC and the results were promising, especially in classifying learners as "to be in *low* or *high* completion". This information could be used to personalise a MOOC by making personal recommendations that would encourage behaviours that are correlated with high completion; for example, recommending people to follow.

In the course of this research a series of outcomes have arisen that make specific contribution towards the literature in technology enhanced learning and understanding social participation in MOOCs. The following points present the contribution of the publications that have been published as a part of this research:

- Since MOOCs have recently emerged, the academic studies about MOOCs have been stated only after 2011. Studies about personalised MOOCs are even newer. In Sunar et al. (2015c), I have completed a comprehensive literature analysis on personalised MOOCs. I have observed in this study that the idea of personalised MOOCs are discussed since 2011, however, the implementations of the idea have been especially reported since 2013. This work was presented in the International Conference on Computer Supported Education in Lisbon in 2015. The expanded version of this paper has been published as a book chapter in Sunar et al. (2015b). (Partly reported in sub-Section 8.4.3).
- In order to give insight into the participants' behaviours in a social-constructivist MOOC, a series of investigations has been carried out in this research. An initial study reported in Sunar et al. (2015a) was conducted to investigate the recurrent peer interactions by analysing learners' social networks. The analysis has demonstrated that most of the participants in online discussions posted once. Hence, peer interactions between learners were remarkably low in comparison to the number of comments posted to the online discussion board. It was also observed that an extremely small minority have recurrent interactions with their peers. Their interactions patterns show that if learners interacted with each other once, it appears likely that they will interact again in subsequent weeks.
- The FutureLearn MOOC platform provides a variety of social affordances to facil-

itate knowledge construction through conversations. One of them is the discussion forum which is very common in MOOCs, although there are various different designs of forums provided by MOOC providers. FutureLearn does also provide a *follow* function that people may be familiar with from their prior experience with Twitter and Instagram. The study analysed in Sunar et al. (2016) conducted an investigation on how learners interacted with the social affordances and how these interactions sustain engagement. In order to carry out this study, a descriptive statistical analysis was applied on the data. One of the contributions of this study is that the *follow* feature on the platform has a potential value of gathering information on learners' performance in the course. The contributions of this investigation can be used by the course providers to facilitate the discussion forum threads. (Partly reported in Chapter 4).

- Inferential statistical analysis was then applied after the descriptive statistical analysis. Learners' behaviours were modelled as *behaviour chains* based on their patterns of peer interactions and frequency of social activities. The modelled *behaviour chains* were then used for developing a prediction model. The results indicated that the modelled behaviours are better predictors for predicting course completion than raw features such as numbers of social interactions initiated. Other researchers may benefit from the results by adopting the features according to the available social affordances in their own MOOC platform. (It is my intention to publish on this topic after I finish my PhD.)
- Comprehensive analyses on the use of learning analytics in patterning learners' behaviours and predicting learners' course performance are presented in this thesis. Since these are very trendy topics in the MOOCs studies, researchers may benefit from the analysis of state-of-the-art techniques and up-to-date research findings. As a related area, possible educational interventions by using learning analytics and prediction models are also discussed in this thesis. The outcomes of these critical analyses indicate that learning analytics are especially useful for understanding the learners' engagement with the course. In addition, prediction models are commonly used for detecting learners who are at risk of leaving the course. There are a number of proposals that suggest use of learning analytics and prediction models for providing personalised educational interventions. (The critical analysis of prediction models used by MOOC studies has been partly reported in Chapter 6).

1.6 Outline of the Thesis

This introductory chapter has presented the motivation, objectives, and contribution of the thesis and provided some of the preliminaries, which will be relied upon in later chapters. How MOOCs emerged, what kind of advantages they may offer and what kind of deficiencies they currently have are discussed in this thesis. The lack of social interactions during the course and its high dropout rates are especially addressed.

In order to tackle the problems and research questions that are presented in this chapter, some prior literature analyses and statistical analyses have been completed. A prediction model has also proposed based on the prior studies, which are explained in the coming chapters. Figure 1.3 illustrates the logical flow of the chapters. It also summarises the genre and topic of the content of each chapter.

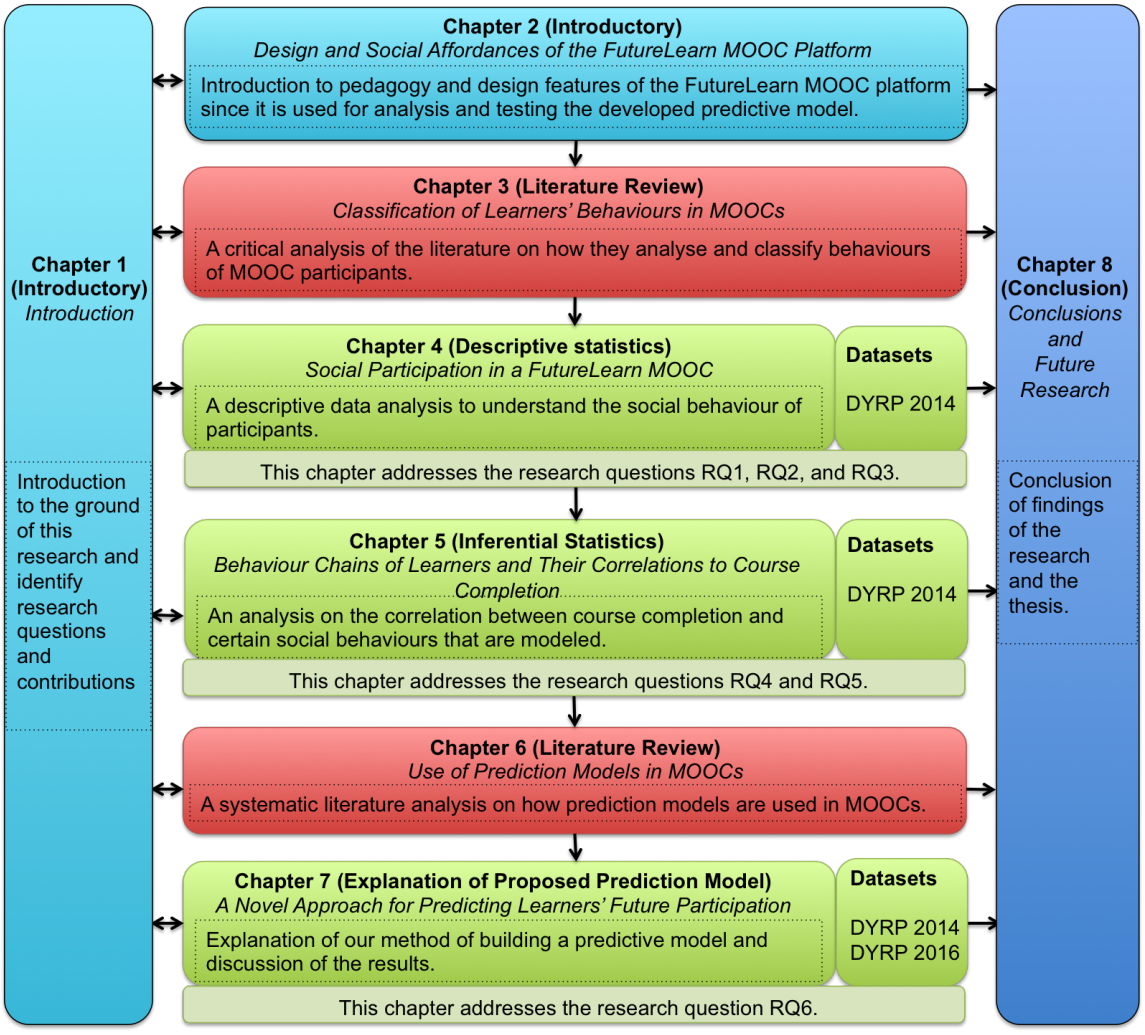


FIGURE 1.3: The structure of the thesis and interrelations between chapters.

The following points summarise the research presented here by providing references

to the related chapters in this thesis.

- Deeper discussions on the social constructivist learning approach and introduction to the FutureLearn MOOC platform including its social features (*post*, *reply*, *follow*) and datasets generated for the partner institutes are provided (Chapter 2).
- A systematic literature analysis on how other MOOC researchers analysed and classified the behaviours of MOOC participants and which techniques they have used is presented in Chapter 3.
- A descriptive statistical analysis in order to understand participants' contribution to online discussions and their use of the *follow* feature, which is one of the unique social feature on the platform, is completed. Some statistical learning analytics tools to visually present findings were used to distinguish learners' social presence and their performance on course completion (Chapter 4. It has been published in Sunar et al. (2016)).
- A study is conducted in order to describe the sequence of social behaviours of participants, which are defined with the aid of characterised use of social affordances on the platform, which is explained in Chapter 4. I then exploited inferential statistical techniques to do correlation analyses on course completion and behaviour chains. The findings imply that the behaviour chains could contribute to the prediction of course completion in MOOCs (Chapter 5).
- A critical literature analysis on implementation of predictive models in MOOCs in order to understand the state-of-the-art techniques that are used by MOOC researchers is presented in Chapter 6 (It has been published in Sunar et al. (2016)).
- The Random Forest Model and Support Vector Machine techniques have been chosen to build a model to predict MOOC participants' course completion performance by using the pattern of their social engagement in the course. Chapter 7 discusses the strength and weakness of this approach.

Chapter 8 discusses findings from the studies that have been completed and concludes the thesis (Section 8.4.3 has been published in Sunar et al. (2015c)).

Chapter 2

Design and Social Affordances of the FutureLearn MOOC Platform

2.1 Research Context

According to the latest statistics, over 58 million participants enrolled in 6850 courses have been delivered by over 700 universities in 2016[1] and the numbers are rapidly growing. FutureLearn, the UK-based MOOC platform owned by The Open University, has launched a year after the US-based Coursera, Udacity and EdX. With its over 5 million enrolled participants, FutureLearn is the fourth largest MOOC platform after Coursera, edX, and XuetangX[2]. FutureLearn has over 100 partners authoring courses including the University of Southampton[3]. As the platform claims on their website: "FutureLearn aims to pioneer the best social learning experiences for everyone, anywhere."

The Open University UK has been providing opportunities for distance learning since the late 1960s. In the early model of distance learning, students received texts and reading resources via postal service. The Open University later provided the students with radio and TV broadcasting to make sure that learners have numbers of opportunities to have best possible learning experience (Casey, 2008). Distance learning now mainly implemented online. In the UK, there has been a strong research community focus on the best way of design and implementation of online learning environments

[1] https://www.class-central.com/report/moocs-stats-and-trends-2016
[2] https://www.class-central.com/report/futurelearn-2016-review
[3] https://www.futurelearn.com/about-futurelearn

and affordances associated with learning activities (Clough et al., 2009; Laurillard, 2013; Ferguson and Sharples, 2014). The FutureLearn MOOC platform is designed by a team where eminent professors who have excessive experience in online learning, involved in. The design of FutureLearn reflects the over 50 years experience in distance learning. FutureLearn has been designed to promote learners to have best possible social learning experiences.

Section 2.2 discusses the pedagogical approach of FutureLearn and presents the design features and social affordances on the platform in a greater detail. Section 2.3 describes the structure of datasets that are provided by FutureLearn to its partners, and discusses the advantages and disadvantages of the current data structure. Section 2.4 wraps up the chapter. To follow up this chapter later in the thesis, Chapter 4 and Chapter 5 discuss the analysis of how participants engaged in the course by making use of the social affordances on FutureLearn.

2.2 Underlying Pedagogy of FutureLearn

Researchers have discussed the most suitable pedagogy for MOOCs to enable learners to have the best possible online learning experience (e.g. Bali (2014); Guàrdia et al. (2013); Mackness et al. (2013); Ferguson and Sharples (2014)). FutureLearn is seeking to develop a pedagogy that works at massive scale. The design team led by Professor Mike Sharples has implemented a social-constructivist learning theory based on Laurillard's conversational framework (Laurillard, 2013). Ferguson and Sharples (2014) highlight the advantages of using this framework in MOOCs as:

- deriving from a theory of learning rather than instruction,
- designing for interactions to be mediated with and through technology,
- embracing variety of approach to learning, such as direct instruction, networked learning, reflection, and inquiry.

Such an affordance led approach is valuable given the observation made by Brown and Voltz (2005). Additionally, Eradze and Laanpere (2014) observed that the objective and underlying pedagogical approach of a platform has an effect on the design of the platform and courses. An example of the impact of this constraint can bee seen in instructivist MOOCs where usually provide internal discussion forums as an opportunity for social interactions. However, such discussion forums are typically operated as simple Q&A threads. Consequently, those forums are a place for collectively gathering and sharing information rather than collaboratively producing new content.

Social constructivist and connectivist MOOCs, on the other hand, typically promote the use of discussion forums and other social media tools to enable participants to learn from their peers and support the processes of collaborative knowledge construction. The FutureLearn platform is designed to promote successful conversations providing participants with links between the visible repository of learning resources and a set of integrated tools which enable commenting, responding and reflection (Ferguson and Sharples, 2014). This is further discussed in the next section (Section 2.2.1).

2.2.1 Course Design in FutureLearn

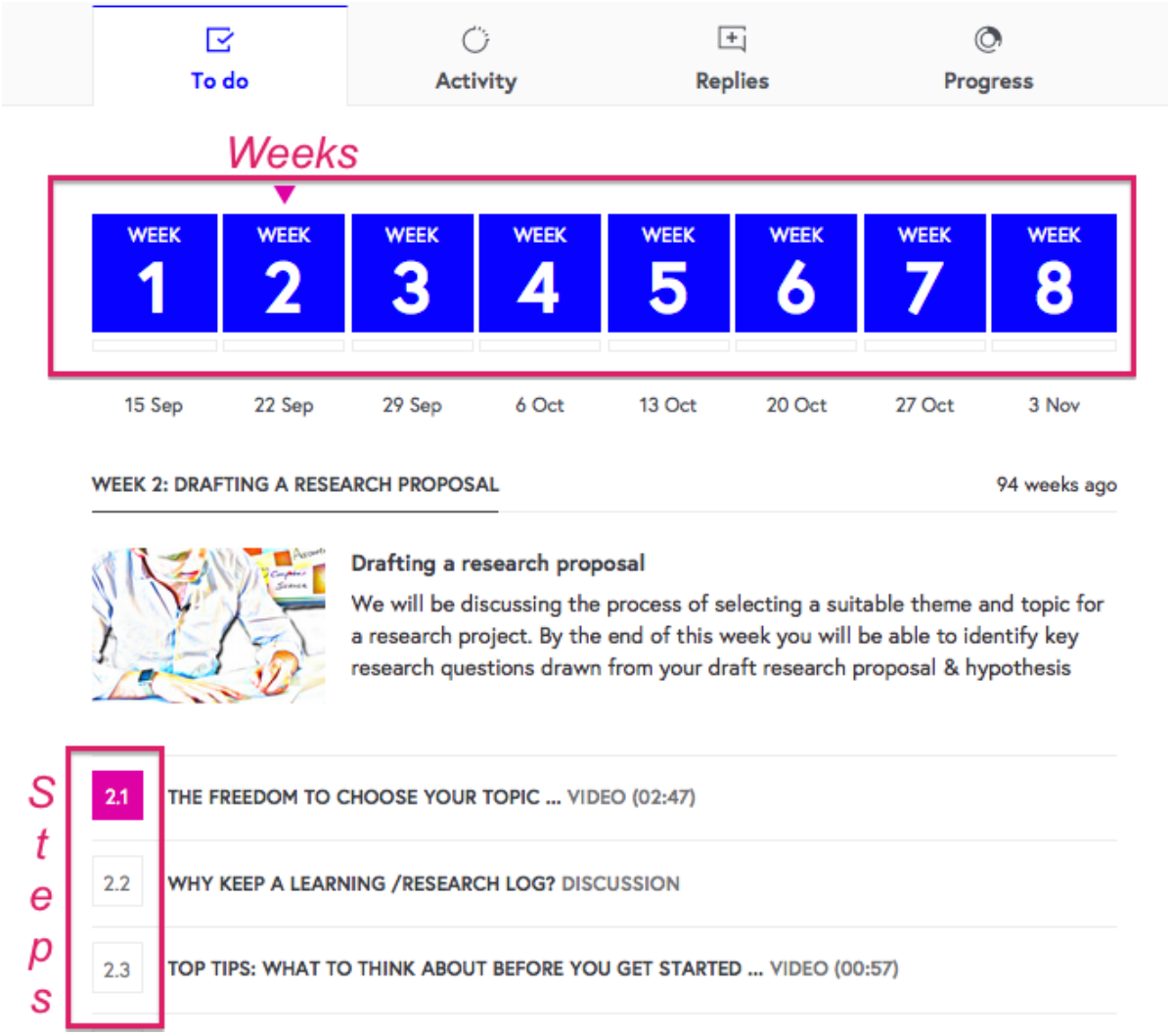


FIGURE 2.1: FutureLearn MOOCs are structured around weeks and series of steps associated with weeks.

FutureLearn MOOCs are structured in weekly components, each component containing a series of steps associated with that week. Each component contains an ordered

series of several video, and written lectures, additional resources such as links, learning activities and self/peer assessments or computer assisted assessment style quizzes. The recommended route for learners to study is a logical sequential progression in steps, but it is not compulsory. Learners can choose i) any order, ii) any steps, or iii) not to complete some steps during their MOOC study experience. They manually mark each *completed* step as they progress. Figure 2.1 illustrates the general view of a week on a FutureLearn MOOCs page.

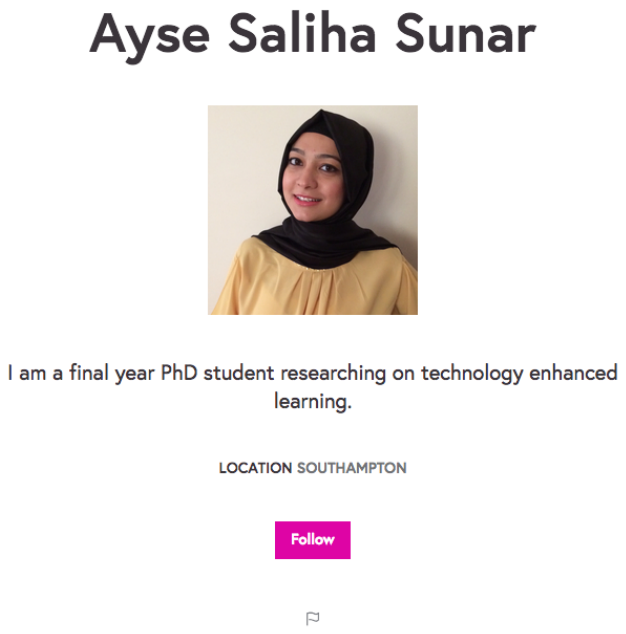


FIGURE 2.2: A participant's profile page with the option of *Follow*.

2.2.2 Social Affordances in FutureLearn: post, scroll down, like, reply, follow, filter, notify

A design feature specific to FutureLearn is its approach to prompting online discussions (Ferguson and Sharples, 2014). Each step in a week has an associated discussion thread, which realises as Twitter-like threads which enable the learners to scroll down and read sequentially through any set of associated comments. A learner can *like* a comment and *reply* to any specific comment. Additionally, learners are able to *follow* other participants in the platform by using the *follow* button. When other participants click on a participant's name on the discussion thread, the person's profile is opened with the option of *Follow*, located in the bottom of participant's profile as shown in Figure 2.2. It is also possible to follow a participant by clicking on *Follow* button appearing at the top right side of their comment on the discussion thread

(Figure 2.4). It is reasonable to expect that people will be aware of how to use these features because of their prior experience of Twitter, Youtube, Instagram and Facebook.

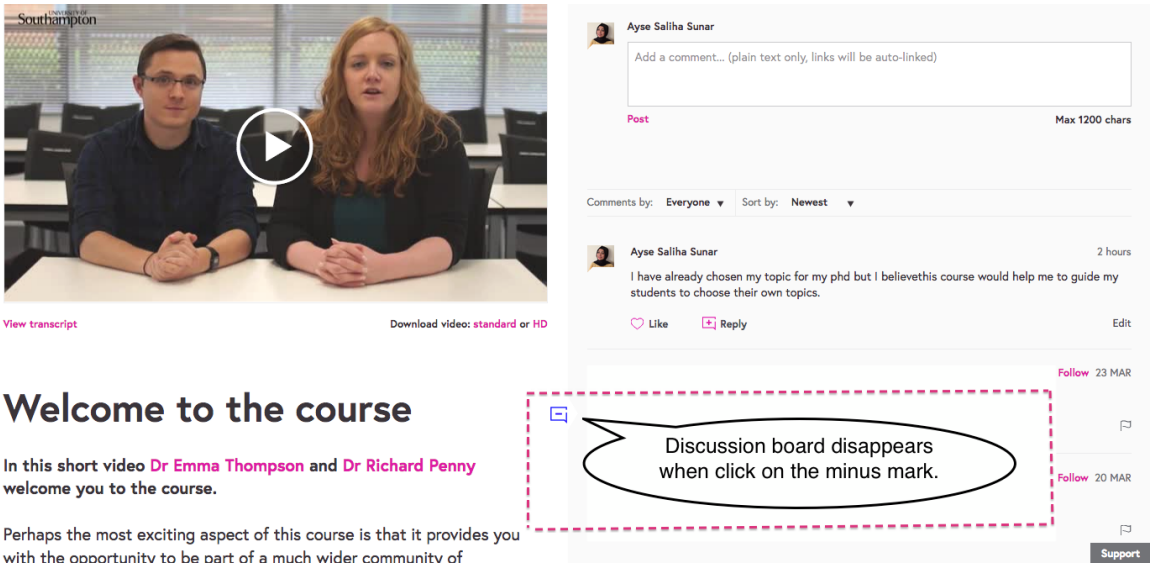


FIGURE 2.3: Discussion thread next to the course content (The 2017 version).

Figure 2.3 shows the discussion thread located next to the course content. It is possible to hide the discussion thread by simply clicking on the "minus" mark.

The design of the platform enables learners to see specific comments (i.e. comments posted by the people that they follow or most liked comments) by simply clicking on following or most liked options located next to everyone option. This allows learners to comment, reflect, share and respond. Note that the updated design of FutureLearn now has "newest" and "oldest" options to sort comments. Figure 2.4 shows the social affordances of the 2014 version of the platform's design.

2.2.3 Social Actions and Roles of Learners in FutureLearn

Table 2.1 summarises the labels that are used in this thesis for roles, social actions and status of a learner. We only consider *follow, post,* and *reply* as a social actions since *like* is not traceable from the datasets.

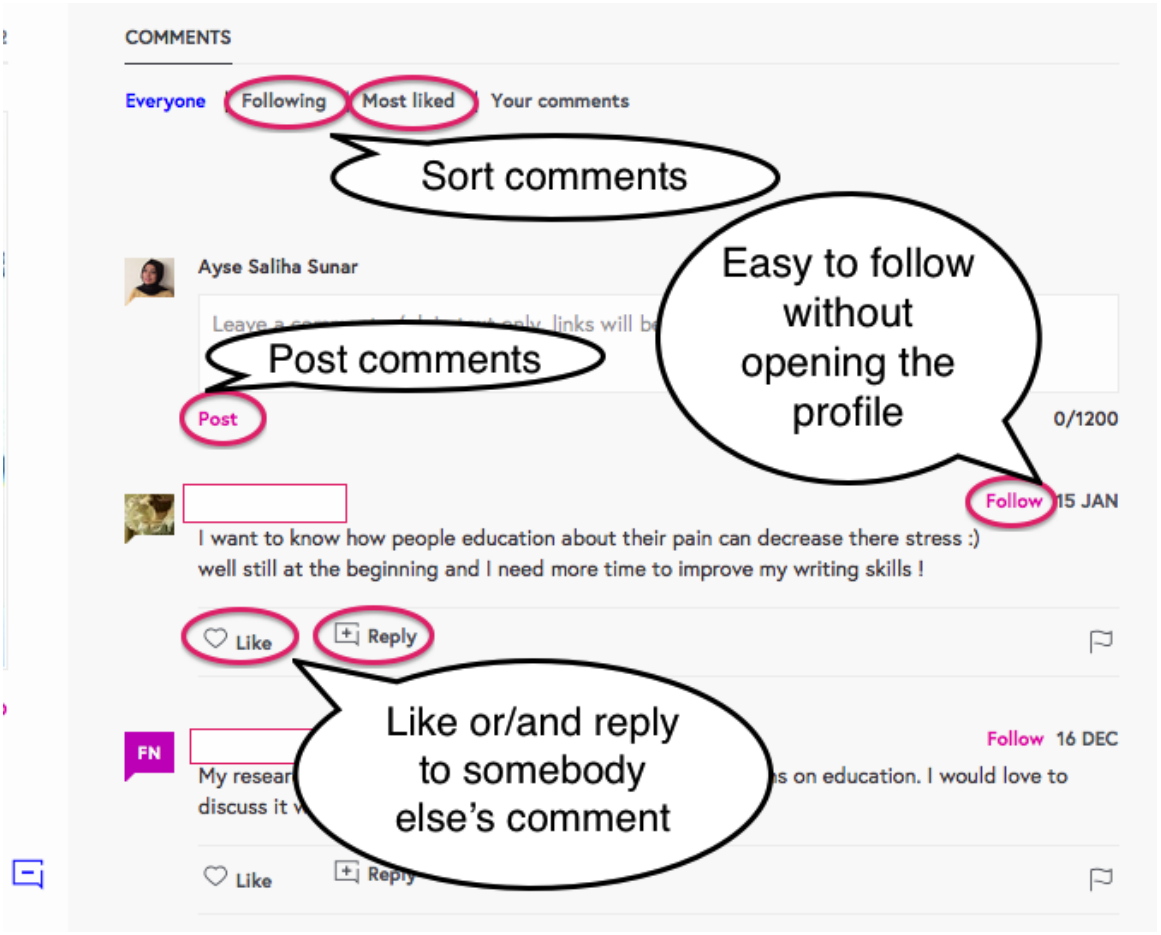


FIGURE 2.4: The FutureLearn platform, highlighting social affordances within discussion thread (The 2014 version).

2.3 Provided Datasets by FutureLearn

In order to support their partners in analysing learners' performance and MOOC participation, FutureLearn provides large amount of anonymised data which is generated from the participants' demographic and online activities during courses. Table 2.2 summarises the types and attributions of the datasets. FutureLearn provides the partner institutions with the standard datasets (*enrolments, end of course, step activity*, and *comments*) of their own courses.

In this research, the University of Southampton's *Developing Your Research Project* MOOC, which ran from the 15th September - 9th November 2014, was used for the analyses[4]. The datasets provided by FutureLearn are a snapshot of the participants' activities observed from the 15th September - 22nd November 2014. The source data

[4]This research is ethically approved by the University of Southampton. The ID for the ethics approval is 9995.

