Characterising Enzymes for Information Processing: Towards an Artificial Experimenter

Chris Lovell, Gareth Jones, Steve R. Gunn, and Klaus-Peter Zauner

School of Electronics and Computer Science,
University of Southampton, UK, SO17 1BJ
{cjl07r,gj07r,srg,kpz}@ecs.soton.ac.uk

Abstract. The information processing capabilities of many proteins are currently unexplored. The complexities and high dimensional parameter spaces make their investigation impractical. Difficulties arise as limited resources prevent intensive experimentation to identify repeatable behaviours. To assist in this exploration, computational techniques can be applied to efficiently search the space and automatically generate probable response behaviours. Here an artificial experimenter is discussed that aims to mimic the abilities of a successful human experimenter, using multiple hypotheses to cope with the small number of observations practicable. Coupling this approach with a lab-on-chip platform currently in development, we seek to create an autonomous experimentation machine capable of enzyme characterisation, which can be used as a tool for developing enzymatic computing.

1 Introduction

Elementary molecular computing has been composed of synthetic molecules used as logical operators [1]. Biomolecules too, for example DNA [2] and enzymes [3], have been employed as Boolean logic gates. However, given the structural complexity of enzymes, and recognising the influence the chemical environment has over enzymatic behaviour, it would appear that enzymatic behaviour is not limited to simple Boolean logic behaviour. Instead, by moving beyond mimicking digital electronics, characterising the response behaviour of enzymes could support new modes of information processing, and ultimately facilitate the application of enzymatic computers [3]. However, resources are typically very limited compared to the large parameter spaces, preventing detailed investigation of behaviours. Effective choice of experiments and a physical platform that minimises resource requirements per experiment, would therefore be desirable.

In consideration here is an artificial experimenter, comprising of a set of machine learning techniques that analyse experimental observations to propose possible hypotheses and determine experiments to perform to test those hypotheses. A key problem for the artificial experimenter is how to produce hypotheses from small amounts of observations. Additionally, as experiments can in some cases be unreliable, resulting in erroneous observations not representative of the true underlying behaviours present, the development of hypotheses must also
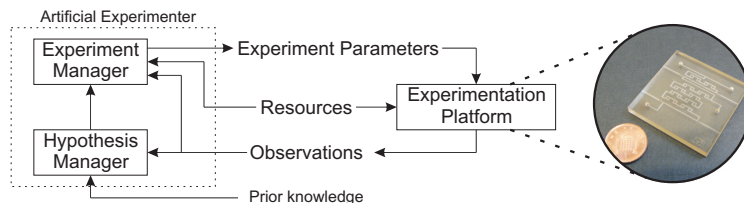
Fig. 1. Flow of experimentation between an artificial experimenter and an automated experimentation platform. A prototype of the lab-on-chip platform in development is shown.

take into consideration the validity of an observation. We use a multiple hypotheses technique, whereby observations that are considered by the artificial experimenter to be potentially erroneous, are considered both erroneous and valid in parallel through multiple competing hypotheses, until further experimentation can provide clarity. The ability of a computational system to consider many thousands of hypotheses simultaneously, provides an advantage over a human experimenter, who can contemplate only a much more limited number of hypotheses.

Also in development is a lab-on-chip platform, where multiple laboratory functions, for instance on-chip optical absorbance measurement, pumping and mixing, are fully automated [4]. The platform uses microfluidic technology to minimise consumption of chemicals per experiment [5]. As shown in Fig. 1, the artificial experimenter described here can be coupled to such an automated platform to allow for autonomous experimentation. In the following, existing techniques for autonomous experimentation are considered in Section 2, the specification of requirements and prototype methods for an artificial experimenter are presented in sections 3, 4 and 5, with initial results considered in Section 6.

2 Existing Approaches

Active learning provides many techniques for determining the next experiment to perform [6, 7]. However, these techniques often look at a simplified view of experimentation, for example where the problem is that of binary classification and there is no noise in the classification [8]. These techniques typically provide mathematical guarantees, such as converging to the optimal hypothesis or decreases in the prediction error. However, first the assumptions made and second the high number of experiments required by these approaches, make them infeasible for problem domains of interest here. Approaches that do look at more complicated problems, such as response prediction with noisy observations, are evaluated based on the use of several hundred experiments per parameter dimension [9, 10], not suitable for the present problem.

