Opportunities for formal methods in the analysis of pervasive adaptation

Vladimiro Sassone

ECS, University of Southampton

Brussels, 26 January 2007

Conclusion

- Without formal models, there is no dealing with non-trivial notions.
  - True in all realms of human activity: complex artifacts require blueprints of all sorts;
  - Yet, mysteriously some believe one could design, build, maintain, service, upgrade, evolve, . . . systems of the complexity of ecologies just by implementing the right support.

    But to build a house is not sufficient to design good bricks . . .
- Without practice, there is no useful formal method.
  - models conceived in abstract are not abstract models, just cathedrals in the desert.
- Without new tools, there is no realistic formal model of adaptation.
  - this is a hard concept, that includes research areas big per se that we cannot handle (eg complex systems, emergent behaviour, . . .)

Conclusion

- Without formal models, there is no dealing with non-trivial notions.
  - True in all realms of human activity: complex artifacts require blueprints of all sorts;
  - Yet, mysteriously some believe one could design, build, maintain, service, upgrade, evolve, . . . systems of the complexity of ecologies just by implementing the right support.

    But to build a house is not sufficient to design good bricks . . .
- Without practice, there is no useful formal method.
  - models conceived in abstract are not abstract models, just cathedrals in the desert.
- Without new tools, there is no realistic formal model of adaptation.
  - this is a hard concept, that includes research areas big per se that we cannot handle (eg complex systems, emergent behaviour, . . .)

Conclusion

- Without formal models, there is no dealing with non-trivial notions.
  - True in all realms of human activity: complex artifacts require blueprints of all sorts;
  - Yet, mysteriously some believe one could design, build, maintain, service, upgrade, evolve, ... systems of the complexity of ecologies just by implementing the right support.

    But to build a house is not sufficient to design good bricks ...
- Without practice, there is no useful formal method.
  - models conceived in abstract are not abstract models, just cathedrals in the desert.
- Without new tools, there is no realistic formal model of adaptation.
  - this is a hard concept, that includes research areas big per se that we cannot handle (eg complex systems, emergent behaviour, ...)

Conclusion

- Without formal models, there is no dealing with non-trivial notions.
  - True in all realms of human activity: complex artifacts require blueprints of all sorts;
  - Yet, mysteriously some believe one could design, build, maintain, service, upgrade, evolve, . . . systems of the complexity of ecologies just by implementing the right support.

    But to build a house is not sufficient to design good bricks . . .
- Without practice, there is no useful formal method.
  - models conceived in abstract are not abstract models, just cathedrals in the desert.
- Without new tools, there is no realistic formal model of adaptation.
  - this is a hard concept, that includes research areas big per se that we cannot handle (eg complex systems, emergent behaviour, . . .)

Conclusion

- Without formal models, there is no dealing with non-trivial notions.
  - True in all realms of human activity: complex artifacts require blueprints of all sorts;
  - Yet, mysteriously some believe one could design, build, maintain, service, upgrade, evolve, . . . systems of the complexity of ecologies just by implementing the right support.

    But to build a house is not sufficient to design good bricks . . .
- Without practice, there is no useful formal method.
  - models conceived in abstract are not abstract models, just cathedrals in the desert.
- Without new tools, there is no realistic formal model of adaptation.
  - this is a hard concept, that includes research areas big per se that we cannot handle (eg complex systems, emergent behaviour, . . .)

OK for models, but why formal?

- Formal models cover the entire spectrum of development:
  - specification, validation, implementation, verification, ...
- Emerging behaviour typical of adaptation needs statements of 'fitness for purpose':
  - need languages to specify those and techniques to verify them
- Critical applications need certification of key properties:
  - safety guarantees, performance standards, quality indicators, ...

OK for models, but why formal?

- Formal models cover the entire spectrum of development:
  - specification, validation, implementation, verification, . . .
- Emerging behaviour typical of adaptation needs statements of 'fitness for purpose':
  - need languages to specify those and techniques to verify them
- Critical applications need certification of key properties:
  - safety guarantees, performance standards, quality indicators, . . .

OK for models, but why formal?

- Formal models cover the entire spectrum of development:
  - specification, validation, implementation, verification, ...
- Emerging behaviour typical of adaptation needs statements of ‘fitness for purpose’:
  - need languages to specify those and techniques to verify them
- Critical applications need certification of key properties:
  - safety guarantees, performance standards, quality indicators, ...

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, and what's so unique about adaptation?

- work with changing and unpredictable environments;
- work with lack of relevant information;
- work with tight resource constraints;
- work with internal hypothesis built from observations;
- work by making hypothesis and testing them against observations of the world;
- and, most importantly,

goal-orientation, autonomy, re- and pro-activeness.

OK that, but what tools for adaptation?

processes $\mapsto$ **agents**??

goal-oriented

autonomous

reactive

proactive

Reject intelligence-based and bio-inspired approaches per se.
Accept when backed by solid mathematical tools, like eg:
data mining, learning, evolutionary and genetic algorithms, Bayesian learning, . . .