TABLE 2.1: Specific functionality and features in the FutureLearn MOOC platform

ROLES
educator: Course designers or mentors. Mentors are the experts from the ground i.e. PhD students are responsible for monitoring discussions during a MOOC. León et al. (2015) examine how mentors intervene during discussions on in FutureLearn.
learner: Participants not from the educator team
ACTIONS
follow: Action of following someone in order to be informed of comments posted by that specific learner
post: Action of posting a comment to a thread
reply: Action of replying a comment in a thread
STATUS
follower: Follows other participants in a MOOC
followee: Is followed by a participant in a MOOC
poster: Posts to discussion threads
replier: Replies a comment
course (overall) completer: Completes at least 50% of the all steps of the course. If not, classified as **course non-completer**. We chose 50% as reference since this is a part of the criterion that FutureLearn uses to define learners who *fully participated*. FutureLearn had two kinds of statements at the time this research was conducted, which were: statement of participation (The learner has marked over 50% of the steps on a course as complete) and certification of participation (The learner has marked over 80% of the steps on a course as complete). The type of certifications has been changed in 2016. In this research therefore the course completion has been divided into the two level: • *satisfactory:* Completion of the steps more than 50% but less than 80%. • *high:* Completion of more than 80% of the steps.
week completer: Completes at least half of the steps in a particular week. If not, classified as **week non-completer**.

which was analysed was a subset drawn from the standard datasets: *enrolments, end of course, step activity*, and *comments*.

This course had been chosen due to its availability at the time this research was conducted. Additionally, the number of socially active learners (1892 people) was larger than the other courses (Exploring Our Oceans MOOC: 1357, Archaeology of Portus MOOC: 1843, Web Science MOOC: 766) at the time.

FutureLearn also provided a *followings* dataset upon our request. This *followings* dataset contained *follow* interactions amongst participants between the first day of FutureLearn and 2015-09-16 09:45:43 UTC, tracking around 1.2 million relationships in the platform. I examined those data associated with the instance of the DYRP course selected for this study (2927 items). The DYRP participants who already initiated *follow* interactions in a previously run MOOC are also included. However, it should be noted that any two participants could take part in more than one MOOC, which are run in the same time period with DYRP. In this case, it is difficult to distinguish that during which course the learner decided to follow the other. Since the dataset does not include the information of which link directed a learner to follow someone, I was only able to draw a subset from the *followings* dataset by using time and learner ID information.

Table 2.2: List of FutureLearn Datasets and their Attributes (The 2014 version)

End of Course Stats
Overall participation rates in a MOOC i.e. number of those enrolled in the course and those who left the course
Enrolments
Enrolment records of participant Attributes:*learner_id, enrolled_at, unenrolled_at*
Step Activity
Number of steps completed by learners i.e. those checked the "completed" mark Attributes: *learner_id, step, week_number, step_number, first_visited_at, last_completed_at*
Comments
Records on the forum activities. This dataset identifies who posted: whether it was a reply, post timestamp, content and number of likes received. Attributes: *id, author_id, parent_id,step_ text, timestamp, likes*
Followings
Records on *follow* relationships amongst participants Attributes: *followed_user_id, follower_user_role, follower_user_id, follower_user_role, created_at*

2.4 Summary

FutureLearn having currently over 6 million participants is a MOOC platform which takes social-constructivist learning approach. Accordingly, the design of the platform provides social affordances that enable participants to share their opinion, reflect their opinion on others, and interact with each other. The platform provides associated discussion threads in each learning unit (called *steps*) where participants share their comments with other fellow learners and reply to somebody else's comments. In addition, participants can follow the discussions that are posted by those participants who they follow.

Even though the platform provides these affordances for escalating social learning through social engagement, there is a need for investigating its real impact on learning. In fact, there are certain things that we cannot know, such as if a learner read the posts from others, what is the motivation for them to follow others, and if the social experience helps them to stay longer on the course. However, there are some factors that can be measured by using the available data. For example, the correlation between social engagement and course completion, the frequency of certain behaviours, who follows whom and so on. Chapters 4 and 5 present the study conducted to make sense of social engagement of learners by using the generated datasets from the participants' online activities.

The chapter also provided some prior information about the frame of datasets which are used in the analysis and visual representation of the platform that may be necessary to comprehend the research that is presented in this thesis.

Chapter 3

Analysis and Classification of Learners' Behaviours in MOOCs

3.1 Introduction

Understanding learners' current progress, identifying their needs, and anticipating their future performance are vital for designing and implementing an effective educational interventions (Chatti et al., 2012). Observing learners' gestures and performance is relatively easier and fast in a classroom environment, but is challenging in distance online learning. Especially if the scale of learning is massive as it is in MOOCs, identifying such features of each individual is not currently feasible. Studies have analysed the value of speech and gesture recognition in collaborative online learning (Nihonyanagi et al., 2014; Nakano et al., 2015), however, it is not currently applicable to MOOCs.

Course creators e.g. teachers, mentors, course providers conduct regular surveys of their learners progress in order to try to diagnose learners needs, and provide them with customised interventions to meet their needs, evaluate the course, and alter or improve the course design in order to improve the quality of a course. Hung and Zhang (2008) investigate the use of data mining techniques in online teaching. Their study indicates that ways in which learners behavioural patterns can be identified by using data mining techniques. They suggest that there is an important role in online education for this approach which can enable educators to intervene in learning processes. Therefore, researchers make enormous efforts to understand learners' behaviours in online education so that it would be possible to track learners and to evaluate their

performance (Zaiane and Luo, 2001). Classification of learners behaviours is one of the methods that facilitate making predictions about learners performance and allows us to provide them with virtual help if necessary (Romero and Ventura, 2007).

The remainder of this chapter further focuses on these approaches specifically in the context of MOOCs. Section 3.2 presents motivations, state-of-the-art techniques and findings of the studies in the recent literature. Section 3.3 summarises the findings that are presented and concludes the chapter.

3.2 Critical Analysis on Classification of MOOC Learners' Behaviours

Researchers have established approaches to classify course participants mainly based on the activities and achievements which can be used by providers of the courses in managing their learners' experience in a course. This allows researchers to classify each individual as part of a fraction of the whole group without precisely identifying the learning progress of each individual.

Table 3.1 shows that main motivation of identifying and classifying learners' behaviours in MOOCs can be divided into the three categories:

1. to gain insight into learners' engagement in courses
2. to predict participants' future performance in courses
3. to make interventions in participants' learning activity when it is necessary

Table 3.1: Three main motivation to classify MOOC learners' behaviours.

Motivation of classifying behaviours	***Studies***
To better understand learners' engagement in a course	Milligan et al. (2013), Coffrin et al. (2014), Yang et al. (2014a), Ferguson and Clow (2015), Sharma et al. (2015), Gelman et al. (2016), Wang et al. (2016)
To make predictions about learners' future performance in a course	Coleman et al. (2015), Xu and Yang (2016)
To make interventions in learners' learning process in a course	Kizilcec et al. (2013), Anderson et al. (2014), Gillani et al. (2014), Hmedna et al. (2017)

The motivation of this research can be categorised as both *To better understand learners' engagement in a course* and *To make predictions about learners' future performance in a course* since this research uses the knowledge from the insight of learner behaviours to predict their course completion.

Table 3.2 analyses numbers of examples from the literature about how learners' classification is handled using course participation of learners by researchers.

Commonly used approaches to reach a categorisation apply statistical methods. For example, through learning analytics and machine learning techniques based on participants' behaviours predominantly via data derived from logs of their interactions, and data about which pages and links participants visited, which is also known as *clickstream data*.

The analysis of the literature (see Table 3.2) shows that four main behaviours of learners in a course are usually taken into consideration to apply statistical methods.

- Behaviours in videos: length of watching, pauses etc.
- Behaviours in discussion forums and other social media tools if available: the forum page visits, contributions to discussions etc.
- Behaviours in assignment submissions: timely submission etc.
- Behaviours in the course structure: progress as measured by sequence of links clicked on during the interactions with the course

Some researchers take a single behaviour, some take multiple behaviours. For instance, Kizilcec et al. (2013) and Coffrin et al. (2014) consider timely assessment submissions for classifying learners whereas Xu and Yang (2016) and Gelman et al. (2016) include learners' visit to forums and wiki pages as well.

In order to understand the learners' behaviours from different angles, some researchers exploit additional information extracted from survey and questionnaire data. For example, Hmedna et al. (2017) aim to identify learners' preferences and learning styles to provide them with an appropriate learning resource recommendation. Therefore, they asked the learners to fill in a questionnaire about their preferences. Then, they applied machine learning techniques to learners' browser histories to measure their behaviours.

Facilities of the platforms have strongly influenced the approach to categorising participants. For example, Anderson et al. (2014) used *forum badges* for identifying level of course engagement of learners, which is a rarely-used social facility on MOOC platforms.

Since this research applied statistical methods to some data collected on FutureLearn, I would like to further analyse the study of Ferguson and Clow (2015) which investigates the commitment of MOOC learners on the FutureLearn platform.

The authors investigated the patterns of engagement and disengagement with the same method that was applied by Kizilcec et al. (2013) for Coursera MOOCs and looked to see if the participation was influenced by design and pedagogy of the platform. Kizilcec et al. (2013) analysed engagement patterns based on learners' behaviours in videos and assessments.

However, Ferguson and Clow (2015) indicated that learners' behaviour in discussions is also an important factor in FutureLearn since the knowledge is jointly constructed through conversations in a social-constructivist MOOC. Therefore, Ferguson and Clow (2015) take into account i) active engagement with course discussion alongside with ii) active engagement with course content and iii) active engagement with course assessment to investigate the engagement patterns. The authors observed that participants engaging with the comments show a more extensive engagement with the course materials and assessments.

In this research, I am interested in some mathematical modelling to have insight into learners' social behaviours which would be valuable for further interventions. This research therefore has taken into consideration i) active engagement with course discussion and ii) active engagement with course content to investigate the impact of social engagement on course completion.

In their study, Ferguson and Clow (2015) considered social behaviours as only posting comments to threads, however, this research considers posting an original comment and replying to somebody else's comment as different social behaviours. The *follow* behaviours of learners are also considered as one type of social behaviour.

3.3 Summary

In conclusion, MOOC researchers use different techniques to understand their participants' online behaviours and classify them. Researchers use numbers of different factors to identify the patterns of behaviours in the course. Some of those factors are behaviours in social discussions, time spent viewing videos, timely submission of assessments and so on.

Since conversation is one of the core elements needed to construct knowledge in a social-constructivist MOOC, this research mainly focuses on participants' engagement with course content and course discussions.

In this thesis, Chapters 4 and 5 present the novel contribution of this research on applying learning analytics methods to gain insight into social engagement of participants on FutureLearn and Chapter 5 specifically classifies learners based on their level of social engagement and presents an analysis on correlation to course completions.

TABLE 3.2: Use of social activities for classification by researchers

Study	***Findings***
Kizilcec et al. (2013)	Social participation is not considered for classification. Timely assessment submission is considered.
Milligan et al. (2013)	They classically classify learners as *active*, *passive* and *lurker*. Blog and Twitter users are active participants. They build internal and external networks.
Anderson et al. (2014)	Participation in discussions and acquiring forum badge are considered for identifying level of course engagement of participants.
Coffrin et al. (2014)	Social participation is not considered for classification. Assessment submission is considered and learners are classified as *auditors*, *active*, *qualified*.
Gillani et al. (2014)	They explore the communication communities in MOOCs. Learners are classified based on their contribution to two sub-forums (cases and final) such as *discussion initiators, individualist learners, help seekers, community builders, and project support seekers.*
Yang et al. (2014a)	Twenty different factors are set to identify sub-communities in discussions. Most ranked words are identified which are associated with course attrition.
Coleman et al. (2015)	Learners' click-stream data including forum page visits is used. They classify them as: *shopping, disengaging, completing*. No other social contribution is particularly considered.
Ferguson and Clow (2015)	Posting to discussion threads is considered as social behaviour. Participants are divided into clusters (e.g. *mid-way dropouts*) and their engagement patterns are identified (e.g. half in the *mid-way dropouts* cluster contributed to forums).
Gelman et al. (2016)	Number of posts and length of posts are considered. Since forum participation is significant in the first week, this social behaviour is only considered for *introduction* behaviour. Rare forum users are also covered in *sampling* behaviour.
Sharma et al. (2015)	Users' visit to forums and wiki pages is considered along with their quiz submissions.
Wang et al. (2016)	Contents of comments are also considered along with having contributed to discussions for classifying higher-order thinking learners.
Xu and Yang (2016)	They classify learners based on their motivation and grades and predict whether or not they are going to earn certificate. Participants' level of contribution to forums and wiki pages is considered. For example, whether or not a learner is a *passive* or *fully contributor*.
Hmedna et al. (2017)	They classify learners according to learners' learning styles to provide recommendations of appropriate resources for each cluster. They do not use social contributions as a factor.

Chapter 4

Social Participation in a FutureLearn MOOC

4.1 Introduction

There are various ways in which participants respond to and use the social affordances provided by FutureLearn: *post* a comment, *reply* to a comment, *like* a comment and *follow* a fellow participant. As Chapter 3 explains, the pattern of participant interactions could be an important key feature for designers and course providers to understand their learners.

This chapter is designed to answer the first three research questions addressing *investigating participants course completion and their engagement with the social affordances that are provided by a MOOC platform which takes a social-constructivist approach. The research question are stated in Section 1.3 as:* ***RQ1****: How is showing social presence in a MOOC associated to the participant's performance in course completion?*
RQ2*: How do participants interact with social affordances that are provided by the FutureLearn MOOC platform?*
RQ3*: How can we characterise the differences between completion rates comparing follow and discussion contribution behaviours?*

In order to answer these research questions, this chapter aims to investigate the social engagement of FutureLearn MOOC participants by applying learning analytics techniques. The datasets that are introduced in Chapter 2 are initially used for a descriptive statistical analysis. The findings show that the majority of the learners do

not make social contributions. Also, weekly social contributions gradually decrease towards the end of the course. In addition, it is observed that the majority of social contributions have been made by participants who completed at least half of the course steps.

The outline of the chapter is as follows. Section 4.2 presents the analysis and its results. Subsection 4.2.1 gives the general statistical results on the course participation. Subsection 4.2.2 specifically analysis the contributions of participants who follow someone to discussions. Subsection 4.2.3 analyses what percentage of socially active participants completed the course.

This section also investigates if there is any differences between completion rates of learners according to the type of social features that they have used. Finally, Section 4.3 concludes the chapter.

4.2 Analysis and Results

The University of Southampton's *Developing Your Research Project* (DYRP) MOOC, which ran from the 15th September - 9th November 2014 was used for analysis. The datasets provided by FutureLearn are a snapshot of the participants' activities observed from the 15th September - 22nd November 2014. The specific data which was analysed was a subset drawn from the datasets: *enrolments, end of course, step activity*, *comments*, and *followings*. Section 2.3 gives a detailed look to the types and attributions of the datasets for reference. In order to accomplish the analysis, the data was mined by using Python and the outcome was visualised by using GLE, Matlab, and R.

4.2.1 General Statistics on Social Participation

Figure 4.1 summarises the "funnel of participation" in the DYRP course.

After the course was announced, 9855 learners enrolled, 5086 (51.6%) participants actually visited the course pages after the course started.

- Of these enrolled learners, 3852 (39%) completed at least one step.
- 2631 (26.7%) revisited the course and completed further steps.

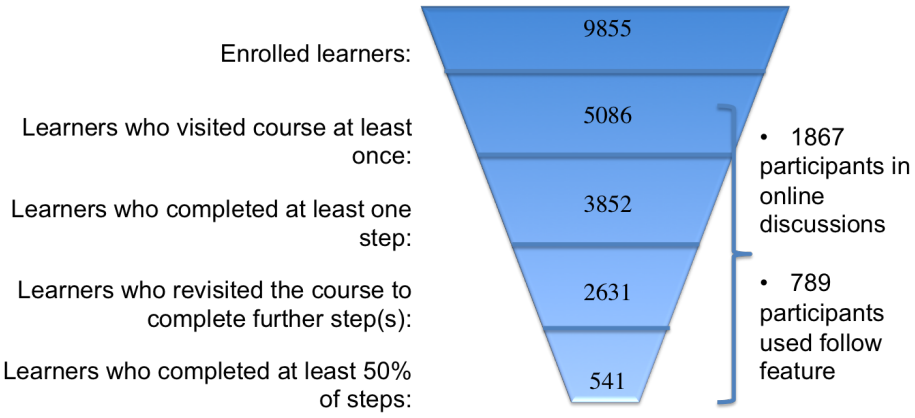


FIGURE 4.1: Funnel of participation as observed in DYRP MOOC (Sunar et al., 2015a).

- In total 1867 (18.9%) learners participated in online discussions by writing at least one comment.
- In addition, 789 (8%) participants followed discussions of one or more other course participants.

In this thesis, participants who interacted with any social affordances i.e. writing a comment, replying to a comment, and following someone, at least once is considered as *socially active*. In order to answer RQ1, which is *How is showing social presence in a MOOC associated to the participant's performance in course completion?*, the correlation between course completion and social presence. I would like to remind the reader here that course completion implies completing more than 50% of the course steps.

The analysis suggests that there is a high correlation (0.50 positive correlation according to Pearson's Correlation Test) between course completion and social presence. This is consistent with previous findings in other similar studies by other MOOC researchers. as numbers of other studies suggested. The analysis indicates that the majority of learners who did not initiate any social activity did not complete any steps at all. Figure 4.2 comparing the step completion of socially active and inactive participants in the course. Figure 4.3 shows the d

In a boxplot graph, lower and upper whiskers show the least and greatest value excluding outliers in the distribution. The median value slicing the box shows the middle of dataset which 50% of data is greater than the median value. Lower quartile (lowest value in the box) shows that 25% of data is less than this value; upper quartile (greatest value in the box) indicates that 25% of data is greater than this value. In Figure 4.2, the box representing the socially inactive learners appears as a line rather

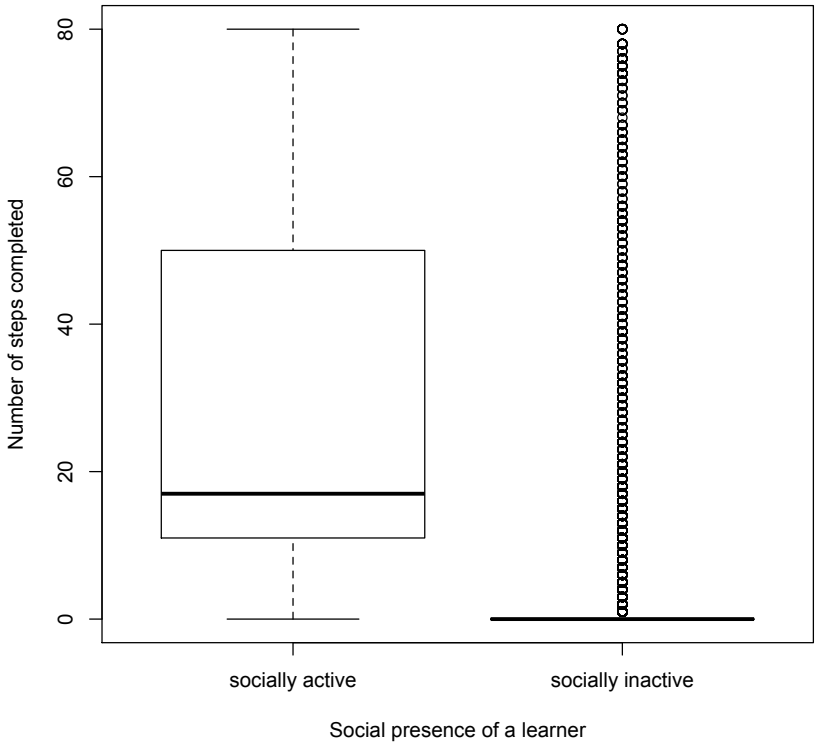


FIGURE 4.2: Comparison of course completion of participants who are socially active and inactive.

than a box. This is because that almost 100% of the distribution have the same value except the outliers which are represented by small circles.

Even though the number of completed steps for each group is various, the majority of social learners completed 10 to 50 steps while the median value is slightly below 20 steps.

Before investigating which behaviours are specifically correlated to the course completion in the next chapter, the remainder of this chapter will investigate how MOOC participants engaged with the social affordances of the course to address RQ2 and RQ3, which are:

RQ2: How do participants interact with social affordances that are provided by the FutureLearn MOOC platform?
RQ3: How can we characterise the differences between completion rates comparing follow and discussion contribution behaviours?

Figure 4.4 and Figure 4.5 illustrate that the volume of follow interactions accompanies

FIGURE 4.3: Number of completed steps by socially active and inactive learners (Total number of steps: 80).

a decline in resources accessed and weekly progress, occurring alongside the previously identified decline in discussion contributions (Sunar et al., 2015a). Figure 4.4 shows the number of activities initiated over the eight weeks. Figure 4.5 illustrates the number of people who initiated those activities. It shows a weekly breakdown of the follow interactions of the 789 learners and discussion contribution of 1867 learners. The largest volume of follow interactions occurred in Week 1, it had the largest number of i) participants who completed course activities; and ii) comments posted to the discussion forums.

Figure 4.6 shows the distribution of learners according to circumstances when they began following someone (before the course, during or after the course concluded). Since some learners had previously participated in a FutureLearn course(s), these learners may have already followed some individuals who also went on to participate in the DYRP MOOC. Hundreds of such participants already had a follow relation when they enrolled in the DYRP MOOC.

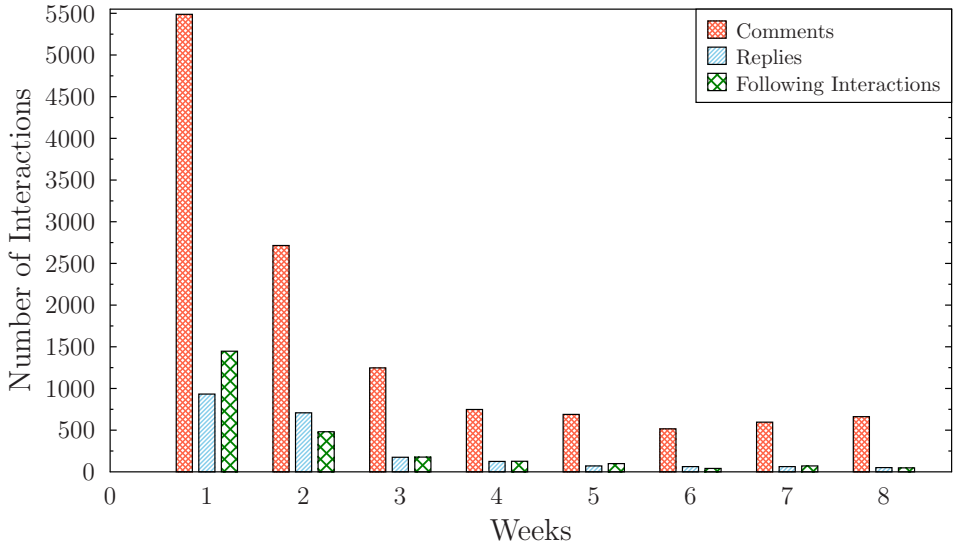


FIGURE 4.4: Volume of weekly social activities: *comments*, *replies*, and *followings*

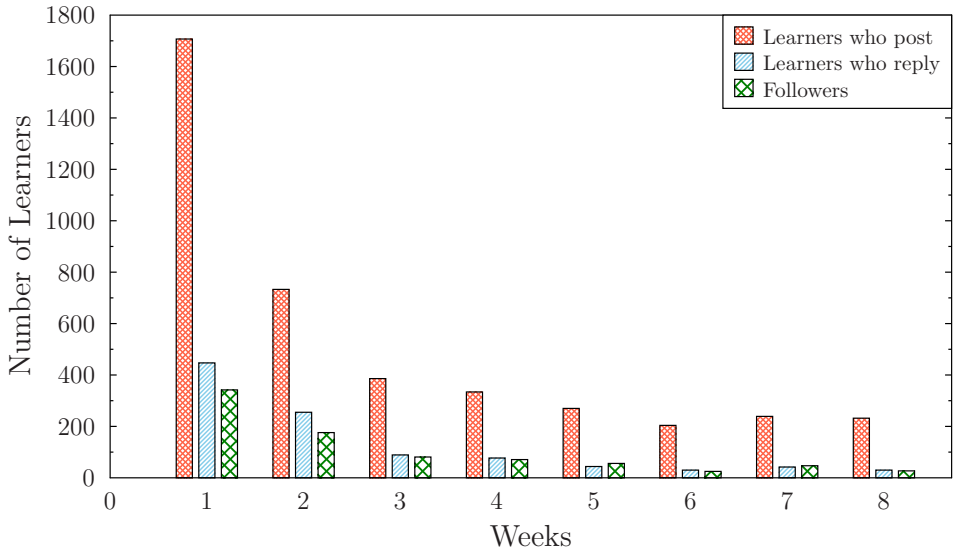


FIGURE 4.5: Volume of weekly participants who are either a *poster*, *replier*, or *follower*

It could be reasonable to assume that since these participants had already taken another MOOC together and had then subsequently enrolled together on the DYRP MOOC, they might be more likely to interact with each other. Nevertheless, our investigation shows that none of these prior experienced FutureLearn MOOC participants ever interacted with each other during the DYRP course. Indeed, interacting with each other in a previous MOOC, showing interest and enrolling in the same course as each others' again does not guarantee that these learners would be interested in each other's comments one more time. Additionally, it is observed that a small number of learners who joined the course late, and started following other learners shortly after

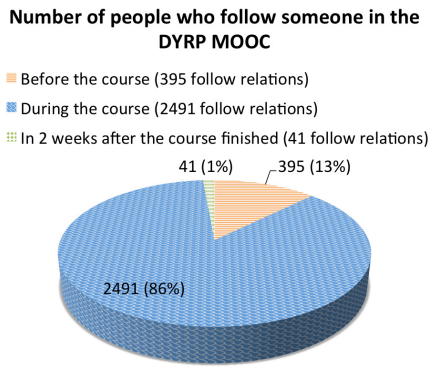


FIGURE 4.6: Learners according to the time they start following somebody.

the official end date of the course (Figure 4.6).

4.2.2 Involvement of Followers in Discussions

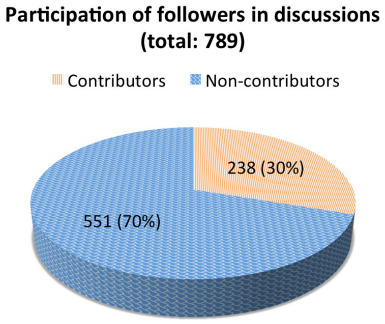


FIGURE 4.7: Proportion of those who followed who contributed to discussions.