Existing artificial experimenter techniques that have been applied to real experimentation, often take more ad-hoc approaches so as to try and mimic the

techniques employed by successful experimenters. For example, the work of Hans Krebs to discover the urea cycle was used as a case study to develop heuristics for scientific discovery in the KEKADA system [11]. One of the heuristics proposed was that of investigating surprising observations, which was also considered an important heuristic in an approach called Scouting [12]. The heuristic of surprise can be applied to the present approach, as obtaining a surprising observation, otherwise described as an observation that disagrees with a hypothesis, suggests the discovery of a behaviour not captured by the hypothesis. However, both the KEKADA and Scouting approaches have limited effectiveness in the present problem domain. KEKADA due to other heuristics requiring explicit a priori domain information, not available in the present domain, and the Scouting approach having a restricted technique for representing the response behaviours required. The requirement of extensive domain information to develop mechanistic hypotheses makes the Robot Scientist approach [13] not applicable for enzymatic characterisation, where the required prior information does not exist and mechanistic hypotheses are not required. An approach using regression to identify response behaviours in electrochemistry, although well suited to identifying interesting phenomena, is restricted by requiring a high number of experiments and not considering the case of erroneous observations [14]. Symbolic regression has also been applied to rediscover physics laws [15], however the technique once more required a large number of observations.

3   Problem Specification

Currently few models of enzyme dynamics exist. Therefore, to evaluate these artificial experimenter techniques, simulated experiments are used to provide a known target function that can be compared against. In formulating the problem framework, an experiment is defined as a set of parameters and actions, represented as vector $\mathbf{x}$. Similarly, the observation for an experiment is represented as vector $\mathbf{y}$, considered here to contain only a single element. Subsequently, the underlying behaviour present in the experiment parameter space can be represented as a function:

$$\mathbf{y} = f(\mathbf{x} + \delta) + \epsilon + \phi \tag{1}$$

with Gaussian noise affecting both the experiment parameters and observations through δ and ϵ respectively. In addition a shock noise term (ϕ) is included that can shift the observation. Shock noise factors in failure within experimentation that yields an observation unrepresentative of the true underlying behaviour. The shock noise can be specified as a percentage of experiments that will lead to erroneous observations, where only that percentage of experiments yield $\phi \neq 0$. This simple yet more general noise model is assumed for testing, as the actual noise model is not known.

Having constructed the problem framework, the goals of the experimentation algorithms can be defined. A theoretical goal is to find an accurate representation of $f(\mathbf{x})$, however in reality, accuracy may be limited by the resources available.

Therefore key aspects, such as any peaks, troughs and sharp changes in behaviour are to be identified, along with the general trend of the behaviour. The existence and knowledge of repeatable behaviours, may form the basis of properties harnessed for enzymatic computation. Later, optimisation at particular points of interest can be used to refine accuracy. Additionally, the artificial experimenter should attempt to identify erroneous observations, to prevent those observations from incorrectly influencing the hypotheses, discussed next.

4   Hypothesis Management

Responses from enzyme interaction experiments may be nonmonotonic and in rare situations include a phase change [3]. Therefore, response hypotheses need to be general so as to allow different nonmonotonic functions, but also flexible to allow for abrupt feature changes such as phase changes. For these reasons, hypotheses are represented here using a smoothing spline. A smoothing spline is an established regression technique that can be placed within a Bayesian framework to provide error bars and does not impose a particular spectral scale [16]. A single hyperparameter (λ) controls the smoothness of the regression, with a higher value representing a smoother output. A hypothesis is therefore determined by the smoothing parameter λ, the set of observations to train from, and weightings on those observations. The weights are determined through the procedure described below to handle erroneous observations.

Techniques such as outlier identification, are typically employed to determine whether it is statistically likely for a particular observation to be valid when compared to all other observations. However, with only a few observations available, these techniques cannot be applied with any acceptable level of certainty. Take the example in Fig. 2. A reasonable suggestion for the given data would be either the linear prediction in h_3 or the curve of h_4, which both would be acceptable outcomes from regression using cross validation or similar to learn the parameters. However, if any of the observations are invalid, those hypotheses become less reasonable.

Instead a multiple hypotheses approach, where different hypotheses not only provide different response curve predictions, but also have different views about the validity of observations, can be employed. By assigning hypotheses a confidence based on how well they represent the observations, all hypotheses can be maintained so as to allow hypotheses that fail to match some early observations, but succeed to match later observations, to recover and become more confident.

Hypotheses are generated as follows. Given a number of previous observations, a set of hypotheses would have been generated with splines with differing smoothing parameters. On obtaining a new experimental observation, the previous hypotheses are checked against the observation to determine whether or not they are in agreement. If the observation lies outside of the 95% confidence interval of a particular hypothesis, h_{original}, the observation is declared as being in disagreement with the hypothesis. To handle the problem of determining whether it is the hypothesis or the observation that is invalid, the disagreeing
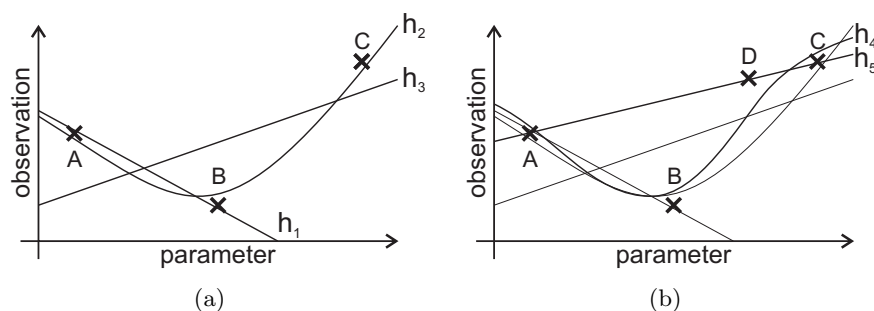
Fig. 2. Validity of observations affecting hypothesis proposal. Hypotheses (lines) are formed after observations (crosses) are obtained. In (a) h_1 is formed after observation A and B are obtained, however the effectiveness of the hypothesis is questioned by further observation C. Further hypotheses h_2 and h_3 consider A, B and C valid, but with differing noise estimates. In (b) D is obtained to test the discrepancy between h_1 and C, indicating that h_1 is unlikely. Hypothesis h_4 considers all observations valid, and hypothesis h_5 questions the validity of B.

observation is then considered both valid and invalid through two refined hypotheses, h_{valid} and h_{invalid} respectively. Both h_{valid} and h_{invalid} are based upon h_{original}, which is left unchanged within the current working set of hypotheses. The suspected erroneous observation is weighted accordingly in the smoothing spline calculation, with a high weight (currently set arbitrarily at 100) in h_{valid} and with a weight of zero in h_{invalid}. The high weight forms a hypothesis that believes the observation to be true and forces the spline to pass near to the observation. The zero weighted observation forms a hypothesis that considers the observation to be invalid and removes the observation from consideration when the spline is trained. Both h_{valid} and h_{invalid} are then added to the working set of hypotheses. With a working set of possible hypotheses, the next task is to determine the confidence of each hypothesis.

4.1   Evaluating a Hypothesis

As discussed previously, erroneous observations make hypothesis proposal a harder task, where any outliers could be due to an erroneous observation or an invalid hypothesis. Subsequently, methods for evaluating these hypotheses also need to take into consideration that the observations obtained may be erroneous. Forcing hypotheses to be evaluated against an erroneous observation, will not lead to the identification of hypotheses that fit the underlying phenomena well. For example, if observation B in Fig. 2 is invalid, but the hypotheses are evaluated against all observations, then hypotheses h_2, h_3 and h_4 will have higher confidences than the actual better fitting hypothesis h_5, as they pass closer to all observations. Ideally, the observations obtained after a hypothesis is created would be used solely for hypothesis evaluation, however we again have to

assume that these will be limited and that their validity will not be guaranteed. As such, using a standard mean squared error approach for evaluation appears not the best strategy. Instead hypotheses should be allowed to ignore a certain amount of selected observations without penalty.