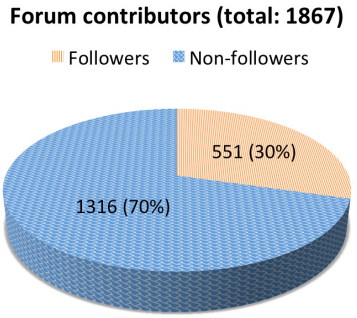


FIGURE 4.8: Comparing discussion contributions between whose who did or did not follow others.

Figure 4.7 to 4.9 examine the contributions to discussions in relation to whether the

participants chose to follow others. Participants' preferences of using social affordances is various i.e. a participant may use all the social features available on the platform or they may choose only one or two.

The majority (70%) of those participants who followed at least one other person contributed to the discussions. At the same time there was a small number of participants who commented extensively but who did not follow any other participants, 70% of all forum contributions were generated by those participants who followed no one.

Figure 4.9 provides an overview of the size of each individual's network in discussion forums and the number of people that they follow.

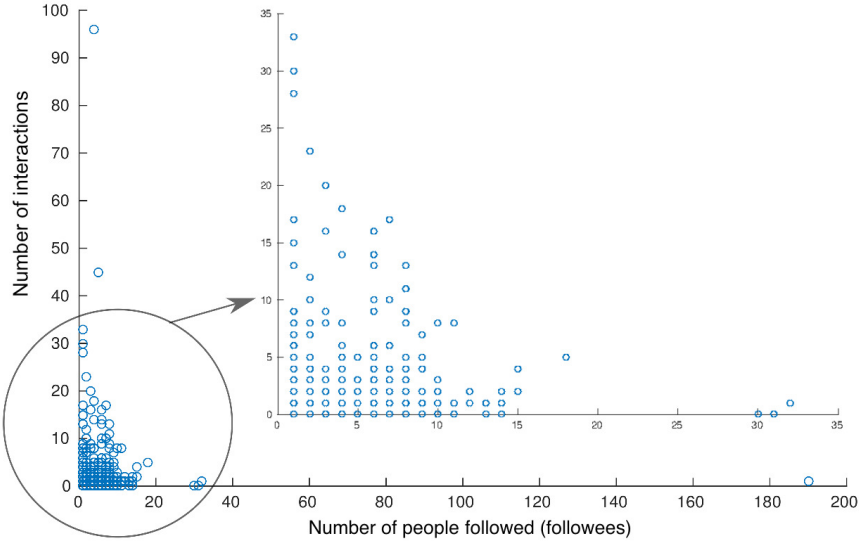


FIGURE 4.9: Comparing the number of people whom a learner follows and the number of people with whom a learner interacted in the discussion forum.

4.2.3 Completion Success of Participants

The next step in the analysis was to examine course completion success amongst the socially active learners. The DYRP MOOC is composed of 80 steps spread across eight separate weeks (see Table 4.1). The number of weekly steps varies. For example, while Week 8 has 13 steps, Week 4 has only 6 steps. In order to provide a consistent representation, the proportion of steps in each week is analysed rather than the actual number of steps. Learners who completed at least 50% of the steps in a week are considered as a *completer* of the week; otherwise, the learner is named as a *non-completer* of the week.

TABLE 4.1: The number of steps in each week.

Weeks	1	2	3	4	5	6	7	8	Total
Steps	11	12	12	6	7	8	11	13	80

Figure 4.10 shows the proportions of completer and non-completers of the course, categorised according to their social activities. Learners are allocated across five categories, which are: i) learners who follow (aka *follower*), ii) followers who contribute to the discussions, iii) followers who do not contribute to discussions (aka *lurker*), iv) learners who contribute to discussions by posting (aka *poster*), v) and posters who do not follow.

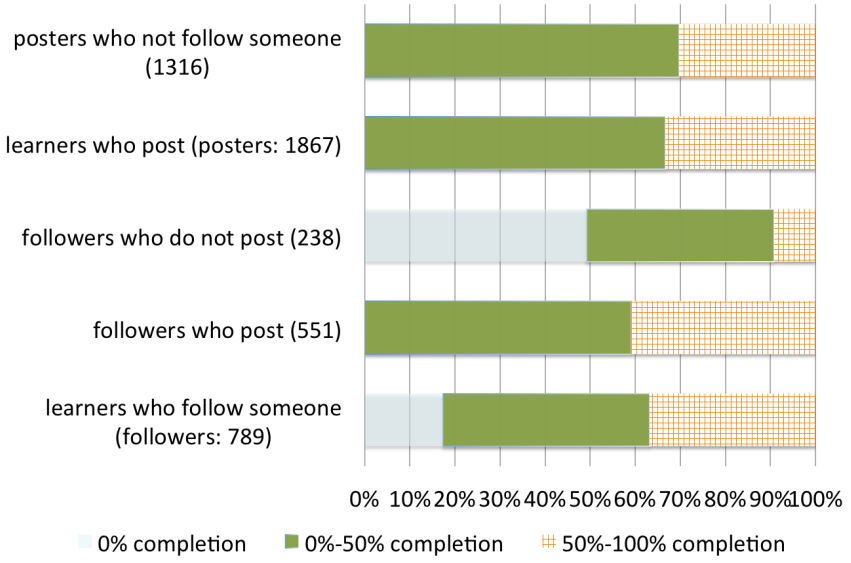


FIGURE 4.10: Proportions of completers and non-completers of learners in different categories.

One important observation is that every learner who posted to a discussion thread completed at least one step in the course. As shown in Figure 4.10, if a learner is socially passive, it is likely that they will complete none of the steps, i.e. over 40% of socially passive followers did not complete any of the steps. The proportion of course completers is high if learners are socially active. Moreover, a larger proportion of course completers (41%) is observed amongst the learners who follow and post. The learners who either post or follow make up a similar percentage, slightly over 30%.

Figure 4.11 plots the ratio of completed steps on a week-by-week basis for the learners. Learners are categorised by three distinct behaviours: i) those followers who post to discussions, ii) those followers who do not post to discussions, and iii) those posters who do not follow. The followings are observed:

- Learners in each of the categories, regardless of whether or not they are an overall course completer, progressed through the individual weekly steps at different rates.
- Fairly high weekly completion rates [from 60% up to 95%] are observed for all learners in each category throughout the course.
- The only exception is in the last week where the average fell to slightly over 30%.
- Followers who contributed to discussion threads completed the highest number of steps and represent the largest proportion of overall completers (Figure 4.10).
- Posters who did not follow anyone completed more of steps than the followers who were socially passive in discussions.

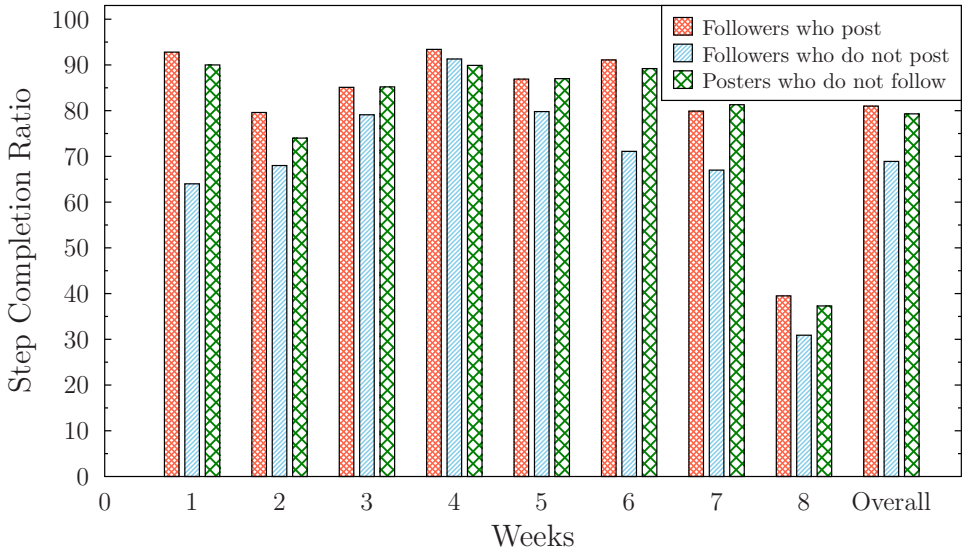


FIGURE 4.11: Average percentages of the completed steps by learners in different categories.

The average step completion in DYRP is 26 of 80 available steps (slightly over the 30% of the steps) (Sunar et al., 2015a). However, *followers* who did not contribute to the discussion forum also performed better than the course average in completed steps (Figure 4.11), implying that follow behaviours of learners could be used as an indicator for predicting their course completion.

Figure 4.12 and Figure 4.13 trace the activities on a week-by-week basis of every participant who was a follower (789 learners). Possible activities include completing a week, following, contributing to discussions, or combination of these activities (see Table 2.1 in Chapter 2).

These activities are shown with the aid of colour code, which has been chosen to remain readable when rendered or printed in black and white. Yellow (code 7, lightest) represents no activity; learners who neither participated in discussions nor fol-

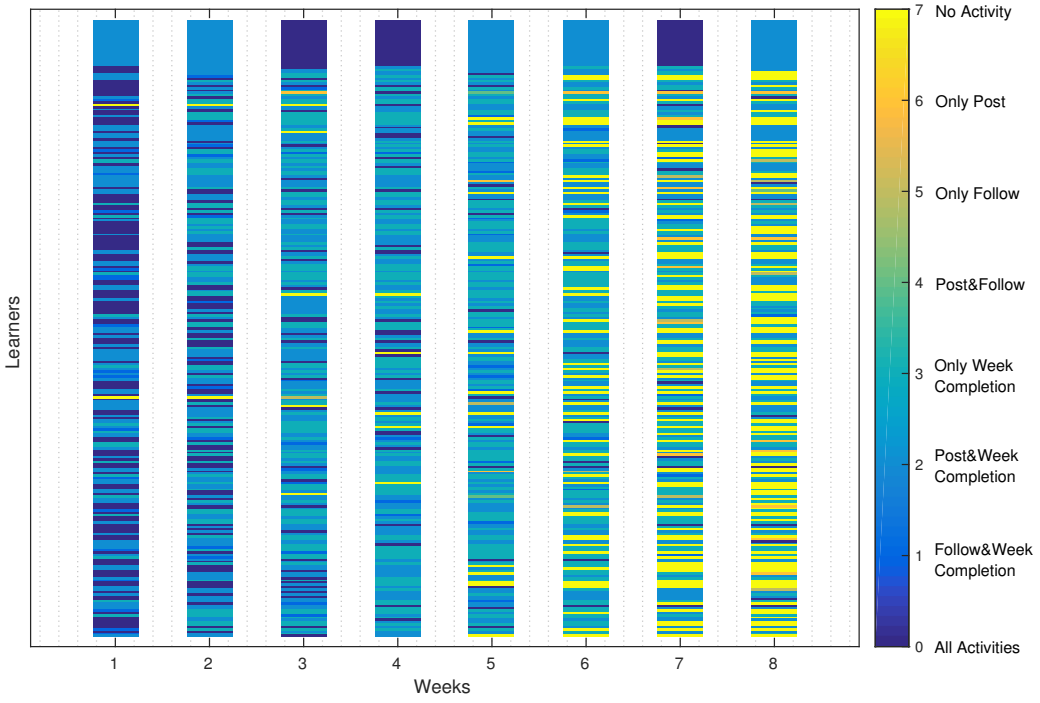


FIGURE 4.12: Activities of completer learners [who followed at least once] throughout the course week-by-week (key on right).

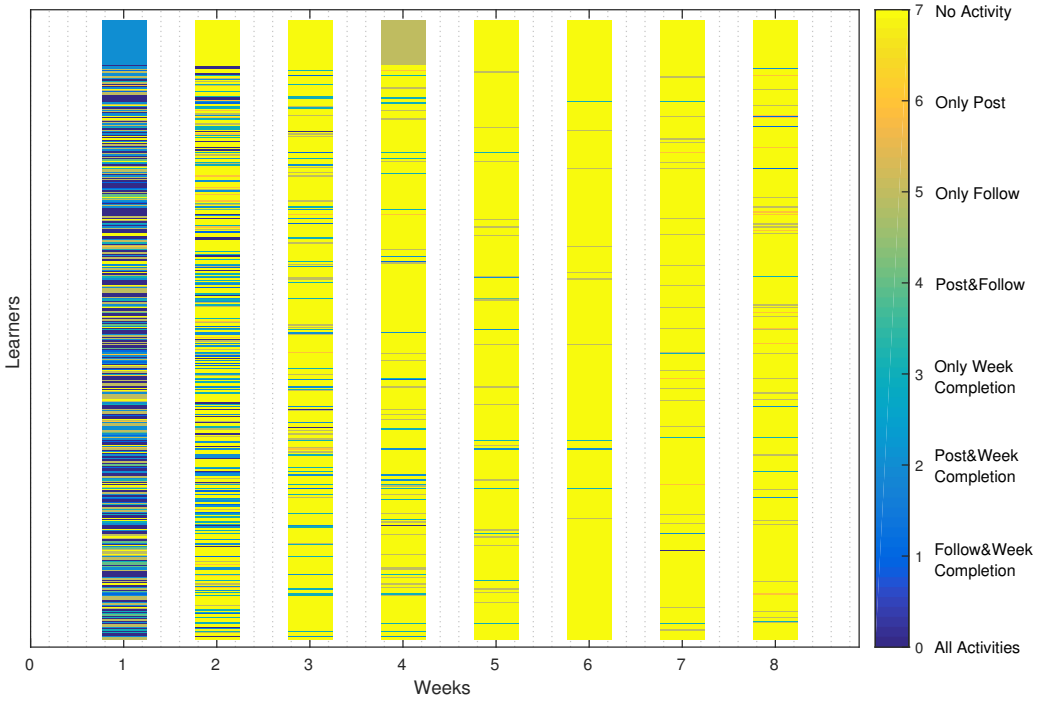


FIGURE 4.13: Activities of non-completer learners [who followed at least once] throughout the course week-by-week (key on right).

lowed anyone and did not complete more than 50% of the steps. Orange (code 6) and lime green (code 5) show learners either contributed to discussions or followed someone, but did not complete the week. Green (code 4) represents socially active non-completer learners. Turquoise (code 3) shows socially passive completer learners. Light blue (code 2) shows the learners who completed the week and were active in the discussions. Blue (code 1) represents learners who completed the week and followed someone. And finally, dark blue (code 0, darkest) shows learners who initiated all the possible activities. In a nutshell, the darker the colour, the more intense the learners' participation.

Figure 4.12 shows the activities of completing followers while Figure 4.13 shows non-completing followers. Although there are some similar behaviours amongst learners, the predominant activity profile for completers and non-completers are distinctive.

Course Completers: The social activeness of the course completers were sustained until Week 6. After Week 6, they showed limited activity. They hardly posted or followed other participants or completed the week. Full participation based on three behaviours (post, follow, step completion) was most prevalent in Week 1.

Course Non-completers: They have also been the most active in Week 1. Their level of activity and weekly completion declined sharply in Weeks 2 and 3 i.e. this is much earlier than course completers. Although no activity was observed in common especially after Week 3, it is still seen that a few of the non-completers kept on following someone or contributing to the discussions or very rarely completing the weeks. It appears that their behaviours are in accordance with the behaviours of lurkers in general discussion forums (van Mierlo, 2014). This guides me to think that the learners who read the comments and followed other participants or only concentrate on completing the steps become lurkers as the 90:9:1 principle proposes.

4.2.4 Mentors in the Data

In FutureLearn MOOCs, a mentoring team involving course designers and mentors, monitors the discussions and intervenes if necessary (León et al., 2015). Each member of the mentoring team has a unique id on the platform as the rest of the participants have. Therefore, course activities of mentors such as social contributions and step completions, have been anonymously collected throughout the course. In the *followings* dataset, the role of participants is specified. Since a mentor may wish to follow the MOOC as a learner while performing their mentoring duty, their data has not

been extracted from the dataset for the analysis.

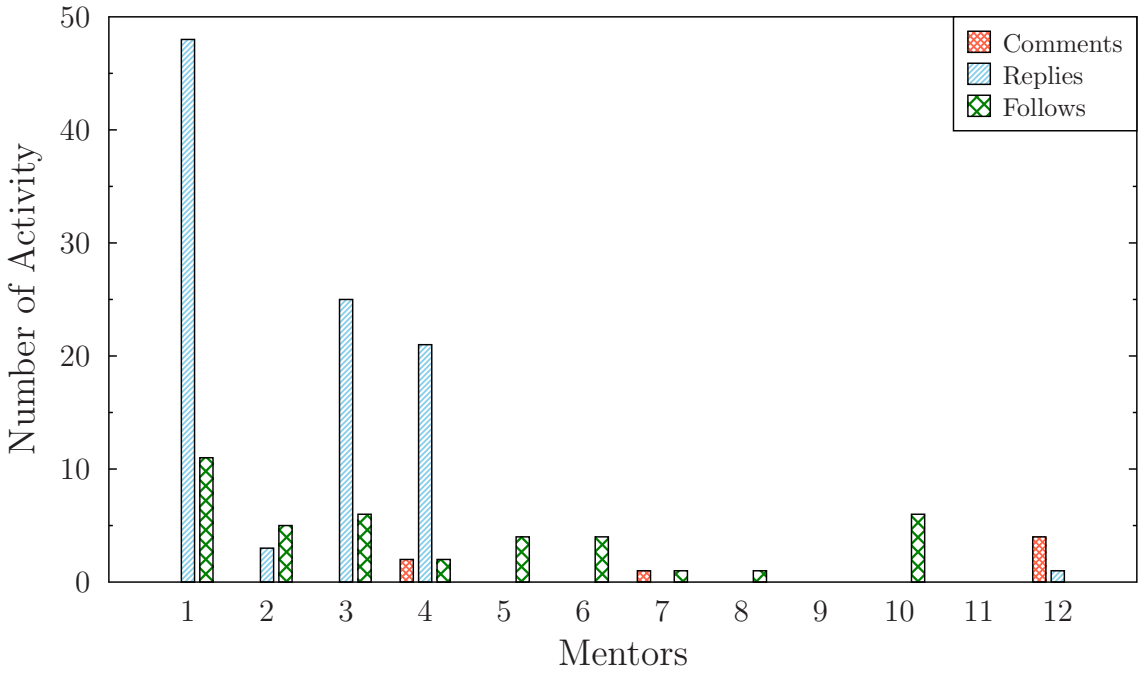


FIGURE 4.14: Volume of social activities of mentors: *comments*, *replies*, and *followings*.

However, it is worth looking at the mentors' engagement with the course. This is because a mentor who chooses not follow the course as a learner but only to mentor the discussions will show a high level of social attendance but no social completion. This kind of behavioural pattern may mislead the development of a prediction model.

In DYRP 2014 MOOC, 12 participants are identified as *educators* in the dataset, which means that they are members of the mentor team. Figure 4.14 shows the social activities of mentors. It is not surprising to see that mentors usually reply to comments rather than post an original comment. Only three mentors occasionally posted comments. It is also observed that most of the mentors follow others.

Figure 4.15 illustrates the course completion ratio of mentors. While a small number of mentors did not complete any of the steps, 8 (67%) mentors completed at least one course step and 4 (34%) of them actually completed more than half of the steps. It is observed that general behaviours of course completion and social attendance of mentors comply with the previously presented observations on the behaviours of all course participants. The mentors who posted comments, replied to many comments, and follow other participants are more likely to complete course: in that they have marked the course steps competed.

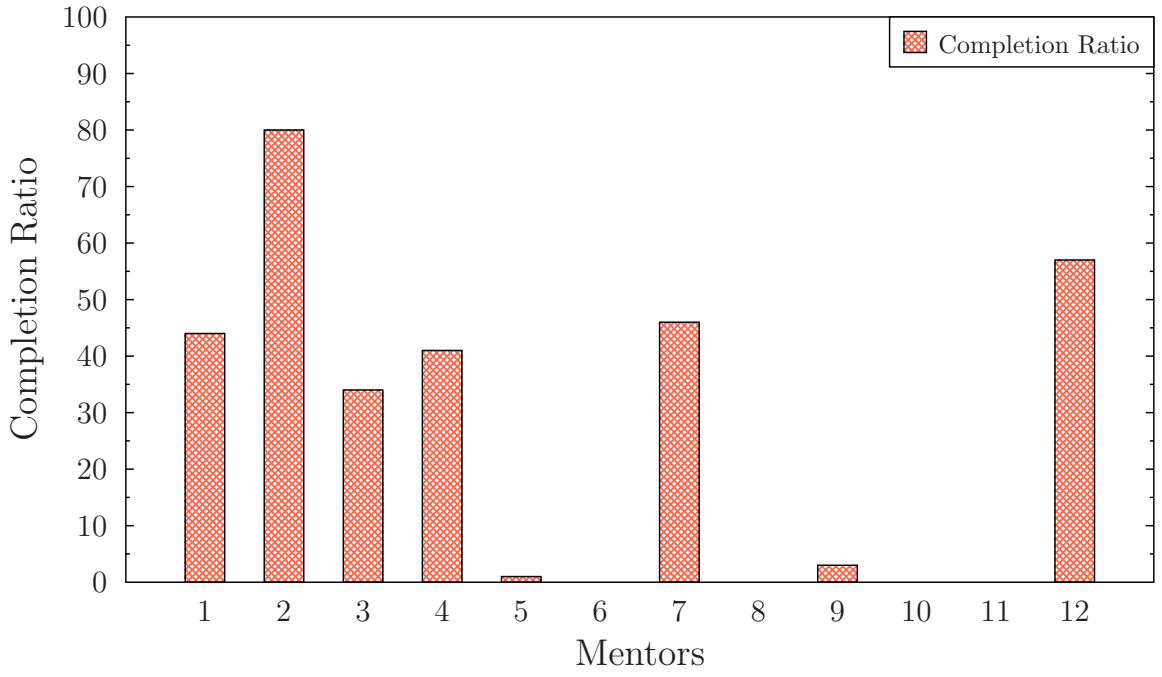


FIGURE 4.15: Course completion ratio of mentors.

4.3 Summary

This chapter analysed social behaviours of learners who exploited the social affordances that are provided by the FutureLearn MOOC platform. The findings of the analysis suggested that

- Participants' preferences vary in respect of using the social features (*follow* and discussion forums).
- The socially active learners more frequently completed the course where the threshold for course completion was set as completing more than half of the individual learning steps.
- Learners who did even a tiniest social contribution to discussions i.e posting a single comment, completed at least one step in the course even though the completion of a single step not necessarily extended to course completion.
- The completion rate of followers who did not contribute to discussions was lower than the completion rate of followers who did contribute to discussions.

Overall the research confirms previous findings that participation in course forums is a good indicator of committed participation in a course, and that learners who fully participate are the most likely to complete.

Furthermore, it is clearly the case that if a learner follows another learner they are

demonstrating that they are actively participating in the course, even if their participation does not extend to making original posts to the course forum. This finding implies that there is a strength in FutureLearn's *follow* opportunity which can be used for learning analytics to give insight about learners' behaviours. This finding can also be valuable for automated or manual facilitation of MOOC forums.

An original contribution of this work is to show that identifying such lurkers provides us with another useful parameter to feed into the model for predicting likeliness to complete. These findings imply a relation between completion and participants' social presence in the course. The next chapter presents a deeper investigation on the correlation between social behaviours and completion of the course.

Chapter 5

Behaviour Chains of Learners and Correlations to Course Completion

5.1 Introduction

In Chapter 4, a basic descriptive statistical analysis on how learners socially engaged in a FutureLearn MOOC was examined. This chapter investigates potential correlation of the level of participants' social engagement with course completion.

In order to answer the research question RQ4, which is *How can we typify the different patterns of participants' social behaviours during a course*, weekly social contributions of learners are investigated and categorised into a variety of interaction chains which typify discernibly different types of social interactions regarding:

- frequency of social attendance over weeks;
- frequency of interactions;
- type of social actions.

In addition, correlation between the categorised social behaviours and course completion is analysed, which address the research question RQ5: *What are the most correlated social behaviours to course completion in a MOOC*. A positive correlation between the categorised social behaviours and course completion may suggest that the level of social engagement could be used as an indicator in order to predict course completion.

Section 5.2 discusses the characteristics of social behaviours according to sequences

of the social actions that have been initiated. Section 5.3 defines behaviour chains of participants and analyses their correlation to course completion. Section 5.4 summarises the findings and concludes the chapter. The findings of this chapter are used to build a predictive model in Chapter 7.

5.2 Characterising Social Behaviours

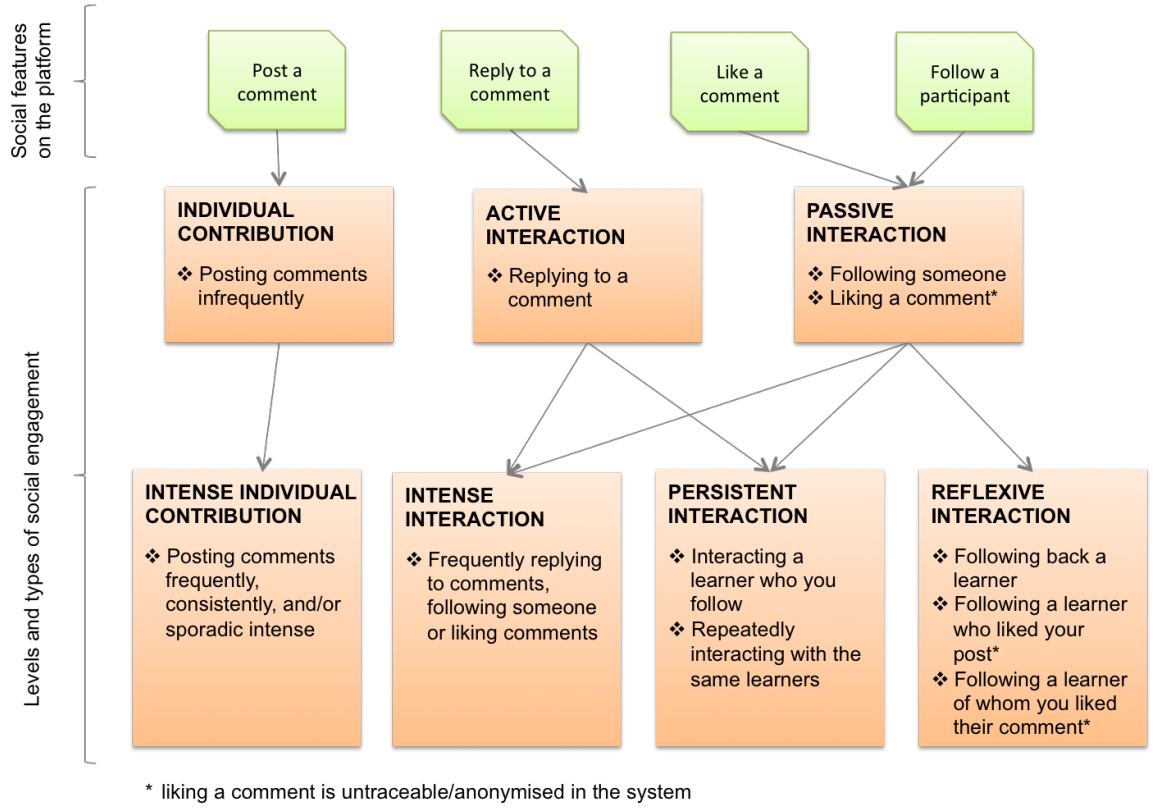


FIGURE 5.1: Categories of social actions and behaviours.