A prediction of the likely percentage error may be determined either through previous experience with the experimental hardware, or through selecting a worst case value that if surpassed, would make the outcome of the experimentation not credible. Using this prediction of the percentage error, we can determine the number of observations a hypothesis can disregard without penalty. The observations a hypothesis may want to disregard are the observations that the hypothesis determines are invalid, so that the hypothesis is not evaluated on its ability to match observations it hypothesises are erroneous. Next a prototype equation for evaluating a hypothesis is considered.

The current prototype evaluation metric revises an existing metric for a evaluating smoothing spline [16], but uses only the observations that trained the hypothesis, so as to act as an overfitting test that gives a high value for a hypothesis that does not overfit the data:

$$
\upsilon(h) = \left(1 + \frac{\sum\limits_{i=1}^{n} \bar{w}_i \left(y_i - \hat{h}(x_i) \right)^2}{\bar{n} \left(1 - \frac{\text{tr}(\mathbf{A})}{\bar{n}} \right)^2} \right)^{-1}
\tag{2}
$$

where $\bar{w}$ is 0 when the observation is declared invalid by the hypothesis and 1 otherwise, $\hat{h}(x)$ is the prediction of a hypothesis for the experiment parameter and y is the actual experimental observation for that same experiment, $\bar{n}$ is the number of observations the hypothesis has declared valid from the training set and $\mathbf{A}$ is the hat matrix from of the smoothing spline.

Those observations not used to train the hypothesis evaluate the hypothesis in a mean squared error approach, presented here inverted to provide a confidence between 0 and 1:

$$
\gamma(h) = \frac{\bar{m}^2}{\bar{m}^2 + \sum\limits_{i=1}^{m} \bar{w}_i \left(\bar{y}_i - \hat{h}(\bar{x}_i) \right)^2}
\tag{3}
$$

where $\bar{m}$ is the number of test observations the hypothesis believes to be valid, $\bar{x}$ and $\bar{y}$ in this instance represent the test experiment-observation pairs, and $\bar{w}$ is 1 if the hypothesis believes the test observation to be valid, 0 otherwise. The confidence is currently defined as:

$$
C(h) = \left(\upsilon(h) + \gamma(h) \right) e(h)
\tag{4}
$$

where $e(h)$ is the penalty term for a hypothesis declaring greater than a given percentage p of observations as erroneous:

$$
e(h) = \begin{cases} \frac{1}{1 + (\bar{n} + \bar{m}) - p(n + m)}, & \text{if } (\bar{n} + \bar{m}) > p(n + m) \\ 1, & \text{otherwise} \end{cases}
\tag{5}
$$

5 Experiment Selection

Experiment selection, the active learning component of a artificial experimenter, should look to explore the experimental parameter space to discover behaviours not captured by the hypotheses, whilst also looking to find evidence to discriminate between competing hypotheses. The experiment selection algorithm should automatically balance between exploration and gaining information to elucidate differences between the hypotheses. This trade-off is described as a exploration-exploitation trade-off [17]. Hypotheses using the above hypothesis proposal scheme will mostly disagree when there are differing views in the validity of observations. In the case that an observation has caused a difference due to the observation being invalid, the experiment selection method should investigate the difference, realise it was caused by an error and then continue to search elsewhere. However, if the difference is caused by hypotheses failing to model the underlying behaviour and the observations are valid, further experimentation should be performed to capture the behaviour.

The prototype approach uses a strategy of placing experiments maximally away from previous experiments as the exploration strategy. For each proposed experiment, the minimum distance to any other previously performed experiment using a Euclidean distance function is calculated:

$$\zeta(\mathbf{x}) = \min_{\mathbf{x}' \in \mathbf{X}} |\mathbf{x} - \mathbf{x}'| \tag{6}$$

For determining discrepancy, the prototype approach looks to place experiments where the variance in hypotheses predictions is maximal, similar to that used, albeit unsuccessfully, in [10]. The unsuccessfulness of the approach described in [10] appears in part due to the sole use of a variance reduction strategy, the low number of hypotheses contemplated in parallel, and the polynomial kernel functions used. Here the variance calculation is weighted by the confidence of each hypothesis, so that weak hypotheses do not overly influence the decision:

$$\xi(\mathbf{x}) = k \sum_{i=1}^{N} C(h_i) \left(\hat{h}_i(\mathbf{x}) - \mu^* \right)^2 \tag{7}$$

To link the exploration and exploitation strategies, this approach sums the normalised values of the evaluations in (6) and (7). It may be worthwhile to develop more sophisticated management of exploration and exploitation, however the approach here is able to demonstrate the desired behaviour. The following is the prototype experiment selection strategy, where hyperparameter Γ controls the preference of exploration over exploitation:

$$\mathbf{x}_{\text{perform}} = \max_{\mathbf{x} \in \mathbf{X}} \left(\frac{\Gamma \zeta(\mathbf{x})}{\max_{\mathbf{t} \in \mathbf{X}} \zeta(\mathbf{t})} + \frac{(1 - \Gamma) \xi(\mathbf{x})}{\max_{\mathbf{t} \in \mathbf{X}} \xi(\mathbf{t})} \right) \tag{8}$$

6   Preliminary Results

To evaluate the approach of hypothesis management and experiment selection algorithms, we consider the number of experiments required to be performed in order to create a good representation of the phenomenon. Additionally we consider how different experiment selection strategies alter the predicted response behaviour under investigation. To evaluate this, we simulate experiments using a target function to return observations from requested experiment parameters. These target functions are designed to be general yet representative of possible responses from enzyme interaction experiments, where we would expect continuous nonmonotonic behaviours, with additional phase changes in some circumstances [3]. The model applies additive Gaussian noise $N(0, 0.5^2)$ to all observations (ϵ in (1)), and has a 5% rate of generating observations distorted by shock noise (ϕ in (1)). Independent variable noise (δ in (1)) has been set to 0, as in early trials altering this value had surprisingly little effect on the abililty of the approach demonstrated here. The hypotheses are restricted to choose λ values of 0.01, 0.001 and 0.0005, determined a priori for use with coded independent variables, as they give good flexibility to the spline.

In Fig. 3 we demonstrate the placement of experiments for a simulated underlying phenomena using the experiment selection algorithm in (8) with $\Gamma = 0.5$. For clarity we separate here the issue of outlying observations caused by the hypothesis being incorrect, and outlying observations caused by the observations being erroneous. In Fig. 3(a) and (b) we develop situations where observations will appear as outliers, but are actually true representations of the underlying phenomena, achieved through the use of a discontinuity between two distinct behaviours. The discontinuity provides observations that are close to each other in the experiment parameter space, yet are significantly different in their observed values, which will yield observations that appear to disagree, but are actually correct. In the example given, when the 9th experiment is performed, the observation will appear to not fit the current hypotheses, as the 1st and 5th observations either side of the 9th observation gave lower observation values.

In Fig. 3(a), a good representation of the underlying phenomena is identified by the approach after 10 experiments, using the confidence metric in (4). In this case the 5th and 9th observations have been automatically weighted, as the observations did not match the predictions of the previous hypotheses. The effect of this weighting is to force the hypothesis to pass through the transition region. Without the weightings, the smoothing spline would produce a hypothesis that averaged through the data points in that region. In Fig. 3(b), the majority of the additional data points obtained after Fig. 3(a) are in the region of the discontinuity. Here the discontinuity produces a large number of hypotheses that have differing views of the observations found in the region of the discontinuity. The exploitation scores of proposed experiments in this region will therefore be high, causing more experiments to be performed in this region. There is still however some exploration occurring as can be seen by observation 12.