The FutureLearn platform affords three major social features: post a comment, reply, and follow. Different behaviours have been observed associated with the patterns of use of these three social features on the platform.

Figure 5.1 characterises the range of social actions and level of social engagement that may be observed in discussions. Course participants can make an *individual contribution* by simply posting comments. Additionally, they can initiate a *passive interaction* by following and liking comments. By this action, participants are effectively building their own learning network. Furthermore, they may contribute to a

comment appearing on the "most liked comments" list by liking. Participants can also extend *participatory interactions* by replying to comments.

Intense individual contribution and *intense interaction* imply that learners continue to frequently post, reply, and like comments and follow other participants.

They may also choose to follow back a learner, when i) they are followed, ii) their comments were liked, or iii) their comments received a reply. This kind of behaviour is defined as *reflexive interactions*.

The final social behaviour described in our research is *persistent interactions* that indicate repeated interactions between the same subset of learners.

An additional feature was added to FutureLearn in January 2016, which enabled learners to be notified when someone replied their comments. However, this feature was not available for the time of the datasets which this study presents. The additional feature may now encourage learners to have more reflexive and persistent interactions. A separate further investigation would need to be conducted to investigate this possible effect.

Learners may have initiated one or a sequence of social actions, which were modelled as behaviour chains. Alphanumeric codes and simple definition of these social actions are listed in Table 5.1.

TABLE 5.1: Definitions of actions observed in a-week-long period in FutureLearn.

Code	***Definition of action***
0	No social actions
C0	Contributing posts via individual comments to discussion threads
C1	Contributing posts via frequent individual comments to discussions
F0	Following a participant
F1	Following numbers of participants
F2	Following a participant after an interaction with that participant
R0	Replying to a comment
R1	Frequently replying to comments
R2	Replying to a comment which was posted by a learner that you follow
R3	Having a recurrent interaction with a learner with whom an interaction has already happened
R4	Having a recurrent interaction with a fellow learner that you follow

Figure 5.2 characterises possible links amongst different types of social actions. The

horizontal blue line (C0, F0, R0, C1, F1, and R1 regardless of order) indicates the three main social actions (*comment*, *reply*, and *follow*) that are observed. The longer the length of the line, the greater the variety of actions.

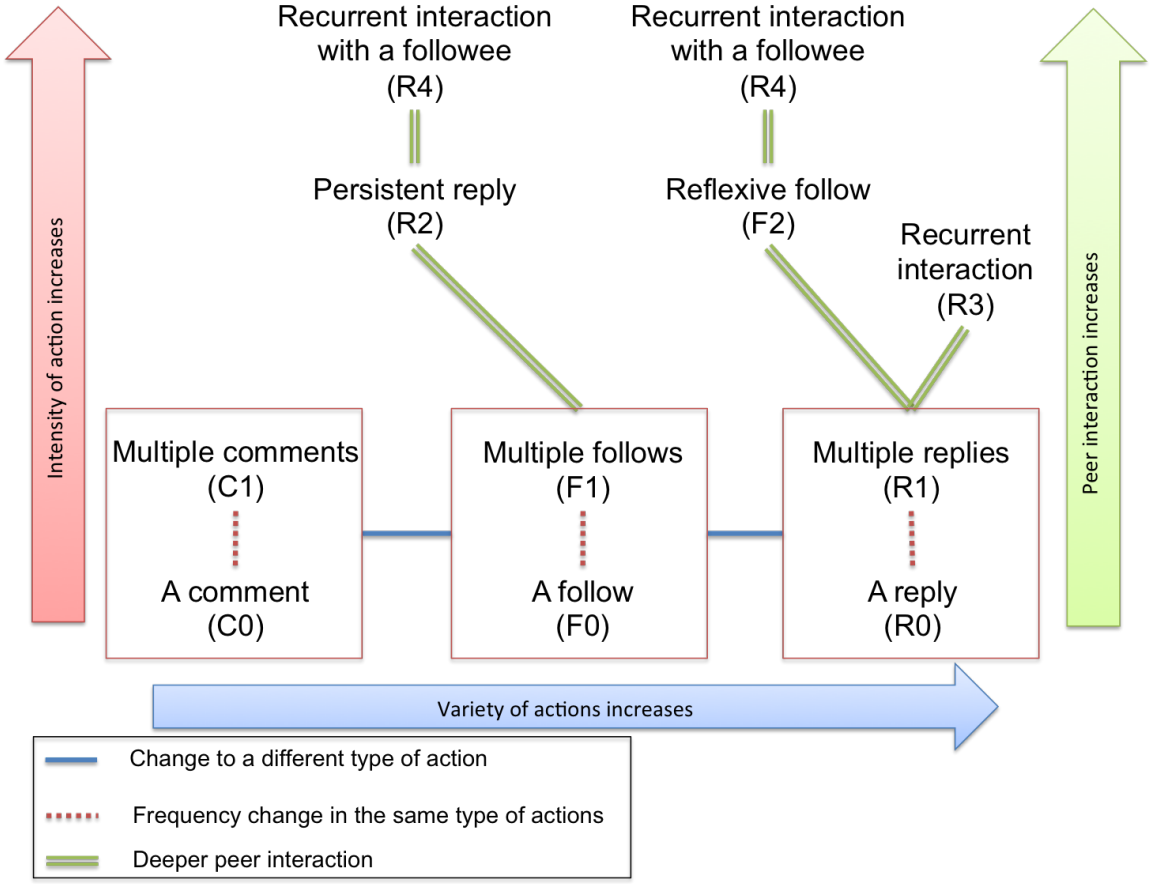


FIGURE 5.2: Extent and variety of participants' social actions.

Red dot dashed lines identify multiple actions such as frequently posting comments to discussions and following more than one person. It is represented by a dashed line since it indicates a change in the same type of actions. The line moves from a single action to a frequent action (e.g. C0 is a single comment; C1 is multiple comments). The straight lines rather indicate the change in type of social actions.

A vertical double green line identifies deeper peer interactions such as following a learner after an interaction with them or vice versa. If the length of vertical chain is longer, it indicates that participants engage in more interactive and persistent interactions with the same fellow learner.

5.3 Observed Behaviour Chains of Participants

Learners build behaviour chains every week that are of different length and consist of different types of social actions. A participant's social behaviour pattern can be categorised according to the behaviour *chains* that they built over the weeks. Table 5.2 shows the behaviour categories that are defined according to participants' behaviour chains over the weeks.

TABLE 5.2: Behaviour categories based on chains that are defined in this study

Simple
Learners who build chains consisting of infrequent individual contribution by writing a single comment (C0), participatory interaction by replying to a single comment (R0), and/or passive interaction by following one person (F0).
Moderately Frequent
Learners who build chains consisting of mainly simple social actions (one or all of C0, F0, R0) but very rarely write multiple comments, follow multiple number of people, and/or writing multiple number of relies to comments (C1, F1, R1).
Frequent
Learners who build chains consisting of intense individual contributions and interactions which are writing multiple comments, multiple replies to comments, and/or following more than one person (C1, F1, R1). Note that this classification does not differentiate how frequent the action is. So, any social actions that happened more than one time over the weeks are classified as frequent.
Persistent Frequent
Learners who build chains consisting of persistent and reflexive interactions alongside frequent contributions (F2, R2, R3, R4). So, if a learner had a reflexive interaction such as following a learner after having a conversation with that learner, this learner is categorised as persistent frequent. Having frequent or simple chains or none in other weeks does not change the category.

Figure 5.3 gives a hypothetical example of a learner's behaviour chains in the course. In the given example, the learner makes infrequent passive interaction by following a single person (F0), frequent individual contribution writing multiple comments (C1) in Week 1 and is socially inactive in Week 2. Then the learner makes infrequent participatory interaction by replying to a single comment (R0) and frequent participatory interaction by replying again to the same comment (R1) in Week 3 and replies to another comment (R0) in Week 4. Later in the course, the learner replies to a comment of a learner with whom had a conversation before (R3) in Week 5 and replies

to a comment of a fellow learner whom the learner follows in Week 7. According to the chains in each week, the overall behaviour pattern of the learner falls into the persistent frequent category at the end of the course.

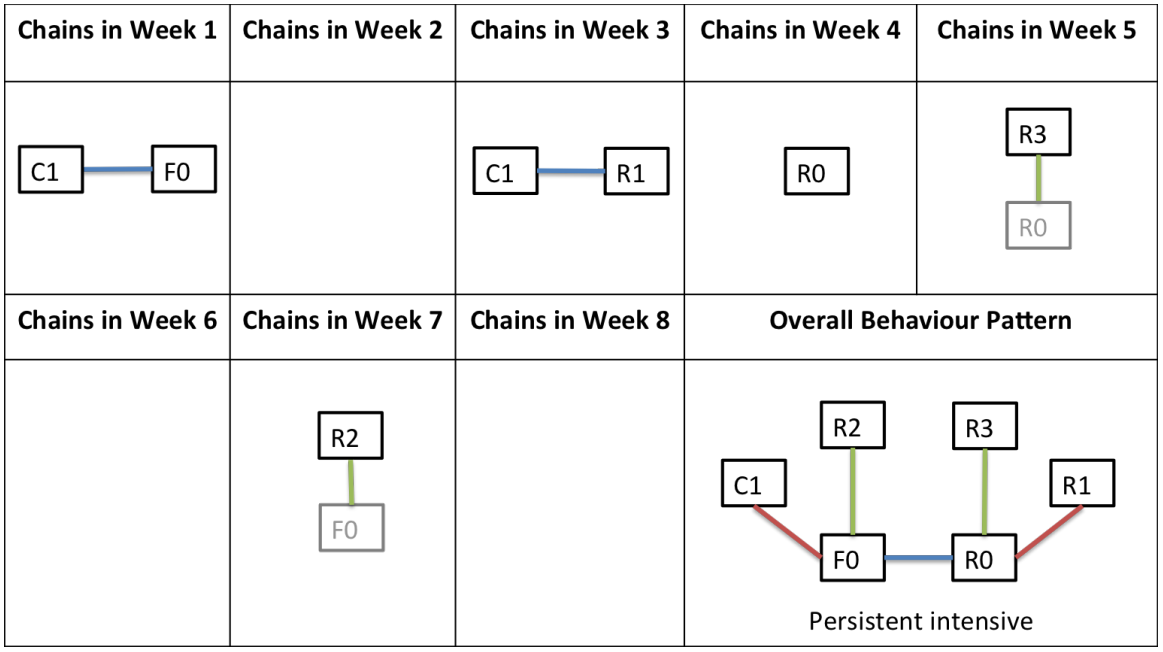


FIGURE 5.3: An example of weekly and overall chains of a learner.

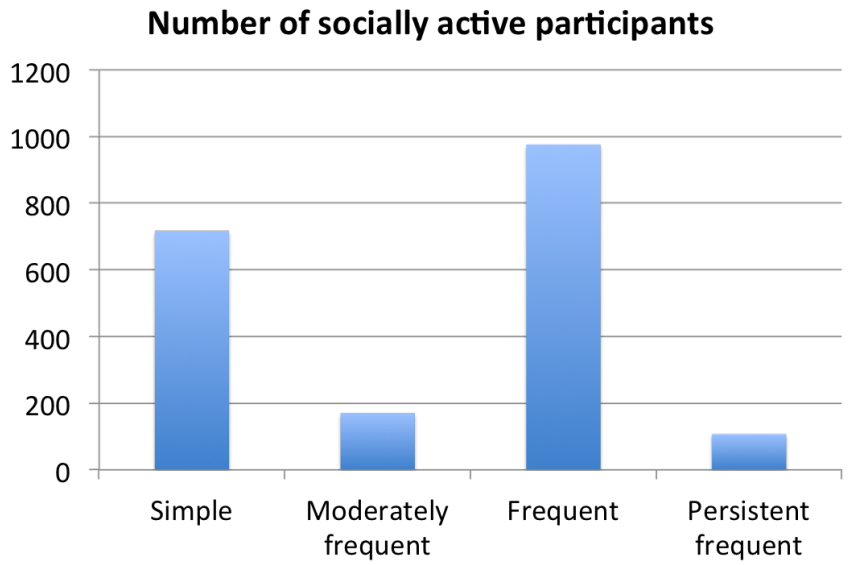


FIGURE 5.4: Number of participants in each group categorised by social chain types.

Category 1: *Simple:* Many learners tended to write a single comment or follow a single person over the weeks. Posting a reply to a discussion was almost never observed as a single behaviour in this category. This could be because overall attendances of

participants who engage in reflexive interactions are generally at a more active level. Additionally, horizontally tied chains such as a learner who posted a comment and followed a participant in a week-long period, are also observed. Figure 5.4 indicates that participants who are in the *simple behaviour* category have one of the largest populations amongst the socially active learners with 717 participants (36.3%).

Category 2: *Moderately frequent:* Learners who demonstrate moderately frequent interaction behaviours typically made individual contributions (C0), some participatory (R0), and/or passive (F0) interactions. This differs from those categorised in frequent who made frequent contributions and interactions (C1, R1, F1). It maybe reasonable to assume that participants in the moderately frequent category were more engaged in the course than the learners who built simple chains, yet they were not intensely engaged when compared with learners who were categorised as *frequent*. The number of participants in this category is 171 out of 1971 (8.7%) (Figure 5.4).

Category 3: *Frequent:* One single behaviour could be observed in a frequent and sporadically intense pattern (C1, F1, and R1). The behaviour of learners who frequently make intense individual contributions by posting multiple comments (C1) and/or intense interactions by posting multiple replies and following more than one person (R1, F1) were categorised to be in the *frequent chains*. A learner could build their chain by initiating any of these behaviours in a week. They do not have to show all of the three frequent behaviours. Of 1971 participants who were socially active, 976 (49.5%) built an frequent chain during their time in the course (Figure 5.4).

Category 4: *Persistent frequent:* Participants having recurrent interactions with a subset of their peers by replying repeatedly to comments of the same peer (R3), interacting with their fellow learners after they follow them (R2) or vice versa (F2), and having repeated interaction(s) with a followee whom the learner has already interacted with (R4) were examined in this category. A learner who is categorised as persistent frequent does not necessarily have to have persistent interactions (R2, R3 or R4) in every week. This type of behavioural chain was observed in participants who continuously interact with the social affordances whether in one single week or over a number of weeks. However, this was the least observed behaviour in the course. Only 107 (5.4%) participants extended their relationship with a fellow participant to a deeper state by repeatedly contacting them during the course (Figure 5.4).

In order to examine the completion rate in the context of the participants' social behaviours, the correlation between completion rate and participants' social behaviours is analysed. Before we proceed to the correlation analysis, the reader is reminded how

this research categorises course completion (Table 2.1 in Chapter 2).

1. *low completion:* Completion of the steps less than 50%.
2. *satisfactory completion:* Completion of the steps at least 50% of the steps but less than 80%.
3. *high completion:* Completion of more than 80% of the steps.

5.3.1 Correlation between Course Completion and Chain Types

The data examined shows that participants' performance in course completion varies according to the type of behavioural chains evidenced by their interactions. Figure 5.5 compares the distribution of learners' course completion represented by a boxplot graph.

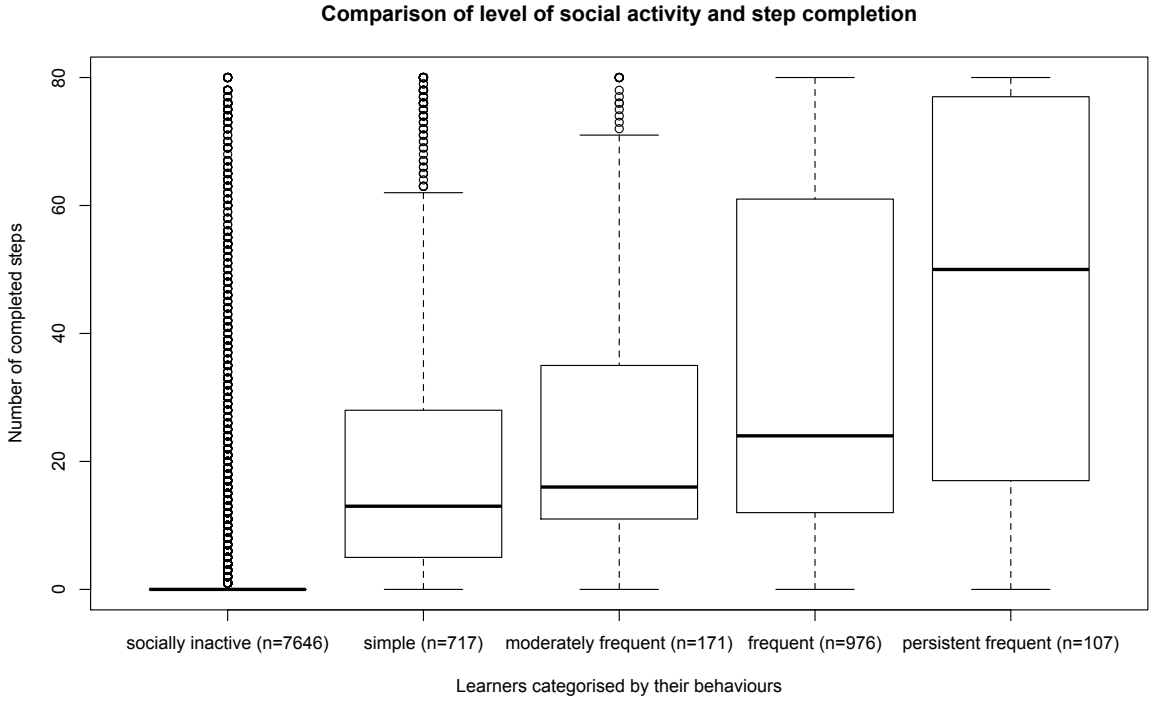


FIGURE 5.5: Boxplot for course completion of learners by their level of social engagement.

The x-axis on the graph represents the participants in each group, differentiating between those who i) were completely socially inactive during the course, ii) built simple chains, iii) tended to build frequent chains, iv) built frequent chains, and v) built persistent frequent chains.

The y-axis indicates the total number of steps that were completed by participants during the course.

The following points could be concluded from Figure 5.5.

- **Inactive participants:** Nearly 100% of participants who were socially inactive did not complete any of the steps. (Figure 4.2 in Chapter 4 demonstrates the same point.)
- **Simple and moderately frequent behaviours:** The participants who built simple (717) and moderately frequent (171) chains completed a relatively shorter range of number of steps. The completion of more than 75% of the learners in these groups remained low. Only a small number of participants who built simple and moderately frequent chains completed more than 80% of the course steps.
- **Frequent behaviour:** The participants who frequently (976) and persistently (107) contributed completed the largest number of steps. Even though their course completion is better than the others, the median value for the learners who built frequent chains (976) is still just slightly over 20 steps, which is not eligible for satisfactory or high completion status.
- **Persistent frequent behaviour:** The participants who built persistent frequent (107) chains (box on the far right) showed outstanding performance. A larger proportion of the learners completed more than half of the steps and the median value for course completion is near to 50 ($\sim$63% of the total steps). However, it is interesting to observe that a far smaller number of the course participants actually initiated persistent and reflexive interactions to build persistent frequent chains.

There are numbers of possible tools which this research can potentially use for measuring correlation. Pearson's correlation is one of the most commonly used statistical tests for measuring correlation. It determines the strength and direction of the **linear** relationship between two variables where the distribution of the population is normal. Since the distribution of our population is not normal, measuring the linear relationship between variables could be misleading.

However, using Spearman's correlation test to calculate correlation is more suitable than Pearson's correlation in this case. Spearman's correlation determines the strength and direction of the **monotonic** relationship between the two variables. In the case of comparing participant behaviour with completion rates in each case, it appeared that as the number of steps completed increases, so did the completion rate. Statistically this would appear to be a monotonic relationship.

Table 5.3 shows the results of correlation tests that have done by both Pearson and

Spearman's correlation tests.

TABLE 5.3: The result of correlation between chain type and course completion.

	Correlation coefficient	***p-value***	***Size of data***
Pearson's	0.52	< 0.001	9617
Spearman's	0.62	< 0.001	9617

These findings show that the course completion monotonically increases as the engagement of the learner gets deeper (from simple chain towards persistent frequent chain). The Spearman's test shows that course completion and behaviour chains are 62% monotonically correlated.

5.3.2 Correlation between Course Completion and Frequency of Social Actions

The previous section investigated the type of social behaviours. This section will further investigate the data of the behaviour chains to identify the frequency of interactions in order to determine whether the frequency of interactions is one of the factors that are associated with course completion.

Figure 5.6 illustrates frequency of social actions and completion status of learners who exploited the social affordances on the platform.

It was observed that learners who very frequently followed discussions and contributed to discussions appeared to have a satisfactory or high completion rate. Learners who made infrequent social contributions were identified in all of the groups, however, they were likely to perform better as the frequency of their contribution increased.

The boxplot graph in Figure 5.7 compares the frequency of social actions and participants' course completion. The majority of learners who completed less than 50% of the course did not show any social presence except for a few outliers (the box on the left).

The boxplot graphic clearly shows that the groups have different median values for the total number of completed steps. The majority of the learners in each group (it is almost 100% for learners in low completion), however, usually initiate 1 or 2 social actions.

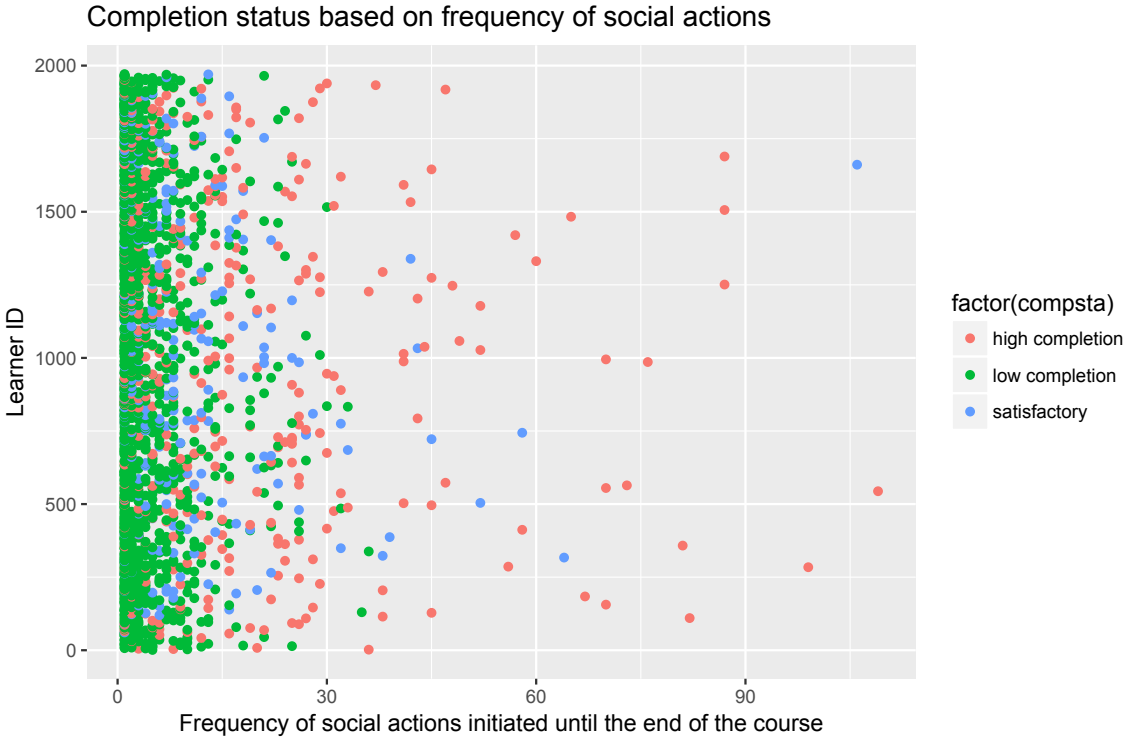


FIGURE 5.6: Frequency of social actions of learners grouped by completion status.

The graph in Figure 5.6 was examined and five threshold values for frequency of social actions were determined in this study as follows:

- **Inactive:** Learners who are socially inactive.
- **Very rare:** Learners who initiated at most 4 social actions.
- **Rare:** Learners who initiated more than 4 but less than 15 social actions.
- **Moderate:** Learners who initiated more than 15 but less than 25 social actions.
- **Frequent:** Learners who initiated more than 25 but less than 35 social actions.
- **High:** Learners who initiated more than 35 social actions. No upper limit.

Table 5.4 breaks down the frequency thresholds and shows the probability of course completion. The course completion and frequency of social actions are 58% positively correlated ($p < 0.001$, N=9617) according to the Pearson's correlation.

5.3.3 Correlation between Course Completion and Continuity to Contribution (Fullness of Chain)

This section will analyse whether or not continuous social participation has an impact on course completion. This part of the analysis is organised to investigate the

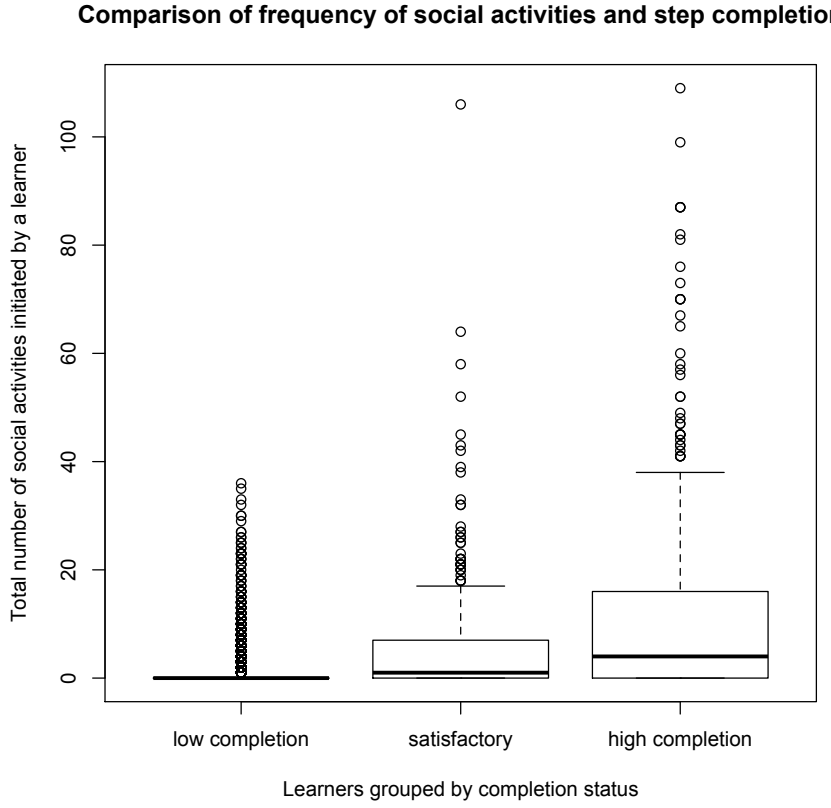


FIGURE 5.7: Boxplot for frequency of social actions of learners grouped by completion status.