In Fig. 3(c) and (d) the case of erroneous observations is considered. Here we consider a simpler underlying phenomena similar to that found in [9] and [10],
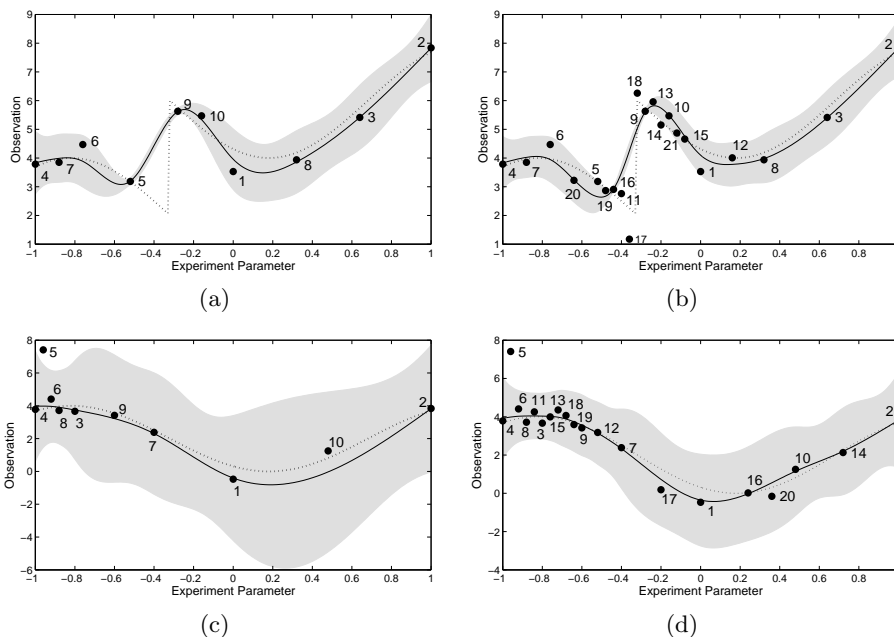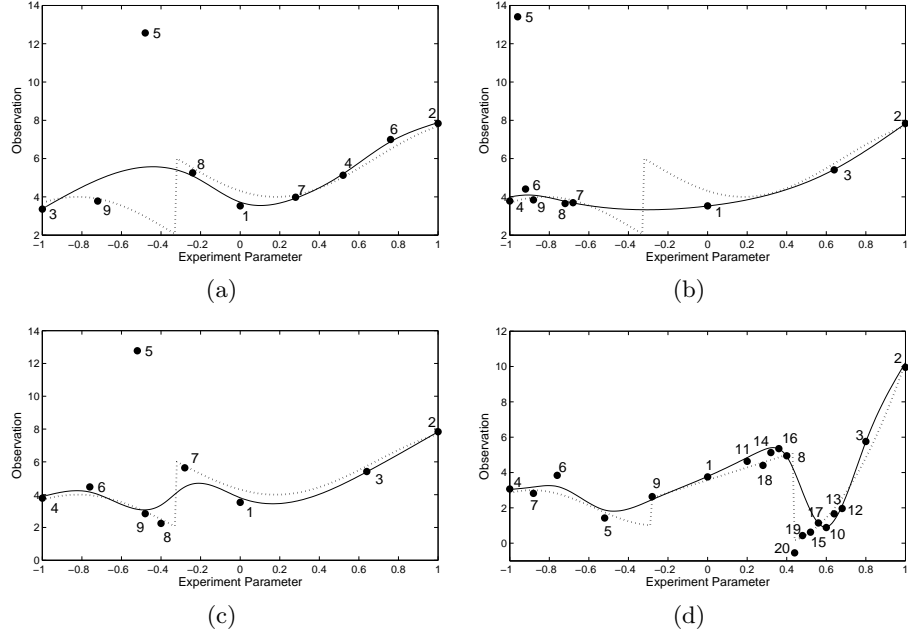
Fig. 3. Most confident hypothesis produced by the artificial experimenter for two target functions. Shown is the most confident hypothesis (solid line) with error bars shaded, underlying phenomena (dotted line) and observations (dots) numbered in order performed. The presence of a discontinuity (a and b) or an erroneous observation (c and d) causes additional experiments to be performed near those features, whilst also ensuring the entire parameter space is explored.

but here the 5th observation provided was an erroneous observation. In this case, there is continued exploration of the space after the 5th observation is obtained. With observations 6, 8 and 11 not supporting the 5th observation, there are fewer confident alternative hypotheses in the region of the erroneous observation, compared to the region of discontinuity in Fig. 3(a) and (b), therefore there is less focus on exploitation promoting experiments.