TABLE 5.4: Probability of course completion according to the frequency of social actions.

Frequency of social actions	***Low completion***	***Satisfactory completion***	***High completion***
Inactive (0)	0.96	0.02	0.02
Very rare (∼4)	0.81	0.09	0.1
Rare (∼15)	0.61	0.18	0.21
Moderate (∼25)	0.34	0.23	0.43
Frequent (∼35)	0.18	0.15	0.67
High	0.02	0.19	0.79

questions:

1. Is there any difference in course completion of participants of those who contributed in only one week and those who contributed continuously for more than one week?
2. Does the level of engagement have an impact on the course completion of the

learners who continuously contributed?

3. Is there any difference in course completion rates of participants who made contributions in consecutive weeks?

Intuitively, the data shown in Figures 4.12 and 4.13 in Chapter 4 suggest that course completers are usually amongst those who are socially active over the weeks. However, the purpose of this investigation was to see if there was any statistical basis for this inference. Taking questions 1 and 2, socially active participants are clustered in three clusters as:

Socially active participants are clustered in three clusters as:

- **one-week-contributors:** those who make social contributions (regardless of whether it is simple, frequent, or persistent frequent) only in one particular week over the duration of the course.
- **continuous and socially less engaged contributors (continuous passive):** those who initiate simple and/or frequent interactions in more than one week.
- **continuous and socially more engaged contributors (continuous active):** those who initiate persistent frequent interactions in more than one week. They are not necessarily amongst those who engaged in reflexive and persistent interactions every week they contributed to discussions. However, learners are considered amongst those i) who reply to their fellows at least once, therefore presumably of those who use personalised tabs in the threads, ii) who follow someone after they interacted, and iii) who repeated these interactions at least once while they continued to make individual and passive contributions during the course.

Around 63% (1229) of the socially active learners only contributed in one week, which is predominantly in the first week in this MOOC. Figure 5.8 shows the distribution of the actions of this subset of participants over the duration of the course. This finding is consistent with the commonly observed attrition pattern in a typical MOOC that the majority of participants attend only the first week (Chapter 4). It is also because many MOOCs ask the participants to introduce themselves in the first week. Another reason could be that some learners who initially intended to take only a week that they were interested in, left after they completed that week.

The behaviour chain of these learners during a one-week social engagement is usually a single type of action by posting a comment or following a learner, or a mix behaviour of following someone, posting a number of comment and replying to a number of comment, which means that they prefer not to be involved in deeper peer interactions.

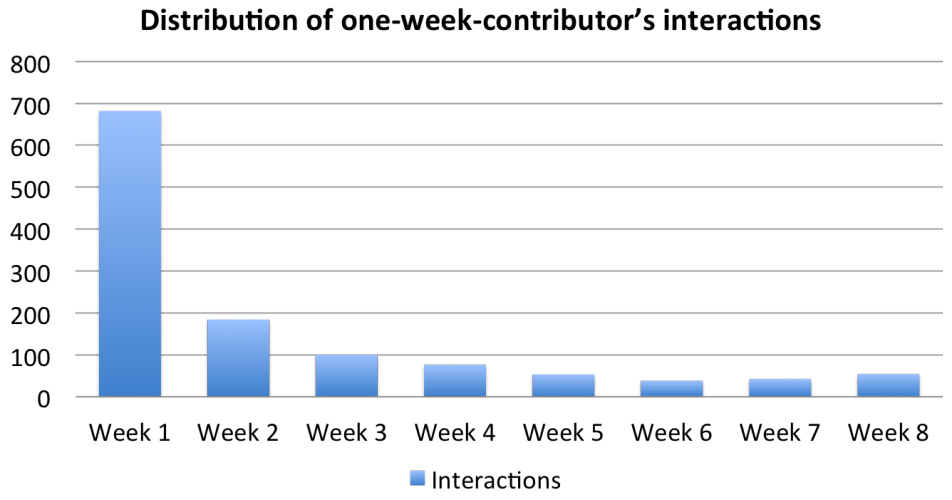


FIGURE 5.8: Distribution of one-week contributors' social actions over weeks.

Persistent peer interactions and recurrent interactions with a followee in discussions are very rarely observed.

The one-week contributors' step completion patterns were also consistent with their one-week contributor behaviours. Typically they completed steps in the very same week with that they were socially active in. Much smaller numbers of exceptional instances show that they completed steps in more than one week. However, their completion and attendance remained low.

While the continuous and socially less engaged learners fall into around 23% (455) of the socially active participants, the continuous and socially more engaged learners are the 14% (272) of socially active participants. Learners who contributed to discussions in more than one week, regardless of the level of their engagement, performed better in completing the course.

Figure 5.9 compares the overall step completion of the participants from different clusters by a boxplot illustration. The range and the median value for the *one-week contributors* are significantly smaller than the other two clusters. However, there is no significant difference between participants whose interactions were i) continuous and socially less engaged, and ii) continuous and more engaged contributors. This indicates that the participants in these two clusters behaved in a similar manner in completing the course.

The data shows that the minimum number of completed steps is 0 for *one-week contributors*; it is 1 for the other two groups. This finding indicates that there were learners who were socially active in a-week-long period but actually never completed a step.

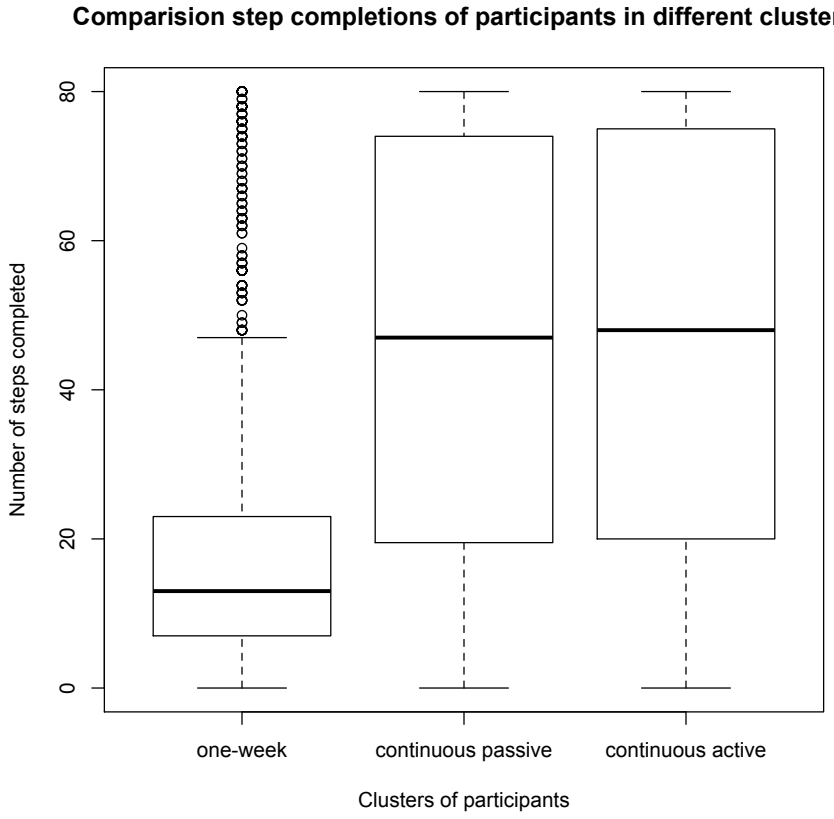


FIGURE 5.9: Comparing the step completions of participants from different clusters.

Small circles out of the range represent the exceptionally higher number of completed steps for each learner (outlier participants). In summary, it is observed that there is a difference between course completion performances of one-week contributors and continuous contributors whereas no significant difference amongst passive and active continuous contributors. This section now continues with the third question which requires an analysis on correlation of course completion to continuous contributions in consecutive weeks.

In order to accomplish the analysis, the following categories are defined to identify completeness of chains in Table 5.5.

The correlation between the completion of the course and completeness of the chain is 0.55 ($p < 0.001$, N=9617) according to Pearson's correlation test. This correlation coefficient value indicates that there is a positive linear relation between sustained contribution over the weeks and course completion. This result is consistent with the findings implied from the data shown in Figure 4.12 in Chapter 4.

Table 5.5: Completeness of chains

No chain
This indicates participants who were socially inactive during the course. Consequently, they had no chain.
One-length-chain
One-length-chain indicates the participants who were *one-week-contributors*. Since they were socially active only in a particular week, they have on-length chain.
Broken chain
Broken chain indicates that a participant made non-sequential social contributions. For example, a participant who was socially active in Week 1 and Week 4 has a broken chain since Week 1 and Week 4 are not sequential weeks.
Unbroken chain
Unbroken chain represents social actions that were initiated in any consecutive weeks. An *unbroken* chain is not necessarily started in Week 1 and ended in Week 8. Any consecutive weekly contributions are considered as *unbroken*, e.g. contributions in Weeks 1, 2, and 3, or in Week 6 and Week 7.

Table 5.6: Probability of course completion according to the fullness of chain.

Fullness of chain	***Low completion***	***Satisfactory completion***	***High completion***
No chain	0.95	0.02	0.03
One-length-chain	0.85	0.07	0.08
Broken chain	0.49	0.24	0.27
Unbroken chain	0.01	0.11	0.88

5.4 Summary

Participants' patterns of social engagement vary over weeks. Although participants may make use of a social affordance, a few participants consistently continue to use the social affordances throughout the duration of the course. Participants may choose to engage in frequent/intense individual contributions, passive/intense interactions, and persistent reflexive interactions (Figure 5.1). Each behaviour type has been modelled as a chain in order to further examine the data.

A statistical analysis of the data suggests that a participant's course completion is strongly correlated to i) type of behaviour chain (62%), ii) frequency of social actions in a chain (58%), and iii) completeness of chain (55%).

These results indicate that high levels of social interactivity could be a good predictor for course completion. The positive correlation between social interactions and course completion is an affirmative answer to the second research question: *Is showing social presence in a MOOC correlated to the participants performance in course completion?*. The findings shown so far also support the hypothesis, which is that *the data extracted from participants' engagement in a MOOC can be used to identify social behaviour patterns of participants and this information can contribute to a model of course completion.*

The findings presented in this chapter show that the participants who completed the course were most likely to be amongst those who were also socially active. However at this point, this research is making no claim that being social causes participants to complete the course. There might be a causal relationship between social interactions and course completion, however, this would need to be investigated by further research investigation.

However, the strength of these findings were taken to be sufficient to provide a basis to build a prediction model for course completion. Chapter 7 uses these features to build a predictive model to anticipate participants' course completions. Moreover, Chapter 6 provides a comprehensive literature review on the use of prediction models in MOOCs.

Chapter 6

Use of Prediction Models in MOOCs

6.1 Introduction

One way in which the observed behaviours of learners in MOOCs discussed in Chapter 4 and 5 might be used more generally as part of a prediction model. For example, learners' social behaviours and completion rates could be used as parameters to predict whether or not a learner will complete the course at the end.

Predictive models are commonly used for making decisions by forecasting outcomes with the aid of statistical models and machine learning (Finlay, 2014). Predictive models are applied in various contexts ranging from health and politics to business and education.

In MOOCs, predicting future participations and dropouts could be useful for detecting the need for educational interventions and the appropriate timing of such interventions. A number of researchers have attempted to apply predictive models. Focus includes anticipating learners behaviours and identifying learners at risk.

In this chapter, Section 6.2 presents available state-of-the-art techniques considering their objectives, prediction methods, the dropout definitions identified in the literature and their notable findings. Section 6.3 summarises and concludes the findings. Furthermore, Chapter 7 proposes our approach to build a prediction model and Chapter 8 discusses the promises of our method.

6.2 Critical Analysis on Prediction Models

Since there is no formal dropout definition (Evans and Baker, 2016), each study implements their own experiments using a variety of definitions. Two most widely-used definitions for dropout are:

1. **Not completed the final week:** If a learner does not engage in the final week's activities, they are assumed to have dropped out of the course. A similar assumption, proposed by some researchers, is that learners are marked as dropped out if they did not submit the final assignments.
2. **No activity during the most recent week** or **No further activities in the following weeks:** This definition differs from the previous in terms of the timing of the dropout. For example, if a learner's last activity is recorded in the fourth week of a six-week-long MOOC, that student is marked as "dropped out in the 4th week".

Several kinds of data are used and collected throughout the duration of a MOOC in order to observe learners' behaviour and develop prediction models (Chapter 3). Typically four types of dataset are available: i) pre- and post- course surveys, ii) clickstream, iii) the results of assignments, and iv) activities in discussion forums. Some researchers use only clickstream data i.e. Amnueypornsakul et al. (2014) and Kloft et al. (2014), others combine the use of clickstream data, assignments and forum data. The studies were examined and selected to identify the strongest indicators that would have the most impact on prediction of dropouts. The following factors are the strong points summarised from the literature. Tables 6.1, 6.2, 6.3, 6.4, and 6.5 give wider and detailed samples from the literature.

- Learners who show even minimal interaction in the forum after Week 1 are unlikely to drop out (Balakrishnan, 2013).
- Learners who start a course earlier and contribute to discussions are less likely dropout than others (Yang et al., 2013).
- Learners who lost their close peers are less likely to continue participating in the course forum (Yang et al., 2014b).
- Learners who join later and participate in the least number of activities drop out (Sinha et al., 2014).
- Assignment submissions are the most predictive (Taylor et al., 2014).
- The length of forum posts is more strongly predictive than the number of posts and responses (Taylor et al., 2014).

- Social integration in Week 1 is strongly correlated with course completion (Jiang et al., 2014b).
- Attrition rates and learners sentiment towards assignments and course materials are correlated (Chaplot et al., 2015).
- Learners' self-statements about their intention are more strongly predictive than demographics (Robinson et al., 2016).
- Choice of approach for training model directly effects the accuracy results. Training on the same course gives overly optimistic accuracy results. In addition, classifier performance is not significantly different in different academic fields (Whitehill et al., 2017).

As presented in Tables 6.1, 6.2, 6.3, 6.4, and 6.5, many studies consider learners' social participation as a factor for predicting dropouts i.e. Yang et al. (2013, 2014b); Taylor et al. (2014); Jiang et al. (2014b); Chaplot et al. (2015). For example, Chaplot et al. (2015) use learners' sentiments extracted from the posts in the forum, while Jiang et al. (2014b) takes into consideration learners' level of activity in the forum in the first week.

This research defines a *dropout* as a learner who has no further activities in subsequent weeks. In addition, completion of less than 50% of the steps is defined as *low completion*. This research aims to predict learners who are going to complete less than 50% of the steps based on social activities. Therefore, this research uses certain parameters of social activities as an indicator in a prediction model.

However, the design of the platform has influence on the selection of parameters. For example, some researchers use opening a new forum topic as an indicator where the design of the discussion platforms are similar to traditional forums. The design of FutureLearn offers a Twitter-like discussion boards where the comments posted by participants flow, which is different from the commonly-used style as discussed in Chapter 2. Therefore, posting an original comment and replying to a comment are considered as parameters of social activities instead of opening a new forum topic. Also, *follow* feature is used, which is very unique among the presented literature since this social function is not afforded by many MOOC platforms.

6.3 Summary

This investigation on how prediction models were used in MOOCs indicated that researchers have been building predictive models to anticipate drop out rates and

to identify the learners at risk so that timely interventions could be possible. The findings suggested that researchers have used different methods for implementation of the prediction models such as logistic regression, random forest, natural language processing and probabilistic models. To implement these models, numbers of sources that generates information about learners' activities are used. For instance, click-stream data, assessment, and forum activities are commonly used. The next chapter will demonstrate the prediction model developed in this research to predict attrition, which uses the data that were generated from the social activities of participants on a FutureLearn MOOC.

TABLE 6.1: State-of-the-art techniques for predicting learners' participation in MOOCs.

Study	Focus	Prediction Model	Datasets
Balakrishnan (2013)	1) Predicting course attrition, 2) Patterns of learners' behaviours	Hidden Markov Model	Clickstream, Assessments, Forum activity
Yang et al. (2013)	Social factors on dropouts	Survival Model	Forum activity
Yang et al. (2014b)	Peer influence on learners' retention	Survival Model	Forum activity
Sinha et al. (2014)	1) Learners' activities' patterns, 2) Predicting course attrition	Baseline Ngram Model, Graph Model	Clickstream, Forum activity
Taylor et al. (2014)	Predicting course attrition	Logistic Regression	Clickstream, Assessments, Forum activity, Wiki revisions
Halawa et al. (2014)	Students at risk of dropout	Least Mean Square (LMS)	Clickstream Assessments
Ramesh et al. (2014)	Predicting learners' survival	Probabilistic Soft Logic	Clickstream, Assessments, Forum activity
Jiang et al. (2014b)	Predicting earning certificates	Logistic Regression	Assessments, Forum activity
Amnueypornsakul et al. (2014)	Predicting learners' retention in a week	Support Vector Machine	Clickstream
Sharkey and Sanders (2014)	Predicting course attrition	Random Forest Model	Clickstream
Kloft et al. (2014)	Predicting course attrition	Fisher Scoring, Support Vector Machine	Clickstream
Chaplot et al. (2015)	Predicting course attrition	Artificial Neural Network	Clickstream, Forum activity
Mi and Yeung (2015)	Predicting course attrition	Recurrent Neural Network	Clickstream, Forum activity
He et al. (2015)	Students at risk of dropout	Logistic regression	Clickstream, Assessments

TABLE 6.2: State-of-the-art techniques for predicting learners' participation in MOOCs (continued to Table 6.1).

Study	Focus	Prediction Model	Datasets
Robinson et al. (2016)	1) Predicting learners' success before a course starts, 2) Intention to earn certificate	Natural Language Processing	Pre-course self-assessment
Li et al. (2016)	Predicting course attrition	Multi-view Semi-supervised Learning	Clickstream
Liang et al. (2016)	Predicting course attrition	Gradient Boosting Decision Tree	Clickstream
Whitehill et al. (2017)	Predicting course attrition	Deep Neural Network	Course surveys, Clickstream
Bote-Lorenzo and Gómez-Sánchez (2017)	Predicting course attrition	Stochastic Gradient Descent	Clickstream, Assessments
Hlosta et al. (2017)	Students at risk of dropout	Tree Boosting XGBoost	Clickstream, Assessments, Forum activity

TABLE 6.3: Milestones of dropout definitions used and remarkable findings of these studies.

Study	Dropout Definition	Findings
Balakrishnan (2013)	No activity in the most recent week	Learners who rarely/never check their progress page leave the course earlier. Those who show even minimal interaction in the forum after the first week are unlikely to dropout early.
Yang et al. (2013)	Not completed the final week	Learners who start the course earlier and contribute to discussions are less likely to dropout the course than others who do not.
Yang et al. (2014b)	No activity in the most recent week	Learners who lost their close peers are not likely to continue participating in discussion forums.
Sinha et al. (2014)	No activity in the most recent week	Recency and frequency of learners' activities would be used to predict learners' pathway. Learners who join courses later and do not participate in many activities usually leave courses.
Taylor et al. (2014)	No further assignment or assignment submission	The most recent four weeks are predictive. Submitting assignments is the most predictive. The length of posts is more predictive than the number of posts and responses in the forum.
Halawa et al. (2014)	1. Absence for a period exceeding one month 2. View fewer than 50% of videos	Dropout is strongly related to one type of bad persistence pattern i.e. learners who are absent 14 days or more are red-flagged.
Ramesh et al. (2014)	No activity in the most recent week	The middle phase of a course is the most important phase to monitor students' activity for prediction of dropout.
Jiang et al. (2014b)	Not completed the final week	Social integration with a learning community in Week 1 is strongly correlated to completion.

TABLE 6.4: Milestones of dropout definitions used and remarkable findings of these studies (continued to Table 6.3).

Study	Dropout Definition	Findings
Amnueypornsakul et al. (2014)	No activity in the most recent	Features related to quiz attempts and submissions are reasonable predictors in a given week.
Sharkey and Sanders (2014)	No activity in the most recent week	Extracted 15 different data features related to learners' engagement and activity are strong predictors for dropout.
Kloft et al. (2014)	No activity in the most recent week	Predictions are better measured at the end of a course.
Chaplot et al. (2015)	No activity in the most recent week	There is a correlation between attrition and attitude of learners towards course materials and assignments.
Mi and Yeung (2015)	1. No activity in the final week 2. No activity during the most recent week 3. No activity in the coming week	Prediction of dropout is a sequence classification problem. LSTM outperformed all other methods tested.
He et al. (2015)	No activity in the most recent week	Early alerts to identify students at risk of not completing is important for interventions.
Robinson et al. (2016)	Not completed the final week	Learners' self-assessment is a better predictor than demographics.
Li et al. (2016)	Absence for more than 10 days	Separately training the system for each type of behaviour achieves better prediction accuracy.
Liang et al. (2016)	No activity in the most recent week	Individual's engagement and total engagement in the course can be used for prediction.
Whitehill et al. (2017)	No activity in the most recent week	Data training approach is very crucial for developing a accurate model. Training on the same course gives overly optimistic accuracy results.

TABLE 6.5: Milestones of dropout definitions used and remarkable findings of these studies (continued to Table 6.4).

Study	Dropout Definition	Findings
Bote-Lorenzo and Gómez-Sánchez (2017)	No activity in the most recent week	Watching lectures, solving finger exercises, and submitting assignments are good predictors for predicting the decrease of students' engagement.
Hlosta et al. (2017)	No activity in the most recent week	6 days before the any deadline might be suitable for applying interventions.

Chapter 7

A Novel Approach for Predicting Learners' Future Participation

7.1 Introduction

Researchers have combined different types of datasets, such as click-stream data, course surveys, assignment performances, and discussion forum activities, to have greater insight into learners' behaviours and success (Shahiri et al., 2015; Dutt et al., 2015). They have analysed the relationships between learners' behaviours in MOOCs and their course completion rates to:

- identify possible reasons for low retention rates (Khalil and Ebner, 2014);
- provide necessary help to learners (Sunar et al., 2015c);
- predict learners' future behaviours before they happen (Shahiri et al., 2015).

Since different MOOC platforms take different pedagogical approaches and offer distinctive technological affordances, researchers use a range of parameters from MOOC learners' online behaviours to predict their performance. The FutureLearn MOOC platform, which was used in this study, takes a social-constructivist approach for designing MOOCs as explained in Chapter 2.

Chapter 4 already presented the analysis on social engagement of FutureLearn MOOC participants to have a greater insight into social behaviours of learners in their learning networks, Chapter 5 presented the novel idea of presenting participants' engagement as behaviour chains according to the type of patterns of social actions. It is observed that pattern of social engagement is correlated to success in course completion. To the

best of our knowledge, no researchers are currently using learners' forum interactions combined with their *follow* behaviours to predict possible dropouts, which is widely discussed in Chapter 5.

This chapter is designed to answer the research question **RQ6**: Can we use these correlated behaviours in order to predict participants' course completion?. Prediction models are presented to predict course attrition by using participants' interactions with social affordances on FutureLearn. Section 7.2 presents the feature set which has been used for predictions. Section 7.3 describes the selected classifiers that the model has been tested on. Section 7.4 presents the implementation of the model and compares the results. Section 7.5 discusses the advantages and the disadvantages of the model. In the end, Section 7.6 summarises the findings and concludes the chapter.

7.2 Feature Set Selection

A feature set is a subset of the features in the dataset and extracted features from the dataset, e.g. the *behaviour chain type* feature was extracted from the data in this study. Feature subset selection can have a positive affect on the performance of machine learning algorithms by enhancing the performance of learning algorithms, reducing the hypothesis search space, and, sometimes reducing the storage requirement (Hall and Smith, 1997). One of the methods to select the feature set is *correlation-based feature selection* which filters the most correlated features.

Table 7.1: Attributes in the selected feature set for the construction of the prediction model.

behaviour chain type	**frequency of social actions**	**completeness of chain**	**course completion status**
inactive, simple, likely intensive, intensive, persistent intensive	An integer value	zero, one-length-chain, broken, unbroken	low, satisfactory, high

Chapter 5 explained *behaviour chains* that were extracted from the data of which FutureLearn generated from the participants' online activities during the course. It is also shown in Chapter 5 that some features extracted from the data such as frequency

of social actions and continuity to contributions are correlated to the course completion. Table 7.1 shows the feature set selected for use in prediction model construction.

7.3 Machine Learning for Classification

The prediction model in this research was developed to predict which class a participant would fall into at the end of the course i.e. *low completion, satisfactory completion, high completion*. There are numbers of different classifier algorithms that have been used in MOOC research as revised in Chapter 6.

In order to train and test the model, Random Forest Model and Support Vector Machine algorithms were selected. Random Forest Model was chosen since it is very simple to implement and interpret. Support Vector Machine was chosen because of its better performance with high-dimension feature sets like the dataset that was used in this research.

7.3.1 Random Forest Model

A *Decision Tree* is an algorithm used to make decisions using a tree-like model. A tree is split into branches on an attribute value. According to the path from the root to the leaf, the decision tree predicts the class of the input value.

The random forest classifier consists of a collection of decision tree classifiers where each classifier is randomly generated using a random subset of input variables, and each tree casts a unit vote for the most popular class to classify an input variable (Breiman, 2001).

Pal (2005) identifies some of the advantages of the random forest classifier as follows:

- The random forest classifier requires two parameters only to be set. Whereas a number of userdefined parameters are required for support vector machines, which are a type of classifier.
- It provides the relative importance of different features during the classification process, which can be useful in feature selection.

7.3.2 Support Vector Machines

Support Vector Machine is a discriminative classifier which uses numbers of hyperplanes and finds an optimal hyperplane to categorise samples in training data (Cortes and Vapnik, 1995). This algorithm is especially good at working with high-dimension feature space.

7.4 Implementations of Prediction Models

7.4.1 Imbalanced Data Problem

The data that has been used in this study was imbalanced. The number of people who did not complete the course outnumbered by a very large percentage those who did complete. In this case, classifiers tend to be overwhelmed by the large classes and ignore the small ones (Chawla et al., 2004). One possible solution for such imbalanced data problem is re-sampling the data by taking the same number of records from all classes.

Table 7.2 shows the number of people in each class categorised by the course completion status. Since the lowest population is 377 (in satisfactory completion), 377 participants were randomly selected from each category. Thus, the sample data for implementation includes 1131 randomly selected participants.