In Fig. 4 we demonstrate problems caused by using alternate active search strategies. Figure 4(a) considers a fully explorative search similar to that of reducing error bar uncertainty, where the erroneous observation causes part of the phenomena to be missed. In Fig. 4(b) a fully exploitative strategy similar to that used in [10] results in experiments being located near the erroneous observation that has caused the hypotheses to differ. Whilst Fig. 4(c) and (d) demonstrate how the approach described here is able to form reasonable predictions of the underlying phenomena after a small number of experiments.

Finally in Fig. 5, we compare the reduction in error between the most confident hypothesis and the true underlying behaviour, for the present active experiment selection technique and a passive technique. In a passive technique,

Fig. 4. Effect of different experiment proposal strategies. Shown is the most confident hypothesis representation (solid line) of a simulated underlying phenomena (dotted line) with observations (dots) numbered in the order performed. An erroneous observation causes the explorative search (a) and exploitative search (b) to fail to characterise the target function. In (c and d) the mixed strategy is shown for two different underlying phenomena and shock noise in (c).

the artificial experimenter cannot select the next experiment to perform, in this case the learner is presented with observations from random experiments. The underlying behaviour used in Fig. 4(a-c) is again used with Gaussian noise $N(0, 0.5^2)$. In Fig. 5(a) there is no shock noise and in Fig. 5(b) any 1 of the 20 experiments can be erroneous. Here we demonstrate that in both scenarios, the present experiment selection technique outperforms the passive technique. The most significant advantage of the present approach is in the early stages of experimentation, which is ideal in a situation where we may have only at most 10 experiments per dimension. As expected, performance is worse when shock noise is applied. However, as an erroneous observation can occur at any stage of experimentation and is distorted by Gaussian noise $N(10, 1)$, those runs where the erroneous observation occurs in the first few experiments will suffer more than those where the erroneous observation is later, causing larger error. In this instance the random technique performs far worse, with the average mean squared error being 30 compared to just under 5 for the active strategy, as shown in the inset of Fig. 5(b). Regardless of when the erroneous observation occurs, the results demonstrate that the hypothesis management technique recovers to
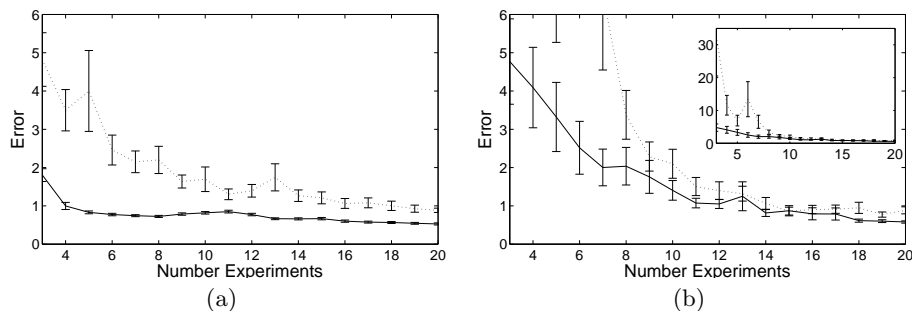
Fig. 5. Error reduction between most confident hypothesis and the target function over a number experiments. Shown is the average error after 100 trials for the active strategy (solid line) and passive strategy (dotted line) for the single discontinuity behaviour, with no shock noise in (a) and 1 of the 20 experiments in each trial is erroneous in (b).

provide reasonable hypotheses. However, by employing the active strategy, this recovery can occur quicker than through random experiment selection.

7 Discussion

The union of machine learning and automated laboratory hardware can allow for effective investigation of complex experiment parameter spaces. In particular the lab-on-chip automated platform being developed in parallel will significantly reduce the amount of chemical resources required per experiment [4]. Whilst the artificial experimenter utilises relatively cheap computational resources, to gain as much information as possible from small sets of observations.