TABLE 7.2: The number of people in each class of course completion.

	socially active	**socially inactive**	**total**
Low completion	1372	7312	=8684
Satisfactory comp.	222	155	=*377*
High completion	399	179	=578

7.4.2 Training Data: k-fold cross-validation

In order to split the training and testing data, the *hold-out* method is commonly used, which divides the data into two at a certain rate e.g. 70% for training; 30% for testing. However, the *hold-out* method does not use the records in the testing set for training since this method relies on a single split of data (Arlot et al., 2010).

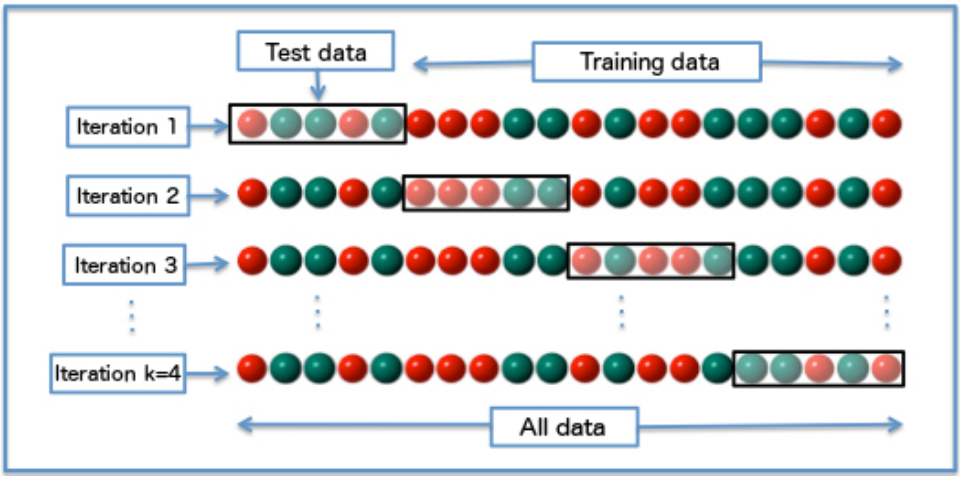


FIGURE 7.1: Illustration of k-fold cross-validation with k=4 (source: Wikipedia).

To deal with this limitation, researchers prefer to use *k-fold cross-validation* which splits data to k and iteratively uses different partitions of the data as a training and testing set. The *k-fold cross-validation* method gives a more accurate estimate of model prediction performance. To depict the model, 7.1 illustrates the data partition in a 4-fold cross validation.

In this research, 10-fold cross validation method is used for data training.

7.4.3 The Workflow of Implementation

This section briefly explains how the implementation process has been carried on in order to complete the implementation of the presented model. Table 7.3 presents the stages in the implementation process and the aim of each stage.

7.4.4 Results

7.4.4.1 Testing with Random Forest Model

In order to implement the prediction model, the instant package *randomForest* of the *R* statistical analysis tool was used.

Table 7.4 shows the performance of the implemented model with Random Forest. The confusion matrix shows the classification of the samples and error rates for each category.

Table 7.3: Workflow of the implementation of machine learning algorithms which have been chosen.

Stage I: Implementation with RFM	**Aim of the stage**
A Random Forest Model was trained with the chosen feature set (*behaviour chain, frequency, fullness*) and tested on the DYRP 2014 MOOC	to see the performance of Random Forest Model
A Random Forest Model was trained with the raw attributes (*the numbers of comments, replies, follows*) and tested on the DYRP 2014 MOOC	to see if the features extracted from the modelled behaviours were a better predictor than the raw attributes in the data
A Random Forest Model was trained with the raw attributes and the chosen features together and tested on the DYRP 2014 MOOC	to see if the performance of the model has changed
Stage II: Implementation with SVM	**Aim of the stage**
A Support Vector Machine was trained with the chosen feature set and tested on the DYRP 2014 MOOC	to compare the performance of different classifiers on the prediction
Stage III: Testing on a data from a different MOOC	**Aim of the stage**
The trained model on the DYRP 2014 MOOC was tested on the DYRP 2016 MOOC	to see if the model is compatible with other MOOCs

For example, the model correctly predicted 270 of the samples who are in *low completion*, 61 of them were mispredicted as the class *satisfactory completion*, and 1 of them were mispredicted as the class *high completion*. Therefore, the error rate for the prediction of the class *low completion* is 0.18, in other words, 82% of the samples in the *low completion* class were correctly predicted. In the confusion matrix, the diagonal entries of the matrix show the correctly predicted samples in the test dataset.

According to the results, the overall error rate of the model is 45% which means that the model 55% correctly predicted the completion status of participants. The model especially correctly predicted participants in the low completion categories (82%, error rate: 18%), however, mispredicted most of the samples who completed the course at a high rate (37%, error rate: 73%).

TABLE 7.4: Confusion matrix of the implemented Random Forest model.

	Low completion	**Satisfactory completion**	**High completion**	**Class error**
Low completion	270	61	1	0.1867470
Satisfactory completion	161	193	17	0.4797844
High completion	88	126	79	0.7303754

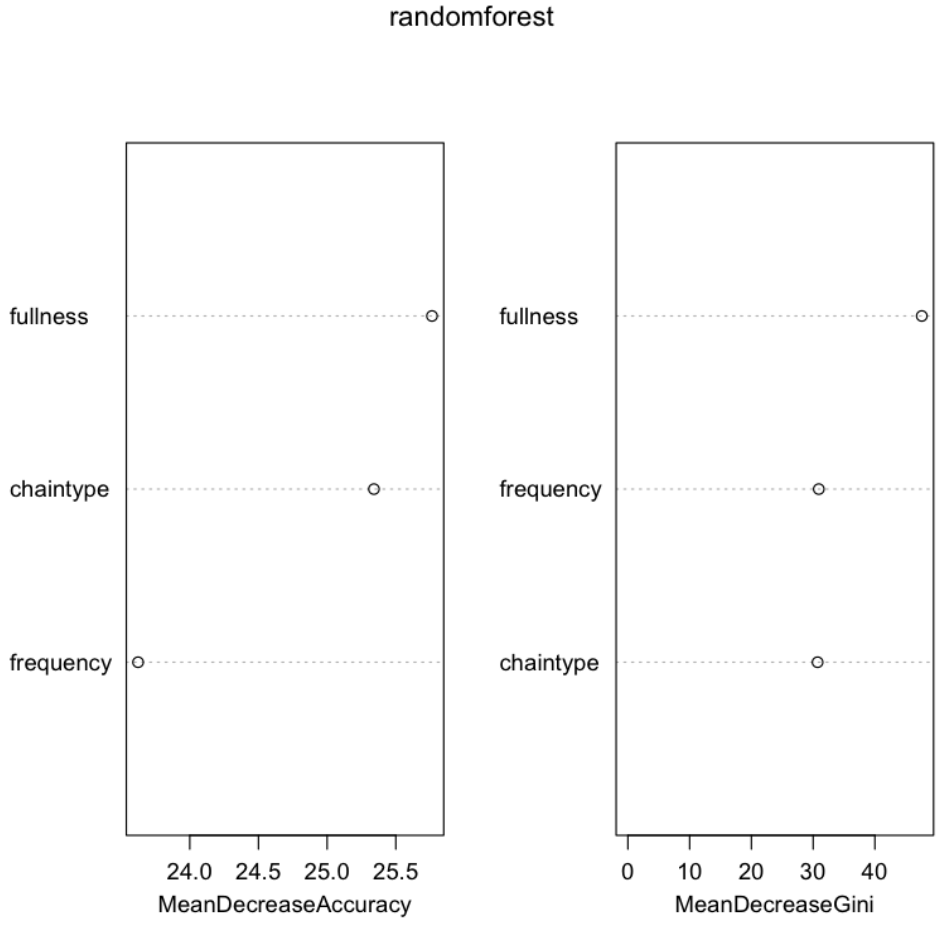


FIGURE 7.2: Mean accuracy of the implemented Random Forest model.

Figure 7.2 shows the impact of variables in the feature set on the model. The graph on the left in Figure 7.2 shows the mean decrease in accuracy of each variable which indicates the importance of the variable for classification of the data. In other words, mean decrease accuracy shows how much the accuracy would decrease by removing the associated feature. According to the results, *fullness* and *chaintype* are more

important than *frequency* for the classification.

The graph on the right in Figure 7.2 shows variable importance based on the Gini impurity index used for the calculation of splits during training. It shows the importance of each variable in a split, but not in the whole tree. A strong Gini importance of a variable does not always mean that it is an important variable for the classification.

TABLE 7.5: Confusion matrix of the implemented Random Forest model applied to the feature set containing raw comment, reply, follow attributes.

	Low completion	**Satisfactory completion**	**High completion**	**Class error**
Low completion	13	250	69	0.9608434
Satisfactory completion	44	304	23	0.1805930
High completion	84	194	15	0.9488055

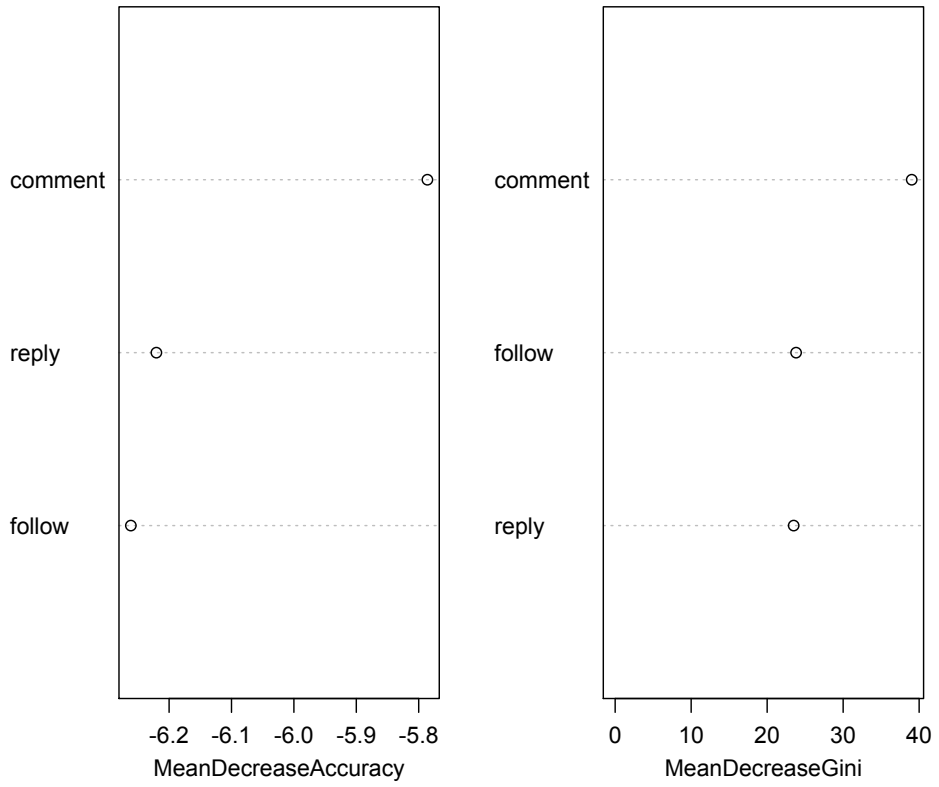


FIGURE 7.3: Mean accuracy of the implemented Random Forest model applied to the feature set containing raw comment, reply, follow attributes.

7.4.4.2 Performance of the raw variables in the classification with Random Forest Model

In order to clearly show the impact of the behaviour chains that were defined in this research, a Random Forest model is trained with the raw attributes of social behaviours i.e. the number of comment behaviours, the number of follow behaviour, the number of reply behaviour. The estimated overall error (OOB) indicates that the model is 67% likely to fail to predict completion of the course.

Table 7.5 shows the confusion matrix of the results. The model trained with raw attributes especially fails in predicting participants in the *low* (error rate: 0.96) and *high* (error rate: 0.95) completion classes. For instance, the model correctly predicted only 13 samples in the *low completion* class whereas it mispredicted 250 of them as *satisfactory* and 69 of them as *high* when they were supposed to be in the *low completion* class. In contrast to its failure on predicting low and high completion classes, the model was successful on predicting learners who completed the course at satisfactory level. The error rate for this class is 0.18.

This results indicate that the behaviour chains, which is an attempted interpretation of an analysis on participants' use of social affordances, is a better indicator in predicting course completion than quantity of the each social actions.

Since the model trained with raw variables performed well at predicting learners in satisfactory completion class, this time the model is trained with raw and selected features together to see if the performance of the model has been improved.

TABLE 7.6: Confusion matrix of the implemented Random Forest model applied to the feature set containing raw variables and selected feature attributes.

	Low completion	**Satisfactory completion**	**High completion**	**Class error**
Low completion	274	49	9	0.1746988
Satisfactory completion	166	168	37	0.5471698
High completion	94	63	136	0.5358362

Training the model with all the features together has slightly improved the result on predicting the learners who completed the course at satisfactory and high level but it

still performed best at predicting learners who completed less than 50% of the course steps.

7.4.4.3 Testing with Support Vector Machine

The instant package *e1071* in R was used for the implementation of Support Vector Machine.

Table 7.7 shows the performance results of implementation of Support Vector Machine.

TABLE 7.7: Confusion matrix of the implemented Support Vector Machine model.

	Low completion	**Satisfactory completion**	**High completion**
Low completion	310	187	98
Satisfactory completion	64	214	134
High completion	3	17	104

According to the results shown in the matrix showing the prediction performance in Table 7.7, the model especially performed well at predicting learners who had high completion. As it is seen, the model predicted 104 learners in the *high completion* class correctly, but mispredicted 20 of them: 3 of them as *low* and 17 of them as *satisfactory*.

7.4.4.4 Comparison of the Results by Models

Table 7.8 compares the results of the implemented Random Forest Model and Support Vector Machine algorithms.

According to the results, it is difficult to conclude which model performed better. The implemented Random Forest Model has performed very well to predict the learners who had low completion whereas the implemented Support Vector Machine correctly predicted the learners who had high completion. However, both models failed to predict learners who had satisfactory completion, which implies the completion of more than 50% of the course steps but less than 80% of the course steps.

TABLE 7.8: Comparison of the results of RFM and SVM.

	Random Forest Model	**Support Vector Machine**
Prediction of learners who had low completion	81% correctly predicted (acc: 66%)	52% correctly predicted (acc: 47%)
Prediction of learners who had satisfactory completion	52% correctly predicted (acc: 47%)	46% correctly predicted (acc: 52%)
Prediction of learners who had high completion	27% correctly predicted (acc: 32%)	84% correctly predicted (acc: 52%)

The learners who completed less than 50% of the course steps (*low completion*) were more likely to be less engaged with the social features as Chapter 4 and Chapter 5 presented. This could be the reason why Random Forest Model easily split the data and correctly predicted 81% of the samples in the low completion class.

In addition, the analysis presented in Chapter 5 indicated that some variables have monotonic correlation with course completion, rather than a linear correlation. This could be the reason why the Support Vector Machine model was successful at some predictions but the Random Forest Model failed.

Futhermore, due to the flexible learning environment of MOOCs, learners are completely free to choose whether or not actively participate in discussions and complete the steps. Therefore, there will always be some learners whose behaviour will be impossible to predict. For example, a determined learner who completed the course at a high level but was completely socially passive during the operation of the course or a learner who was socially engaged but left the course earlier. These kind of users maybe unpredictable from their social activities and it may cause misclassification. In order to reduce the error rate, learners' behaviours should be analysed more from different angles.

7.5 Discussion of the Results

7.5.1 Testing on Different MOOCs

Whitehill et al. (2017) investigate the accuracy of the results of MOOC dropout prediction models and suggest that the accuracy is related to which data is used for training the model. The authors suggest that training the classifier on the same course or iterations of the same course may lead to extremely optimistic accuracy estimates.

The MOOC that is used in this study is an eight-week course. However, most of the courses that are authored by the University of Southampton lasts in either four-week or less. The *unbroken* chain attribute related to completeness of chain would indicate attendance in every week in a short-period course. This, consequently, would give more precise but biased results. Therefore, to test the accuracy of the presented model, another run of the same 8-weeks DYRP MOOC has been used.

The model was trained on a run of DYRP MOOC in 2014 as it is explained in Section 7.4.2 and was tested on a 2016 run of DYRP MOOC where 6550 participants enrolled on. Table 7.9 shows the results of prediction from the Random Forest Model which was tested on the DYRP 2016.

TABLE 7.9: Confusion matrix of the implemented Random Forest Model which was tested on the DYRP 2016 MOOC.

	Low completion	**Satisfactory completion**	**High completion**
Low completion	5173	100	114
Satisfactory completion	834	103	142
High completion	16	16	52

The results indicate that the model performed similarly to the previous test. It correctly predicted most of the samples in the *low completion* class. Out of the samples in the *low completion* class, 5173 were predicted correctly, 100 of them were predicted as *satisfactory completion* and 114 of them were predicted as *high completion*.

Similarly to the results of the test on the same course, the model failed on predicting

the learners who had *satisfactory completion* class and predicted most of them (834) as *low completion*.

These findings imply that the proposed model can be compatible with other courses. To further improve the model, it should be tested on a different MOOC which is not an iteration of the same MOOC, and MOOCs from the different MOOC platforms.

7.5.2 Timely prediction for timely intervention

Prediction models are typically used for educational interventions as Chapter 6 provides some examples from the literature. According to the results of the model presented in this chapter, a learner could be red-flagged when they stop social contributions and drop the intensity of their contributions.

In this thesis, completion performance of learners was predicted at the end of the course. The model can also be implemented weekly to predict learners at risk of leaving the course. Additionally, other machine learning models such as the Hidden Markov Model which is very suitable to work with data structured as time-series, could be implemented. With these kind of weekly implementations, timely prediction during the course's operation could be possible. It should be noted again that this research does not concern the implementation of any educational interventions and does not claim that interventions cause course completion. However, the results imply that the implemented prediction models could be valuable in use for interventions to help participants stay on the course.

In order to show how the predictive power changes as the weeks of the MOOC progress, the data was rearranged weekly and the performance of the Random Forest Model was observed on the same participants who were already randomly selected for the tests (see sub-Section 7.4.1).

Figure 7.4 shows the weekly performance of the prediction model. The highest error rate is in the very first week by almost 60% misprediction. From Week 2 to Week4, the error rates remained between 50% and 54%. The error rates of the predictions made at the end of Week 6 and Week 7 were under 50% which are close to the lowest error rate which was observed at the end of the course (Week 8).

According to the results presented in sub-Section 7.4.4, the Random Forest Model was especially good at predicting those who were not going to complete the course at the end of the Week 8. When the model was run weekly, it is observed that the

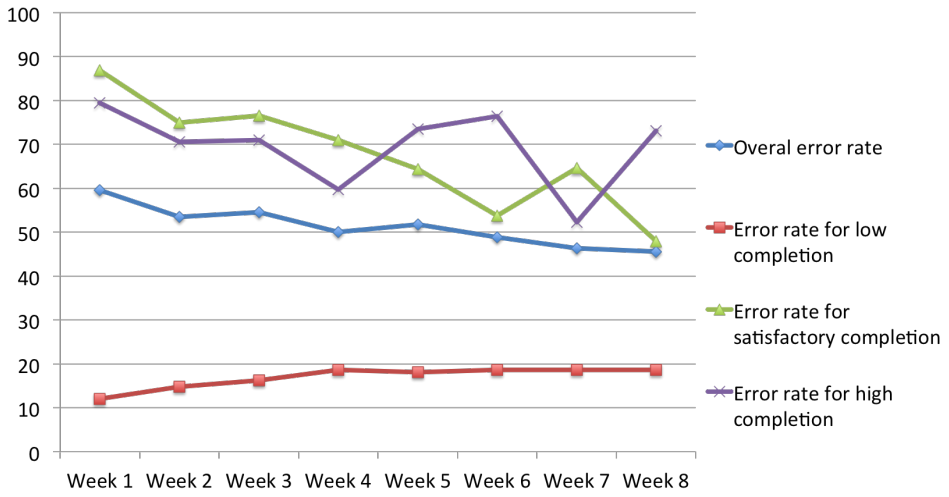


FIGURE 7.4: Error rates of the weekly implemented Random Forest model.

model correctly predicts the participants who would be in *low completion* (at the error rate 12%-18%). The performance of predictions on the participants who would be in *satisfactory* or *high completion* get betters throughout the course. However, these results indicate that it is possible to detect the participants who are not going to complete the course at the end of the first week. This is a very valuable opportunity to provide personalised help for those participants.

7.5.3 Additional Feature Extraction from Participants' Behaviour

FutureLearn gathers plain information about learners such as how many comments they posted, to whom they replied, when they posted, and which steps they completed. In this thesis, the gathered records were analysed to interpret and model learners' behaviours. The results indicated that the modelled behaviours were more predictive than the simple statistical information about learners' behaviours.

However, the model that was proposed in this thesis is not the only way of interpretation of learners' engagement with the course. Different features could be extracted from learners' behaviours and those could perform better in predicting. For instance, MOOC authors who provide shorter MOOCs could replace the *fullness of chain* feature with another behaviour which is more typically observed in a short term MOOC.

7.6 Summary

Our predictive model showed that the continuity of social participation of a participant, the frequency of their social actions, and type of their social contribution could be a strong predictor for course completion. The implemented Random Forest Model correctly predicted most of the test samples in *low completion* class while the implemented Support Vector Machine correctly predicted most of the test samples in *high completion* class. Both models failed in predicting samples in *satisfactory completion* class.

The most predictive indicator is the completeness of the chain which indicates the continuity of social participation over weeks. The findings also indicate that the behaviour chains which were modelled according to the pattern of interactions with social affordances, are much more predictive than the amount of interactions with social affordances.

The results of the predictive model suggest that it is possible to anticipate when a learner would dropout of the course based on their social engagements and it would be possible to take measurements to encourage the learner to continue on the course. The next chapter provides examples from the literature how prediction models could be used for educational interventions and discusses the further improvements of the model.

Chapter 8

Conclusions and Future Research

In 2016, over 58 million students around the world participated online in 6850 MOOCs organised by over 700 universities. MOOCs offer world-wide accessible online contents typically including video lectures, readings, quizzes along with social communication components on a platform that allows people to have their own personal experience with a course. Each individual, therefore, has their own pattern of engagement.

Learning designers orchestrate MOOC content to engage learners at scale and retain their interest by carefully mixing videos, lectures, readings, quizzes, discussions and activities. For example, in order to promote successful engagement, learners in the UK-based FutureLearn MOOC platform have opportunities to share and reflect on opinions by *posting* comments, *replying*, or *following* discussion threads. FutureLearn takes a social-constructivist approach where learning is constructed through conversations. Therefore the design of discussion threads is a very core component on the FutureLearn platform.

The research presented in this thesis investigated how MOOC learners socially engaged in a FutureLearn MOOC and the impact of engagement on course completion.

The following subsections answer the research questions presented in Chapter 1 and discuss the potential value and the limitation of the research and provide suggestions for future work to improve the research.

8.1 The Results of this Research

8.1.1 Participants' Engagement with Social Affordances on FutureLearn

In order to investigate how participants' social contributions were associated to course completion, two research questions were put in Chapter 1:
RQ1: How is showing social presence in a MOOC associated to the participant's performance in course completion?
RQ2: How do participants interact with social affordances that are provided by the FutureLearn MOOC platform?

The course analysed in this thesis had 9855 enrolled learners who showed initial interest, however only 51.6% (5086 participants) actually visited the course pages at least once after the official start date of the course. Amongst these 5086 participants, 1867 participants (36.7%) contributed to discussions by writing at least one comment and 789 (15.5%) followed at least one person on the platform during the duration of the course (Figure 4.1 in Chapter 4).

Considering the total number of enrolled learners, the proportion of the learners who interacted with a social affordance is relatively small. The data represented in Figures 4.4 and 4.5 shows that the participants socially engaged by mostly posting a comment to the discussions. The number of replies posted and the number of follow interactions initiated were much smaller. Additionally, the volume of social activities was the highest in the very first week of the course.

Chapter 4 presents an analysis of participants interactions with the social affordances provided on a social-constructivist MOOC platform. The findings showed that course completion and being socially active was 50% positively correlated. It was observed that participants who post a comment to discussions were highly likely to complete the course.

In addition, the data represented in Figures 4.12 and 4.13 shows that the participants who completed more than half of the course steps typically were socially active in multiple weeks throughout the duration of the course, whereas the participants who completed less than half of the steps were typically socially active in the first two weeks.

As the follow-up question, the third research question was stated as: ***RQ3****: How*

can we characterise the differences between completion rates comparing follow and discussion contribution behaviours?

The descriptive statistical analysis presented in sub-Section 4.2.3 differentiated the following behaviours of participants who followed or posted a comment to discussion:

- **Comment:** Learners who made even the smallest social contribution to discussions i.e posting a single comment, completed at least one step in the course even though the completion of a single step did not necessarily extend to course completion.
- **Follow:** Learners who exploited only the *follow* feature as a social behaviour, did not necessarily complete a step. However, the majority of the learners who only used the *follow* feature completed a larger number of course steps than the average number of course steps completed by the all learners in the course.
- **Follow & Comment:** The completion rate of followers who also posted a comment to discussions is higher than the completion rate of followers who only used the *follow* feature.

8.1.2 Identifying and Modelling Different Social Behaviours

After identifying some different behaviours of socially active learners, the next research question is interested in: ***RQ4:*** How can we typify the different patterns of participants' social behaviours during a course?

In order to make more sense of the data regarding to learners' social activities, the pattern of learners' engagement with the social affordances were analysed. As one of the novel contributions of this thesis, a *chain* model for representing learners' social engagement and peer interactions was proposed. Four main behaviour chains were defined according to the depth of peer interactions and frequency of social actions as: *simple*, *moderately simple*, *intensive*, and *persistent intensive*.

Comparing course completion of the learners clustered by behaviour chains, it is observed that the median value for the total number of completed steps by a learner is bigger once the learner makes deeper peer interactions. These findings are of our interest because they suggest a root to an additional research on prediction models. This result implies that there could be a positive correlation between the type of behaviour and course completion which could contribute to a prediction model.

8.1.3 Prediction of Course Completion based on Learners' Social Behaviours

The fifth question stated in this thesis was: ***RQ5****:* What are the most correlated social behaviours to course completion in a MOOC?

In order to answer this question, correlation between modelled behaviours and course completion were statistically tested. According to the findings, it was observed that course completion is positively correlated to the type of modelled behaviours. This correlation implies that learners who actively interact with other fellow participants e.g. replying to comments and following others, are highly likely to be amongst the course completers. However, the findings also indicate that the category of learners who have actively engaged with their peers are a very small proportion of the enrolled course participants.

Additionally, statistical correlation tests show that when learners who frequently exhibit of social behaviours and make social contribution continuously in consecutive weeks, their behaviours are also positively correlated to course completion.

In order to investigate if these correlated behaviours could successfully predict learners' course completion, the next and final question has been raised: ***RQ6****:* Can we use these correlated behaviours to predict participants' course completion?

Machine learning techniques were applied to the dataset to learn from learners' past behaviours and make predictions of their course completion. The features that were extracted from the modelled behaviours are type of behaviour chains, frequency of behaviours, and the continuity of behaviours (fullness of chain). The dataset structured according to these behavioural features were trained and tested with the two chosen algorithms which were the Random Forest Model and Support Vector Machine classifiers. Both the algorithms performed well and correctly predicted the majority of the samples in the dataset.

However, predictions sometimes failed when they predict someone who actually completed the course as non-completer. Since some learners who completed the course without any social interactions, the classifier failed on predicting. This problem could be overcome by combining prediction based on social interactions with predication based on other data known about the user. For example we could ask learners about their intention to complete the course before the course starts. Moreover, a clickstream data could be helpful to understand if a learner is viewing the pages before

the learner marks the step as *completed*. However, the datasets that we currently have do not let us know this kind of information.

8.2 Limitations of this Work

As already explained throughout the thesis, this research considers learners' interactions with social affordances on the platform, which are writing a comment, replying to a comment, liking a comment, and following cohorts. In the available dataset structure, however, having detail information about learners' *like* behaviours was not possible. The dataset only includes the total number of likes that a comment had. No information regarding who liked the comment was gathered. Learners' *like* behaviours could also have given some insight about learners' social engagement and peer interactions and its effect on overall course performance.

Furthermore, the dataset showing the *follow* interactions does not specify in which course a learner started following another. If the date that the *follow* relationship started and the date of the run of the course were matched, it was assumed that the *follow* relationship happened during this course. However, there is always a possibility that the learners enrolled on a different MOOC which runs parallel and that learners followed each other because of their interactions in the second MOOC. Unfortunately, the dataset does not allow us to know which link directed the learners to follow each other. If we had this information, the accuracy of the model could have been improved.

In this research, the correlation analyses were done and the prediction model, based on the correlated features, has been developed. It would be nice to build a recommender system which promotes peer interactions based on the results of the prediction model. Although the ethics approval was given by the University of Southampton, the FutureLearn MOOC platform did not allow us to build an external recommendation tools for the reason that the data could have been used for extracting personal information. Since FutureLearn also does not allow any third party to add any built-in technology in the platform, it was not possible for this research to use the findings to build a personalisation service on FutureLearn.

Therefore, this research only makes a claim about the correlation between course completion and being socially engaged in the platform. According to statistics, the participants who completed the course were mostly amongst those who were socially active. However, this research does not bring any evidence on that being socially active on the platform *causes* completion of the course. In order to understand the

reason and motivation of participants to use social affordances, pre-course and post-course questionnaires could be useful. Possible findings of a recommender system and questionnaires could show there is also a causal relationship between being socially active and course completion. The findings of this research, however, are not sufficient to claim this. Having said that, it is not unreasonable to consider that there might be some causal link between social behaviour and completion. Social interactions encourage and reward participation, engagement and time on task, which in turn might be expected to lead to good completion. In order to investigate this further it would be necessary to have access to the platform to encourage some users to participate on social interactions, and to measure their completion compared to a group who had not had the same encouragement. This work was not possible within FutureLearn as explained above.

Another limitation of this research is that correctly predicting a hundred percent of MOOC participants' completion is almost impossible. There are different kinds of participants who act in different ways in MOOCs. There are a bunch of learners who go through the whole MOOC without socially interacting. There are learners who do not watch the all videos but submit the assignments at the end. There is another type of learner who comes to the course for one particular subject and only engages with the activities of that week. These examples are also related to the discussions around what is the merit for success in MOOCs. In this research, the completion of a high number of learning objects were defined as a high achievement. This high achievement in course completion was correlated to the social behaviours. However, because of the learners who do not follow the commonly observed patterns, this approach will not predict some learners' behaviours. Especially, the model failed to predict the completion of learners who were neither in the *high completion* nor in the *low completion* classes, but those who were in the *satisfactory completion* class.

Finally, the prediction model presented in this thesis was only tested on the iterations of the same MOOC. The context of the eight-weeks *Developing Your Research Project* MOOC is in education. The participants may behave differently in MOOCs on different subjects such as science and health. Furthermore, learners' engagement patterns might have been effected by the length of the course, which is a quite long MOOC comparing to the average. The model needs to be tested on different types of MOOCs and different lengths of MOOCs.

8.3 Contribution of this Research

The contributions of this research could be summarised as follows:

- MOOC learners who participated socially are more likely to complete the course than others. The *follow* feature which is uniquely integrated into the platform as an implementation of the social-constructivist approach is actually valuable for understanding the engagement of learners in the course.
- The *behaviour chains* are proposed to model patterns of learners' interactions with social affordances. The modelled behaviours have been used in the prediction model as indicators.
- The proposed prediction model was developed to predict learners' completion as *low, satisfactory*, or *high* completion class. The prediction model was particularly successful on predicting the performance of learners in the *low* and *high* completion classes.

The findings of this research could be valuable for different actors in MOOCs. The following points give some ideas on how the results of this research could be used by different actors in MOOCs:

- **The learner:** A dynamic dashboard in conjunction with badges may help learners to change their behaviour and make the learners benefit more from their MOOC study.
- **The learning designers:** The learning designers could take measures to prevent their participants from dropping out the course by detecting learners who are at risk of *low completion*.
- **MOOC platform designers:** The platform designers could use the findings of this research to provide adaptive platform designs for promoting social contributions and peer interactions. Some of the implementations could be recommender systems, gamification features, and additional social affordances.
- **Other researchers:** The idea of behaviour chains could be used/modified for their research in MOOCs in order to understand their participants' behaviours. This may also help us to compare the behaviours of MOOC participants on different MOOC platforms.

  This research was particularly interested in predicting learners' behaviour based on their social engagement. Therefore, the prediction model only used the social behaviour of learners. However the model could be improved by adding other parameters such as participants' video engagement and demographical data.

8.4 Ideas for Future Work

8.4.1 Improving the model on different courses

As discussed in Section 7.5, when a classifier is trained and tested on the same course, the results might be unrealistically optimistic. Since social contribution in consecutive weeks is one of the predictors in the model, testing on a short-term MOOC could give misleading results. Since another set of data from a different eight-weeks course authored by the University of Southampton was not available at the time this research has been conducted, we could not train and test the data on a different MOOC, but a different iteration of the same eight-weeks MOOC.

However, other researchers from different partner institutions of FutureLearn could use the model proposed here and apply it to their data generated from a longer duration MOOC. In order to improve the model, the model could be modified according to behaviour analyses from a shorter duration MOOC as well.

Furthermore, although other MOOC platforms may not offer the same social features as FutureLearn, they collect data from their learners' interactions with the social features that have been provided by their MOOC platform. The researchers who analyse those data could identify similar behaviour chains which was proposed here. The results could let us compare the impact of social engagement on predicting course completion on different MOOC platforms.

8.4.2 Improving the data structure and the range of social affordances on the platform

The *follow* feature is uniquely integrated into the FutureLearn MOOC platform. The findings of this research show that this particular social feature has a value in MOOCs. The data generated from participants' interactions with this social feature could provide insights and potential of understandings of the behaviours of the MOOC participants.

These findings build a case for MOOC providers to integrate social features into their platforms. Learners are likely to be familiar through prior experience of devices such as follow, retweeting, tagging people and so on.

In addition, FutureLearn conducts pre-course and post-course questionnaires to collect demographic data about learners and their feedback about the course. However, additional questionnaires would be useful to know more about why do learners choose (not) to use the social affordances, what are their motivations to interact with a fellow learner, and do the learners actively follow the comments posted by the learners once they started *following* them. This kind of information could be useful to improve the use and effectiveness of social affordances on the platform.

8.4.3 Personalised recommenders in MOOCs

The findings of this research are also relevant to the study of personalised services. Researchers in the field of educational technology are paying huge attention to the widespread adoption of MOOCs learning online. An exploratory literature analysis has been completed to understand the situation in MOOC studies on personalisation (Sunar et al., 2015c,b).

In order to analyse the attention to personalised and adaptive MOOCs, available literature was searched on several academic databases between 2011 and 2016 with the keywords "MOOCs personalisation" and "adaptive MOOCs". The reason for starting with 2011 is twofold. First, 2011 is the first year in which both xMOOCs and cMOOCs were discussed (Daniel, 2012). Secondly, MOOCs had become rapidly and widely used in online learning as reported in the study of Liyanagunawardena et al. (2013). Figure 8.1 shows the number of retrieved results from the search for "adaptive MOOCs" only on Google Scholar on April 11, 2017.

It is observed in the research related to MOOCs that there is growing attention to adaptive MOOCs, especially personalisation of MOOCs since 2013. Researchers used learning analytics techniques for implementing personalisation and adaptation in MOOCs in order to improve user engagement and achievement, reducing drop-out rates. There are numbers of different implemented personalisation services such as personalised feedback (Shatnawi et al., 2014), adaptive content presentation (Sonwalkar, 2013), and personalised recommendation (Agrawal et al., 2015).

Chapter 2 has explained that FutureLearn took a social-constructivist approach which promotes learning through conversations (Laurillard, 2013). Conversations are crucial for learning in a social-constructivist MOOC and promoting conversations may have a good impact on participants' MOOC experience. The findings of this research can also be used to implement a personalised system which may help learners to be socially

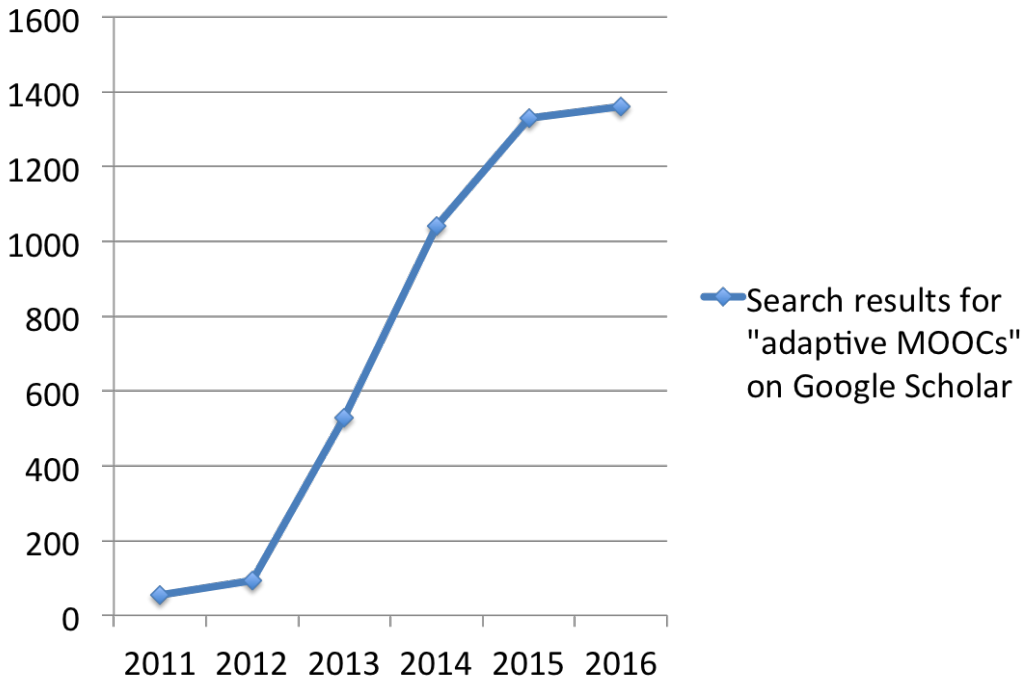


FIGURE 8.1: Research attention to personalisation in MOOCs.

more engaged and to complete a larger numbers of steps. For example:

- **A friend recommender system:** In this study, it is observed that the number of learners who repeatedly interacted with their peers is a very small percentage of the socially active learners.
  A friend recommendation system may help learners to be more socially active. Even though, it is not possible to know at this stage if this kind of system would help learners complete the course and boost the conversations, some widely-used friend recommendation systems e.g. Twitter's *who to follow* and Facebook's *people whom you may know* algorithms could be helpful to integrate a friend recommendation system into MOOCs.
- **A thread recommender system:** The volume of the threads is quite large on FutureLearn. The easy-to-use design of Twitter-like forum threads may also make it possible to miss some older comments which a learner may find interesting. A comment recommendation or an ongoing-conversation thread recommendation may help learners to get engaged with their peers by highlighting the potentially valuable threads.
  For example, a newspaper promoted the *Irish Lives in War and Revolution: Exploring Ireland's History 1912-1923* MOOC three weeks after the course launched. Consequently, a large number of learners joined the course three weeks late. More-

over, there are numbers of studies showing that there are always participants who join the course late even though they are not a huge chunk of people like in the Irish History MOOC. In such cases, it would be practical to point these learners to the discussion threads that were posted earlier. This might help learners to find a space where they can join in the community without becoming frustrated by the large volume of conversations.

8.4.4 Gamification

Another implementation could be an adoption of gamification into MOOCs. Many MOOC researchers have already been applying gamification techniques to their courses such as creating a leader board based on their achievements, delivering badges as they have progressed (Llanos et al., 2016).

Gamification techniques can be used to encourage learners to build a community of learning by interacting with the social affordances on the platform. This kind of additional change in the design of a platform could help learners benefit from the course more.

The findings of the research presented in this thesis may be useful for detecting who should be encouraged by notifications and badges. Additionally, the kind of badges could be designed based on the findings. For example, a *first contribution* badge for their first comment and another kind of *interaction* badges when they reply to their peer for promoting the peer interaction. Even though this research does not provide any evidence indicating that being active or interacting with peers causes completion of the course, the findings logically indicate that this kind of gamification implementation could possibly help learners to complete the course.

8.4.5 An Improved Model For Predicting Completion

The prediction model presented here indicated how modelling of social behaviours could be used to predict completion. The findings of the research showed that modelling behaviour *chains* provided a good prediction model but there is room for further improvement in the accuracy of prediction.

In order to build a more complete and accurate prediction model, other factors could be combined with the model presented in this research. For example, the prediction

model presented here considers the quantity of replies to identify peer interactions, however, semantic analysis could be used as an additional factor to identify the quality of peer interactions. Moreover, further factors beyond social behaviour could be included. Examples might be: any known information about the motivations of the learner; patterns of uses, e.g. frequency and regularity; and completeness of coverage of all elements of the course. Such factors might reveal the intentions of those who do not participate in the social aspects of the course.

Furthermore, this research made predictions by using the data which was collected at the end of the course. So, it did not predict learners' future behaviours at the end of the each week. The model could be tested weekly by restricting the dataset to the certain week using the same machine learning algorithms though, other algorithms which exploit temporal dynamics could be more suitable for predicting time-based events. For example Hidden Markov Models could be used for predicting participants' performance in the coming week by using the data collected at the end of each week. The weekly predictions could also be useful for providing timely interventions for supporting the learners.

8.5 Final Remarks

Even though education is a fundamental human right, it is still a luxury in some regions since millions of people do not have access to education. According to the human rights reports, 72 million children of primary education age are not in school and 759 million adults are illiterate[1]. The causes of the lack of education are various such as war, poverty, and inequalities that originate in sex, health and cultural identity.

Even though the most affected area in the world is Sub-Saharan Africa, the right to education is also a concern in the developed word. There are numbers of statistics showing that students may drop out of university because of high tuition fee[2].

Furthermore, in 2016, the United Nations (UN) declared Internet access as a fundamental human right[3]. As Figure 1.1 shows, however, there is a huge gap between the use of Internet in the developed world and developing world.

There are millions of geographically dispersed potential learners who have diverse

[1] http://www.humanium.org/en/world/right-to-education/

[2] http://theconversation.com/higher-tuition-fees-reduce-the-risk-of-students-dropping-out-of-university-44549

[3] https://www.article19.org/data/files/Internet_Statement_Adopted.pdf

educational backgrounds, learning requirements, languages, and motivations. MOOCs offer a potential to bridge that gap but there are issues of dealing with the diversity, poor Internet infrastructure, and sustainability of business model of MOOCs.

Because of the fast spread of MOOCs, people become enthusiastic with expectations of free university education, less debt after graduation, equal educational opportunities for people all around the world and so on. The newspapers have reflected this enthusiasm with their headlines such as *Instruction for Masses Knocks Down Campus Walls*[4]. As Daniel (2012) stated in his essay: "the discourse about MOOCs is overloaded with hype and myth while the reality is shot through with paradoxes and contradictions."

However, learning at scale is becoming more important in education. Increasingly universities not only provide MOOCs but they also use MOOCs in their typical face-to-face campus education. This practice is a development of the existing practices "flipped classroom" and "blended learning".

In the fifth European MOOC Stakeholders Summit, the CEO of FutureLearn Simon Nelson said that they are offering valuable professional qualifications and CPD accreditation through the platform. He also added that they are working on creating a core set of MOOC programmes so that it is possible to have a degree on particular areas.

Wildavsky (2015) discusses the global potential of MOOCs in the developing world in his study titled *MOOCs in the Developing World: Hope or Hype?*. The author gives an example, which is that EdX has launched a partnership with Facebook to introduce MOOCs technology as a form of cheap mobile learning. The pilot programme will start in Ruwanda.

Another example is that Kiron University in collaboration with edX offered free verified certificates for refugees who successfully completed the EdX courses in Berlin[5].

These examples indicate that there is a growing demand for MOOCs for various purposes. This means that there will be an even larger learning community with diverse educational and cultural backgrounds. In order to answer the need of each individual and improve the MOOC education, widely-applied traditional course design of MOOCs needs to be personalised. Therefore, we strongly need insightful analysis

[4] http://www.nytimes.com/2012/03/05/education/moocs-large-courses-open-to-all-topple-campus-walls.html

[5] http://blog.edx.org/new-partnership-with-kiron-enables-thousands-of-refugees-to-receive-college-credit-online

of MOOC data from different perspectives in order to understand how participants interact in the course and how their behaviours effect their achievements in the course so that we can use the analysis to help the participants.

This research was particularly interested in understanding the social behaviours of MOOC participants when the design of the platform promotes social conversations. Findings from the analysis of the participants' social behaviours contributed to develop a prediction model of course completion. Consequently, the prediction model enables course organisers to identify the participants who are at risk of leaving the course. Even though this research did not provide any personalised system, Section 8.4 offers possible implementations for answering individuals' needs by using the proposed prediction model. I believe the research presented in this thesis contributes to the knowledge of the educational affordances of socially enabled MOOCs and brings attention to the value of the social affordances in MOOCs to enable a personalised MOOC experience.

Appendix A

Documents for Ethics Approval

This study was originally designed for implementing a personalised social recommendation system. The original version of the research required two main methodological approaches: i) analysis of data which is collected form participants' online activities and ii) pre-/post-questionnaire with participants about their experience. The ethics approval was given for the application ID 9995 which is a study using MOOCs authored by University of Southampton and interviewing with the participants. However, as was discussed in the limitation of this work in Section 8.2, it was not possible for us to build a built-in system on FutureLearn. Therefore, although part of the work that I have ethical approval was never conducted, the work that was conducted is a subset of the ethical approval. Figure A.1 shows the approval.

The following pages show the document for application. The important information is highlighted.

Additionally, as a partner institution of FutureLearn, the University of Southampton has a right to do research with the provided **anonymised** datasets. The research ethics of FutureLearn is available on this link: https://about.futurelearn.com/terms/research-ethics-for-futurelearn.

FPSE Ethics Committee
FPSE EC Application Form Ver 6.6

Refer to the *Instructions* and to the *Guide* documents for a glossary of the key phrases in **bold** and for an explanation of the information required in each section. The *Templates* document provides some text that may be helpful in presenting some of the required information.

Replace the highlighted text with the appropriate information.

Note that the size of the text entry boxes provided on this form does **not** indicate the expected amount of information; instead, refer to the *Instructions* and to the *Guide* documents in providing the complete information required in each section. Do not duplicate information from one text box to another.

Reference number: **ERGO/**FPSE**/9995**	Version: 1	Date: 2014-05-09
Name of **investigator**(s): Ayse Saliha Sunar		
Name of supervisor(s) (if student **investigator**(s)): Hugh Davis and Su White		
Title of study: Understanding Social Media Behaviours of Learners		
Expected start date: 2014-05-19	Expected end date: 2014-06-30	

The investigator(s) undertake to:

- Ensure the study Reference number ERGO/FPSE/9995 is prominently displayed on all advertising and study materials;
- Conduct the study in accordance with the information provided in the Study Protocol, its appendices, and any other documents submitted;
- Conduct the study in accordance with University policy governing research involving human **participants** (http://www.southampton.ac.uk/corporateservices/rgo/);
- Submit the study for re-review (as an amendment through ERGO) if any changes, circumstances, or outcomes materially affect the information given;
- Promptly advise an appropriate authority (Research Governance Office) of any adverse study outcomes, changes, or circumstances (via an adverse event notification through ERGO);
- Seek FPSE EC advice in the event of material changes to the study following approval.
- Submit an end-of-study form as may be required by the Research Governance Office upon completion of the study.

REFER TO THE INSTRUCTIONS DOCUMENT WHEN COMPLETING THIS FORM.

PRE-STUDY

Characterise the proposed **participants:**
Participants are who enrolled MOOCs which are offered by the University of Southampton on the FutureLearn. Participants will be from different countries and belong to different age groups.

Describe how **participants** will be approached:
Participants will be invited to take part in the study by e-mail.

Describe how inclusion and/or exclusion criteria will be applied (if any):
All participants who answer the questionnaire will be included in the study.

Describe how **participants** will decide whether to take part:
Taking the questionnaire is completely voluntarily. Participants can decide to take part anytime during the estimated course time from the first day of the course to the last day of the last week.

Participant Information

Provide the **Participant Information** in the form that it will be given to **participants** as an appendix. All studies must provide **participant information**.

Consent Form

Provide the **Consent Form** (or the request for consent) in the form that it will be given to **participants** as an appendix. All studies must obtain **participant** consent. Some studies may obtain verbal consent, other studies will require written consent, as explained in the *Instructions* and *Guide* documents.

DURING THE STUDY

Describe the study procedures as they will be experienced by the **participant**:
On the first day of the online course, learners will be invited to take part in the study. This invitation will guide the volunteer learners to the link of the online survey. Participants can save and exit the questionnaire before completing all. After answering questions, they submit their answers.

Identify how, when, where, and what kind of data will be recorded (not just the formal research data, but including all other study data such as e-mail addresses and signed consent forms):
Data will collected by the iSurvey. All answers will be anonymised and only answers will be recorded for the study.

Participant questionnaire

As an appendix, reproduce any and all **participant** questionnaires or data gathering instruments in the exact forms that they will be given to or experienced by **participants**.

POST-STUDY

Identify how, when, and where data will be stored, processed, and destroyed.
Data will be stored on my personal page on the iSurvey website. After I complete my doctoral study, it will be destroyed by the end of 2017. If Study Characteristic M.1 applies, provide this information in the **DPA Plan** as an appendix instead.

STUDY CHARACTERISTICS

(L.1) The study is funded by a commercial organisation: **No** If 'Yes', provide details of the funder or funding agency:

(L.2) There are **restrictions** upon the study: **No** If 'Yes', explain the nature and necessity of the **restrictions**:

(L.3) Access to **participants** is through a third party: **No**

If 'Yes', provide evidence of your permission to contact them as an appendix.

(M.1) **Personal data** is collected or processed: **No**
Data will be processed outside the UK: **No**

If 'Yes' to either question, provide the **DPA Plan** as an appendix. (Note that retaining e-mail addresses, signed consent forms, or similar study-related **personal data** requires M.1 to be "Yes".)

(M.2) There is **inducement** to **participants**: **No**

If 'Yes', explain the nature and necessity of the inducement:

(M.3) The study is **intrusive**: **No**

If 'Yes', provide the **Risk Management Plan** and the **Debrief Plan** as appendices, and explain the nature and necessity of the intrusion(s) here:

(M.4) There is **risk of harm** during the study: **No**

If 'Yes', provide the **Risk Management Plan**, the **Contact Information**, and the **Debrief Plan** as appendices, and explain the necessity of the risks here:

(M.5) The true purpose of the study will be hidden from **participants**: **No**
The study involves **deception** of **participants**: **No**

If 'Yes' to either question, provide the **Debrief Plan** as an appendix, and explain the necessity of the deception here:

(M.6) **Participants** may be minors or otherwise have **diminished capacity**: **No**

If 'Yes', AND if one or more Study Characteristics in categories M or H applies, provide the **Risk Management Plan** and the **Contact Information**, as appendices, and explain here the special arrangements that will be put in place that will ensure informed consent:

(M.7) **Sensitive data** is collected or processed: **No**

If 'Yes', provide the **DPA Plan** as an appendix.

(H.1) The study involves: **invasive** equipment, material(s), or process(es); or **participants** who are not able to withdraw at any time and for any reason; or animals; or human tissue; or biological samples: **No**

If 'Yes', provide further details and justifications as one or more appendices. Note that the study will require separate approval by the Research Governance Office.

Technical details

If one or more Study Characteristics in categories M.3 to M.7 or H applies, provide the description of the technical details of the experimental or study design, the power calculation(s) which yield the required sample size(s), and how the data will be analysed as appendices.

APPENDICES (AS REQUIRED)

While it is preferred that this information is included here in the Study Protocol document, it may be provided as separate documents.

If provided separately, be sure to name the files precisely as "Participant Information", "Questionnaire", "Consent Form", "DPA Plan", "Permission to contact", "Risk Management Plan", "Debrief Plan", "Contact Information", and/or "Technical details" as appropriate.

If provided separately, each document must specify the reference number in the form ERGO/FPSE/xxxx, its version number, and its date of last edit.

Appendix (i): **Participant Information** in the form that it will be given to **participants.**

Appendix (ii): Questionnaire in the form that it will be given to **participants.**

Appendix (iii): **Consent Form** in the form that it will be given to **participants.**

Appendix (iv): **DPA Plan.**

Appendix (v): Evidence of permission to contact **participants** or prospective **participants** through any third party.