In situations where hypotheses cannot correctly specify the underlying behaviour and high noise exists, artificial experimenters with few observations available can become misled, resulting in poor representation of the underlying phenomena. However, we have shown that by extending a variance based approach to better manage the exploration-exploitation trade-off and considering a larger corpus of possible hypotheses, that such techniques can provide a significant benefit to experimentation with limited resources. This proof-of-principle technique is demonstrated to work in scenarios where existing machine learning techniques will struggle, and that response behaviours can be characterised with a very small number of observations. However, we believe improvements can be made in the hypothesis evaluation and experiment selection techniques.

Currently the artificial experimenter is designed to build models of behaviours it identifies. However, in the future as we become more accustomed to the information processing capabilities of enzymes, we may wish to design target behaviours we require and then inspect the biological system to see if and where they exist. By modifying the kernel functions used in the hypotheses to those that describe the behaviours required, the present approach can be modified to

allow targeted experimentation that searches for particular behaviours, whilst still maintaining the multiple hypotheses benefits of the present approach. Overall, the purpose of this autonomous experimentation machine is to allow the human scientist to redirect their time from monotonous experimentation tasks, to determining computational functionality from the enzyme behaviours identified.

References

1. de Silva, A.P., Uchiyama, S.: Molecular logic and computing. Nature Nanotechnology **2** (2007) 399–410
2. Seelig, G., Soloveichik, D., Zhang, D.Y., Winfree, E.: Enzyme-free nucleic acid logic circuits. Science **314** (2006) 1585–1588
3. Zauner, K.-P., Conrad, M.: Enzymatic computing. Biotechnol. Prog. **17** (2001) 553–559
4. Jones, G., Lovell, C., Morgan, H., Zauner, K.-P.: Characterising enzymes for information processing: Microfluidics for autonomous experimentation. In: 9th International Conference on Unconventional Computation (2010)
5. Whitesides, G.M.: The origins and future of microfluidics. Nature **442** (July 27th 2006) 368–373
6. MacKay, D.J.C.: Information–based objective functions for active data selection. Neural Computation **4** (1992) 589–603
7. Cohn, D.A., Ghahramani, Z., Jordan, M.I.: Active learning with statistical models. Journal of Artificial Intelligence Research **4** (1996) 129–145
8. Freund, Y., Seung, H.S., Shamir, E., Tishby, N.: Selective sampling using the query by committee algorithm. Machine Learning **28** (1997) 133–168
9. Sugiyama, M., Rubens, N.: Active learning with model selection in linear regression. In: SIAM International Conference on Data Mining. (2008) 518–529
10. Burbidge, R., Rowland, J.J., King, R.D.: Active learning for regression based on query by committee. In: Intelligent Data Engineering and Automated Learning
11. Kulkarni, D., Simon, H.A.: Experimentation in machine discovery. In Shrager, J., Langley, P., eds.: Computational Models of Scientific Discovery and Theory Formation. Morgan Kaufmann Publishers, San Mateo, CA (1990) 255–273
12. Pfaffmann, J.O., Zauner, K.P.: Scouting context-sensitive components. In: The Third NASA/DoD Workshop on Evolvable Hardware—EH-2001. (2001) 14–20
13. King, R.D., Whelan, K.E., Jones, F.M., Reiser, P.G.K., Bryant, C.H., Muggleton, S.H., Kell, D.B., Oliver, S.G.: Functional genomic hypothesis generation and experimentation by a robot scientist. Nature **427** (2004) 247–252
14. Żytkow, J.M., Zhu, J., Hussam, A.: Automated discovery in a chemistry laboratory. In: Proceedings of the 8th National Conference on Artificial Intelligence, Boston, MA, AAAI Press / MIT Press (1990) 889–894
15. Schmidt, M., Lipson, H.: Distilling free-form natural laws from experimental data. Science **324**(5923) (2009) 81–85
16. Wahba, G.: Bayesian "confidence intervals" for the cross-validated smoothing spline. J. R. Statist. Soc. B **45**(1) (1983) 133–150
17. Auer, P.: Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research **3** (2002) 397–422