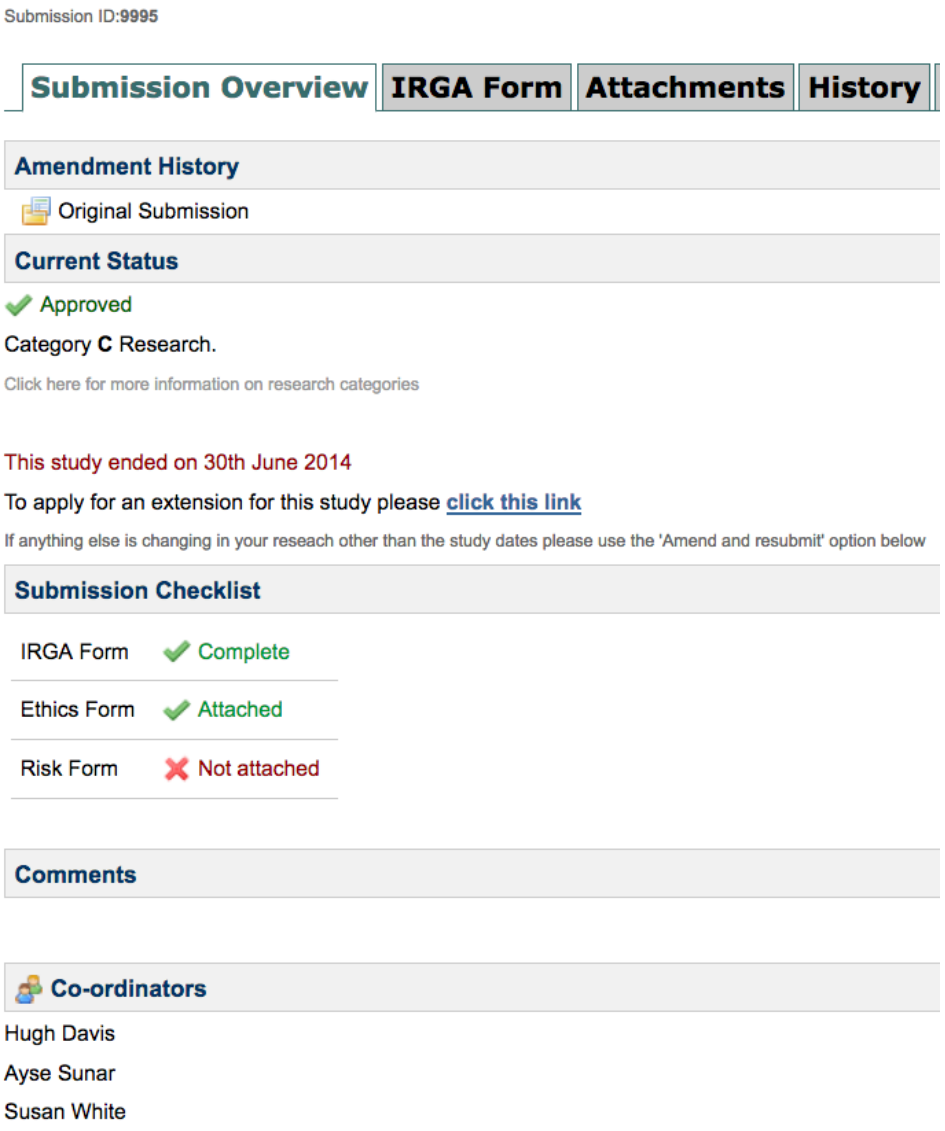
Appendix (vi): **Risk Management Plan**.

Appendix (vii): **Debrief Plan**.

Appendix (viii): **Contact Information**.

Appendix (ix): Technical details of the experimental or study design, the power calculation(s) for the required sample size(s), and how the data will be analysed.

Appendix (x): Further details and justifications in the case of **invasive** equipment, material(s), or process(es); **participants** who are not able to withdraw at any time and for any reason; animals; human tissue; or biological samples.


Submission ID:**9995**

Submission Overview | **IRGA Form** | **Attachments** | **History**

Amendment History

Original Submission

Current Status

Approved

Category **C** Research.

Click here for more information on research categories

This study ended on 30th June 2014

To apply for an extension for this study please **click this link**

If anything else is changing in your reseach other than the study dates please use the 'Amend and resubmit' option below

Submission Checklist

IRGA Form	Complete
Ethics Form	Attached
Risk Form	Not attached

Comments

Co-ordinators

Hugh Davis

Ayse Sunar

Susan White

FIGURE A.1: Ethical approval for the application ID: 9995

Appendix B

Implementation of Machine Learning Algorithms with *R*

Figure B.1 shows the piece of *R* code written for implementing the Random Forest Model.

```
library(randomForest)
library(caret)
library(plyr)
library(ROCR)

set.seed(32323)
rfdata = data.frame(read.csv('randomforestdata.csv', head=TRUE, sep=","))

k = 10 #10fold cross validation
rfdata$id <- sample(1:k, nrow(rfdata), replace=TRUE)
list<- 1:k

prediction <- data.frame()
testsetCopy <- data.frame()

progress.bar <- create_progress_bar("text")
progress.bar$init(k)

for (i in 1:k){
    training <- subset(rfdata, id %in% list[-i])
    testing <- subset(rfdata, id %in% c(i))

    randomforest <- randomForest(as.factor(completion) ~ fullness + frequency + chaintype, data=training,
 importance=TRUE, ntree = 500, trControl = rfControl, tuneGrid = rfGrid, metric = "Kappa", maximize = TRUE)

    temp <- as.data.frame(predict(randomforest, testing[,-1]))
    prediction <- rbind(prediction, temp)
    testsetCopy <- rbind(testsetCopy, as.data.frame(testing[,1]))

    progress.bar$step()
}

result <- cbind(prediction, testsetCopy[, 1])
names(result) <- c("Predicted", "Actual")
write.csv(result, file="predictedrff10fold.csv", row.names = FALSE )

print(randomforest)
summary(randomforest)
varImpPlot(randomforest)
```

FIGURE B.1: The *R* code for implementation of the Random Forest Model.

Figure B.2 shows the piece of *R* code written for implementing the Support Vector Machine.

```
library("e1071")
library(caret)
library(ROCR)

set.seed(32323)
svmdata = data.frame(read.csv('svmdata.csv', head=TRUE, sep=","))
attach(svmdata)
x <- subset(svmdata, select=-completion)
y <- completion

svmmodel1 <- svm(x,y, type='C')
# print(svmmodel1)

pred <- predict(svmmodel1,x)

svm_tune <- tune(svm, train.x=x, train.y=y, kernel="radial", ranges=list(cost=10^(-1:2), gamma=c(.5,1,2)))

svm_model_after_tune <- svm(completion ~ ., data=svmdata, type='C', kernel="radial", cost=1, gamma=0.5)
print(summary(svm_model_after_tune))

pred <- predict(svm_model_after_tune,x)

print(table(pred,y))

```

FIGURE B.2: The *R* code for implementation of the Support Vector Machine.

List of Figures

1.1 Internet users per 100 people over time (source: International Telecommunication Union). . . . 2
1.2 The methodological approaches in this research. . . . 9
1.3 The structure of the thesis and interrelations between chapters. . . . 12

2.1 FutureLearn MOOCs are structured around weeks and series of steps associated with weeks. . . . 17
2.2 A participant's profile page with the option of *Follow*. . . . 18
2.3 Discussion thread next to the course content (The 2017 version). . . . 19
2.4 The FutureLearn platform, highlighting social affordances within discussion thread (The 2014 version). . . . 20

4.1 Funnel of participation as observed in DYRP MOOC (Sunar et al., 2015a). . . . 33
4.2 Comparison of course completion of participants who are socially active and inactive. . . . 34
4.3 Number of completed steps by socially active and inactive learners (Total number of steps: 80). . . . 35
4.4 Volume of weekly social activities: *comments*, *replies*, and *followings* 36
4.5 Volume of weekly participants who are either a *poster*, *replier*, or *follower* . . . 36
4.6 Learners according to the time they start following somebody. . . . 37
4.7 Proportion of those who followed who contributed to discussions. . . . 37
4.8 Comparing discussion contributions between whose who did or did not follow others. . . . 37
4.9 Comparing the number of people whom a learner follows and the number of people with whom a learner interacted in the discussion forum. . . . 38
4.10 Proportions of completers and non-completers of learners in different categories. 39
4.11 Average percentages of the completed steps by learners in different categories. 40
4.12 Activities of completer learners [who followed at least once] throughout the course week-by-week (key on right). . . . 41
4.13 Activities of non-completer learners [who followed at least once] throughout the course week-by-week (key on right). . . . 41
4.14 Volume of social activities of mentors: *comments*, *replies*, and *followings*. . . . 43
4.15 Course completion ratio of mentors. . . . 44

5.1 Categories of social actions and behaviours. . . . 48
5.2 Extent and variety of participants' social actions. . . . 50
5.3 An example of weekly and overall chains of a learner. . . . 52
5.4 Number of participants in each group categorised by social chain types. . . . 52
5.5 Boxplot for course completion of learners by their level of social engagement. 54

5.6 Frequency of social actions of learners grouped by completion status. 57
5.7 Boxplot for frequency of social actions of learners grouped by completion status. 58
5.8 Distribution of one-week contributors' social actions over weeks. 60
5.9 Comparing the step completions of participants from different clusters. 61

7.1 Illustration of k-fold cross-validation with k=4 (source: Wikipedia). 79
7.2 Mean accuracy of the implemented Random Forest model. 81
7.3 Mean accuracy of the implemented Random Forest model applied to the feature set containing raw comment, reply, follow attributes. 82
7.4 Error rates of the weekly implemented Random Forest model. 88

8.1 Research attention to personalisation in MOOCs. 100

A.1 Ethical approval for the application ID: 9995 vii

B.1 The *R* code for implementation of the Random Forest Model. ix
B.2 The *R* code for implementation of the Support Vector Machine. x

List of Tables

2.1 Specific functionality and features in the FutureLearn MOOC platform 21
2.2 List of FutureLearn Datasets and their Attributes (The 2014 version) 22

3.1 Three main motivation to classify MOOC learners' behaviours. 26
3.2 Use of social activities for classification by researchers 30

4.1 The number of steps in each week. 39

5.1 Definitions of actions observed in a-week-long period in FutureLearn. 49
5.2 Behaviour categories based on chains that are defined in this study 51
5.3 The result of correlation between chain type and course completion. 56
5.4 Probability of course completion according to the frequency of social actions. 58
5.5 Completeness of chains 62
5.6 Probability of course completion according to the fullness of chain. 62

6.1 State-of-the-art techniques for predicting learners' participation in MOOCs. . 69
6.2 State-of-the-art techniques for predicting learners' participation in MOOCs (continued to Table 6.1). 70
6.3 Milestones of dropout definitions used and remarkable findings of these studies. 71
6.4 Milestones of dropout definitions used and remarkable findings of these studies (continued to Table 6.3). 72
6.5 Milestones of dropout definitions used and remarkable findings of these studies (continued to Table 6.4). 73

7.1 Attributes in the selected feature set for the construction of the prediction model. 76
7.2 The number of people in each class of course completion. 78
7.3 Workflow of the implementation of machine learning algorithms which have been chosen. 80
7.4 Confusion matrix of the implemented Random Forest model. 81
7.5 Confusion matrix of the implemented Random Forest model applied to the feature set containing raw comment, reply, follow attributes. 82
7.6 Confusion matrix of the implemented Random Forest model applied to the feature set containing raw variables and selected feature attributes. 83
7.7 Confusion matrix of the implemented Support Vector Machine model. 84
7.8 Comparison of the results of RFM and SVM. 85
7.9 Confusion matrix of the implemented Random Forest Model which was tested on the DYRP 2016 MOOC. 86

Bibliography

Aghaei, S., M. A. Nematbakhsh, and H. K. Farsani

2012. Evolution of the world wide web: From web 1.0 to web 4.0. *International Journal of Web & Semantic Technology*, 3(1):1–10.

Agrawal, A., J. Venkatraman, S. Leonard, and A. Paepcke

2015. YouEDU: Addressing confusion in MOOC discussion forums by recommending instructional video clips. In *Eight International Conference on Educational Data Mining*. Stanford InfoLab.

Amnueypornsakul, B., S. Bhat, and P. Chinprutthiwong

2014. Predicting attrition along the way: The UIUC model. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.

Anderson, A., D. Huttenlocher, J. Kleinberg, and J. Leskovec

2014. Engaging with massive online courses. In *23rd international conference on World Wide Web*, Seoul, Korea. ACM.

Arlot, S., A. Celisse, et al.

2010. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79.

Balakrishnan, G.

2013. Predicting student retention in massive open online courses using hidden markov models. Master's thesis, EECS Department, University of California, Berkeley.

Bali, M.

2014. MOOC pedagogy: Gleaning good practice from existing MOOCs. *Journal of Online Learning and Teaching*, 10(1):44–56.

Berners-Lee, T., J. Hendler, O. Lassila, et al.

2001. The semantic web. *Scientific American*, 284(5):28–37.

Bittencourt, I. I., S. Isotani, E. Costa, and R. Mizoguchi

2008. Research directions on semantic web and education. *Interdisciplinary Studies in Computer Science*, 19(1):60–67.

Bote-Lorenzo, M. L. and E. Gómez-Sánchez

2017. Predicting the decrease of engagement indicators in a MOOC. In *Seventh International Learning Analytics & Knowledge Conference*, Vancouver, Canada. ACM.

Breiman, L.

2001. Random forests. *Machine learning*, 45(1):5–32.

Brown, A. R. and B. D. Voltz

2005. Elements of effective e-learning design. *The International Review of Research in Open and Distributed Learning*, 6(1):1.

Casey, D. M.

2008. The historical development of distance education through technology. *TechTrends*, 52(2):45.

Chaplot, D. S., E. Rhim, and J. Kim

2015. Predicting student attrition in MOOCs using sentiment analysis and neural networks. In *Fourth Workshop on Intelligent Support for Learning in Groups*, Madrid, Spain.

Chatti, M. A., A. L. Dyckhoff, U. Schroeder, and H. Thüs

2012. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.

Chawla, N. V., N. Japkowicz, and A. Kotcz

2004. Editorial: special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1):1–6.

Clough, G., A. C. Jones, P. McAndrew, and E. Scanlon

2009. Informal learning evidence in online communities of mobile device enthusiasts. *Mobile learning: Transforming the delivery of education and training*, (1):99–112.

Clow, D.

2013. MOOCs and the funnel of participation. In *3th International Conference on Learning Analytics and Knowledge*, Leuven, Belgium. ACM.

Coffrin, C., L. Corrin, P. de Barba, and G. Kennedy

2014. Visualizing patterns of student engagement and performance in MOOCs. In *4th International Conference on Learning Analytics and Knowledge*, volume 2012, P. 1, Indianapolis, USA. ACM.

Coleman, C. A., D. T. Seaton, and I. Chuang

2015. Probabilistic use cases: Discovering behavioral patterns for predicting certification. In *2nd ACM Conference on Learning@ Scale*, Vancouver, Canada. ACM.

Cortes, C. and V. Vapnik

1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Daniel, J.

2012. Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*.

Downes, S.

2008. Places to go: Connectivism & connective knowledge. *Innovate: Journal of Online Education*, 5(1):1.

Dutt, A., S. Aghabozrgi, M. A. B. Ismail, and H. Mahroeian

2015. Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, 5(2):112.

Eradze, M. and M. Laanpere

2014. Interrelation between pedagogical design and learning interaction patterns in different virtual learning environments. In *International Conference on Learning and Collaboration Technologies*. Springer.

Evans, B. J. and R. B. Baker

2016. Moocs and persistence: Definitions and predictors. *New Directions for Institutional Research*, 2015(167):69–85.

Ferguson, R. and D. Clow

2015. Consistent commitment: Patterns of engagement across time in Massive Open Online Courses (MOOCs). *Journal of Learning Analytics*, 2(3):55–80.

Ferguson, R. and M. Sharples

2014. Innovative pedagogy at massive scale: teaching and learning in MOOCs. In *9th European Conference on Technology Enhanced Learning*, Graz, Austria. Springer.

Fini, A.

2009. The technological dimension of a massive open online course: The case of the CCK08 course tools. *The International Review of Research in Open and Distributed Learning*, 10(5):1.

Finlay, S.

2014. *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*. Palgrave Macmillan.

Gelman, B., M. Revelle, C. Domeniconi, A. Johri, and K. Veeramachaneni
2016. Acting the same differently: A cross-course comparison of user behavior in MOOCs. In *9th International Conference on Educational Data Mining*, Raleigh, NC, USA.

Gerstein, J.
2014. *Moving from education 1.0 through education 2.0 towards education 3.0*, Experiences in Self-Determined Learning. CreateSpace Independent Publishing Platform.

Gillani, N., R. Eynon, M. Osborne, I. Hjorth, and S. Roberts
2014. Communication communities in MOOCs. *arXiv preprint arXiv:1403.4640*.

Guàrdia, L., M. Maina, and A. Sangrà
2013. MOOC design principles: A pedagogical approach from the learners perspective. *eLearning Papers*, 33:1–6.

Halawa, S., D. Greene, and J. Mitchell
2014. Dropout prediction in MOOCs using learner activity features. In *2th MOOC European Stakeholders Summit*, Lausanne, Switzerland.

Hall, M. A. and L. A. Smith
1997. Feature subset selection: a correlation based filter approach. In *International Conference on Neural Information Processing and Intelligent Information Systems*, Dunedin, New Zealand.

He, J., J. Bailey, B. I. Rubinstein, and R. Zhang
2015. Identifying at-risk students in Massive Open Online Courses. In *29th AAAI Conference on Artificial Intelligence*, Texas, USA.

Hernández-García, Á., I. González-González, A. I. Jiménez-Zarco, and J. Chaparro-Peláez
2015. Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, 47:68–80.

Hlosta, M., Z. Zdrahal, and J. Zendulka
2017. Ouroboros: early identification of at-risk students without models based on legacy data. In *7th International Learning Analytics & Knowledge Conference*, Pp. 6–15, Vancouver, Canada. ACM.

Hmedna, B., A. El Mezouary, O. Baz, and D. Mammass
2017. Identifying and tracking learning styles in moocs: A neural networks approach. *International Journal of Innovation and Applied Studies*, 19(2):267.

Hung, J.-L. and K. Zhang
2008. Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*, 4(4):426–437.

Jiang, S., S. M. Fitzhugh, and M. Warschauer

2014a. Social positioning and performance in MOOCs. In *Workshop on Graph-Based Educational Data Mining, EDM 2014*, P. 14, London, UK.

Jiang, S., A. Williams, K. Schenke, M. Warschauer, and D. O'dowd

2014b. Predicting MOOC performance with week 1 behavior. In *7th International Confrence on Educational Data Mining*, London, UK.

Joksimović, S., N. Dowell, O. Skrypnyk, V. Kovanović, D. Gašević, S. Dawson, and A. C. Graesser

2015. How do you connect?: Analysis of social capital accumulation in connectivist MOOCs. In *5th International Conference on Learning Analytics And Knowledge*, Poughkeepsie, New York.

Kennedy, J.

2014. Characteristics of massive open online courses (MOOCs): A research review, 2009-2012. *Journal of Interactive Online Learning*, 13(1):1–16.

Khalil, H. and M. Ebner

2014. MOOCs completion rates and possible methods to improve retention-a literature review. In *EdMedia: World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Tampere, Finland.

Khalil, M. and M. Ebner

2016. What is learning analytics about? a survey of different methods used in 2013-2015. In *9th Smart Learning Conference*, Dubai, UAE.

Kizilcec, R. F., C. Piech, and E. Schneider

2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *3rd International Conference on Learning Analytics and Knowledge*, Leuven, Belgium. ACM.

Kloft, M., F. Stiehler, Z. Zheng, and N. Pinkwart

2014. Predicting MOOC dropout over weeks using machine learning methods. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.

Koller, D., A. Ng, C. Do, and Z. Chen

2013. Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3):62–63.

Laurillard, D.

2013. *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. Routledge.

León, M., S. White, S. White, and K. Dickens
2015. Mentoring at scale: MOOC mentor interventions towards a connected learning community. *EMOOCs 2015 European MOOC Stakeholders Summit.*

Li, W., M. Gao, H. Li, Q. Xiong, J. Wen, and Z. Wu
2016. Dropout prediction in moocs using behavior features and multi-view semi-supervised learning. In *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, Canada. IEEE.

Liang, J., C. Li, and L. Zheng
2016. Machine learning application in moocs: Dropout prediction. In *11th International Conference on Computer Science & Education (ICCSE)*, Pp. 52–57, Nagoya, Japan. IEEE.

Liyanagunawardena, T. R., A. A. Adams, and S. A. Williams
2013. MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning*, 14(3):202–227.

Llanos, D. R., J. Fresno, H. Ortega-Arranz, A. Ortega-Arranz, and A. Gonzalez-Escribano
2016. Applying gamification in a parallel programming course. *Gamification-Based E-Learning Strategies for Computer Programming Education*, P. 106.

Mackness, J., M. Waite, G. Roberts, and E. Lovegrove
2013. Learning in a small, task–oriented, connectivist MOOC: Pedagogical issues and implications for higher education. *The International Review Of Research In Open And Distributed Learning*, 14(4):140–159.

Mi, F. and D.-Y. Yeung
2015. Temporal models for predicting student dropout in massive open online courses. In *IEEE International Conference on Data Mining Workshop (ICDMW)*, Atlantic City, NJ.

Milligan, C., A. Littlejohn, and A. Margaryan
2013. Patterns of engagement in connectivist MOOCs. *Journal of Online Learning and Teaching*, 9(2):149.

Musser, J. and T. Oreilly
2006. Web 2.0. *Principles and Best Practices.[Excerpt]. oO: O'Reilly Media.*

Nakano, Y. I., S. Nihonyanagi, Y. Takase, Y. Hayashi, and S. Okada
2015. Predicting participation styles using co-occurrence patterns of nonverbal behaviors in collaborative learning. In *International Conference on Multimodal Interaction*, New York, NY, USA. ACM.

Nath, K., S. Dhar, and S. Basishtha

2014. Web 1.0 to web 3.0-Evolution of the Web and its various challenges. In *International Conference on Optimization, Reliabilty, and Information Technology (ICROIT)*, Faridabad, India. IEEE.

Nihonyanagi, S., Y. Hayashi, and Y. I. Nakano

2014. Analyzing co-occurrence patterns of nonverbal behaviors in collaborative learning. In *7th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye-Gaze & Multimodality*. ACM.

Pal, M.

2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222.

Papamitsiou, Z. K. and A. A. Economides

2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4):49–64.

Parker, K. R. and J. T. Chao

2007. Wiki as a teaching tool. *Interdisciplinary journal of knowledge and learning objects*, 3(1):57–72.

Ramesh, A., D. Goldwasser, B. Huang, H. Daume III, and L. Getoor

2014. Learning latent engagement patterns of students in online courses. In *28th AAAI Conference on Artificial Intelligence*, Quebec, Canada.

Reeves, T.

1993. Interactive learning systems as mindtools. *Viewpoints*, 2:2–11.

Ristoski, P. and H. Paulheim

2016. Semantic web in data mining and knowledge discovery: A comprehensive survey. *Web semantics: science, services and agents on the World Wide Web*, 36:1–22.

Robinson, C., M. Yeomans, J. Reich, C. Hulleman, and H. Gehlbach

2016. Forecasting student achievement in MOOCs with natural language processing. In *6th International Conference on Learning Analytics & Knowledge*, Edinburgh, UK.

Rodriguez, C. O.

2012. MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning*, 15(2):1–13.

Romero, C. and S. Ventura

2007. Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1):135–146.

Shahiri, A. M., W. Husain, and N. A. Rashid

2015. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72:414–422.

Sharkey, M. and R. Sanders

2014. A process for predicting MOOC attrition. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.

Sharma, K., P. Jermann, and P. Dillenbourg

2015. Identifying styles and paths toward success in MOOCs. *International Educational Data Mining Society*, Pp. 408–411.

Shatnawi, S., M. M. Gaber, and M. Cocea

2014. Automatic content related feedback for MOOCs based on course domain ontology. In *15th International Conference on Intelligent Data Engineering and Automated Learning*, Salamanca, Spain. Springer.

Siemens, G.

2005. Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1):3–10.

Sinha, T., N. Li, P. Jermann, and P. Dillenbourg

2014. Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners. In *EMNLP Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, Doha, Qatar.

Sonwalkar, N.

2013. The first adaptive MOOC: A case study on pedagogy framework and scalable cloud architecture-part i. In *MOOCs Forum*, volume 1, Pp. 22–29. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

Sunar, A. S., N. A. Abdullah, S. White, and H. C. Davis

2015a. Analysing and predicting recurrent interactions among learners during online discussions in a MOOC. In *11th International Conference on Knowledge Management.*, Osaka, Japan.

Sunar, A. S., N. A. Abdullah, S. White, and H. C. Davis

2015b. Personalisation in MOOCs: A critical literature review. In *Lecture Notes in Communications in Computer and Informaiton Science (CCIS)*, Pp. 152–168. Springer-Verlag.

Sunar, A. S., N. A. Abdullah, S. White, and H. C. Davis

2015c. Personalisation of MOOCs: The state of the art. In *7th International Conference on Computer Supported Education CSEDU*, Lisbon, Portugal.

Sunar, A. S., S. White, N. A. Abdullah, and H. C. Davis

2016. How learners interactions sustain engagement: a MOOC case study. *IEEE Transactions on Learning Technologies*, PP(99).

Taylor, C., K. Veeramachaneni, and U. M. O'Reilly

2014. Likely to stop? predicting stopout in massive open online courses. *arXiv preprint arXiv:1408.3382*.

van Mierlo, T.

2014. The 1% rule in four digital health social networks: An observational study. *Journal of medical Internet research*, 16(2):e33.

Wang, X., M. Wen, and C. P. Rosé

2016. Towards triggering higher-order thinking behaviors in MOOCs. In *6th International Conference on Learning Analytics & Knowledge*, Edinburgh, UK.

Wang, Y. and R. Baker

2015. Content or platform: Why do students complete MOOCs? *MERLOT Journal of Online Learning and Teaching*, 11(1):17–30.

Whitehill, J., K. Mohan, D. Seaton, Y. Rosen, and D. Tingley

2017. Delving deeper into mooc student dropout prediction. *arXiv preprint arXiv:1702.06404*.

Wildavsky, B.

2015. Moocs in the developing world: Hope or hype? *International Higher Education*, (80):23–25.

Xu, B. and D. Yang

2016. Motivation classification and grade prediction for MOOCs learners. *Computational Intelligence and Neuroscience*, 2016(4):1–7.

Yang, D., T. Sinha, D. Adamson, and C. P. Rose

2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *NIPS Data-Driven Education Workshop*, Nevada, USA.

Yang, D., M. Wen, A. Kumar, E. P. Xing, and C. P. Rose

2014a. Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *The International Review of Research in Open and Distributed Learning*, 15(5):214–234.

Yang, D., M. Wen, and C. Rose

2014b. Peer influence on attrition in massively open online courses. In *Educational Data Mining 2014*, London, UK.

Yeager, C., B. Hurley-Dasgupta, and C. A. Bliss

2013. cMOOCs and global learning: An authentic alternative. *Journal of Asynchronous Learning Networks*, 17(2):133–147.

Yuan, L., S. MacNeill, and W. Kraan

2008. Open Educational Resources–opportunities and challenges for higher education. Technical report, Joint Information Systems Committee (JISC) CETIS.

Zaiane, O. R. and J. Luo

2001. Towards evaluating learners' behaviour in a web-based distance learning environment. In *IEEE International Conference on Advanced Learning Technologies*, Madison, USA.