Existing approaches: game theory and mechanism design; negotiation (because often the environment is just others like you); evolutionary, probabilistic.

NEXT: A typical example of adaptation of the simplest form.

OK that, but what tools for adaptation?

processes $\mapsto$ **agents**??

goal-oriented

autonomous

reactive

proactive

Reject intelligence-based and bio-inspired approaches per se.
Accept when backed by solid mathematical tools, like eg:
data mining, learning, evolutionary and genetic algorithms, Bayesian learning, . . .

Existing approaches: game theory and mechanism design; negotiation (because often the environment is just others like you); evolutionary, probabilistic.

NEXT: A typical example of adaptation of the simplest form.

Simple Adaptation via Bayesian Learning

The model λ_θ:

- Each principal p behaves in each interaction according to a fixed and independent probability θ_p of 'success' (and therefore $1 - \theta_p$ of 'failure').

The framework:

- Interface (Trust computation algorithm, $\mathcal{A}$):
  - Input: A sequence $h = x_1 x_2 \cdots x_n$ for $n \geq 0$ and $x_i \in \{\mathbf{s}, \mathbf{f}\}$.
  - Output: A probability distribution $\pi : \{\mathbf{s}, \mathbf{f}\} \rightarrow [0, 1]$.
- Goal:
  - Output π approximates $(\theta_p, 1 - \theta_p)$ as well as possible, under the hypothesis that input h is the outcome of interactions with p.

Simple Adaptation via Bayesian Learning

The model λ_θ:

- Each principal p behaves in each interaction according to a fixed and independent probability θ_p of 'success' (and therefore $1 - \theta_p$ of 'failure').

The framework:

- Interface (Trust computation algorithm, $\mathcal{A}$):
  - Input: A sequence $h = x_1 x_2 \cdots x_n$ for $n \geq 0$ and $x_i \in \{\mathbf{s}, \mathbf{f}\}$.
  - Output: A probability distribution $\pi : \{\mathbf{s}, \mathbf{f}\} \to [0, 1]$.
- Goal:
  - Output π approximates $(\theta_p, 1 - \theta_p)$ as well as possible, under the hypothesis that input h is the outcome of interactions with p.

Maximum likelihood (Despotovic and Aberer)

Trust computation $\mathcal{A}_0$

$$\mathcal{A}_0(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h)}{|h|} \qquad \mathcal{A}_0(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h)}{|h|}$$

$$\mathrm{N}_x(h) = \text{“number of } x\text{'s in } h\text{”}$$

Bayesian analysis inspired by $\boldsymbol{\lambda}_{\boldsymbol{\beta}}$ model: $f(\theta \mid \alpha\, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

Properties:

- Well defined semantics: $\mathcal{A}_0(\mathbf{s} \mid h)$ is interpreted as a *probability* of success in the next interaction.

Maximum likelihood (Despotovic and Aberer)

Trust computation $\mathcal{A}_0$

$$\mathcal{A}_0(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h)}{|h|} \qquad \mathcal{A}_0(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h)}{|h|}$$

$$\mathrm{N}_x(h) = \text{“number of } x\text{'s in } h\text{”}$$

Bayesian analysis inspired by $\boldsymbol{\lambda}_\beta$ model: $f(\theta \mid \alpha\,\beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

Properties:

- Well defined semantics: $\mathcal{A}_0(\mathbf{s} \mid h)$ is interpreted as a *probability* of success in the next interaction.

Beta models (Mui et al)

Even more tightly inspired by Bayesian analysis and by λ_β

Trust computation $\mathcal{A}_1$

$$\mathcal{A}_1(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + 1}{|h| + 2} \qquad \mathcal{A}_1(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + 1}{|h| + 2}$$

$$\mathrm{N}_x(h) = \text{“number of } x\text{'s in } h\text{”}$$

Properties:

- Well defined semantics: $\mathcal{A}_1(\mathbf{s} \mid h)$ is interpreted as a *probability* of success in the next interaction.

Beta models (Mui et al)

Even more tightly inspired by Bayesian analysis and by λ_β

Trust computation $\mathcal{A}_1$

$$\mathcal{A}_1(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + 1}{|h| + 2} \qquad \mathcal{A}_1(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + 1}{|h| + 2}$$

$$\mathrm{N}_x(h) = \text{“number of } x\text{'s in } h\text{”}$$

Properties:

- Well defined semantics: $\mathcal{A}_1(\mathbf{s} \mid h)$ is interpreted as a *probability* of success in the next interaction.

Cross entropy

An information-theoretic “distance” on distributions

Cross entropy of distributions $\mathbf{p}, \mathbf{q} : \{o_1, \ldots, o_m\} \to [0, 1]$.

$$\mathrm{D}(\mathbf{p} \,||\, \mathbf{q}) = \sum_{i=1}^{m} \mathbf{p}(o_i) \cdot \log\big(\mathbf{p}(o_i)/\mathbf{q}(o_i)\big)$$

It holds $0 \leq \mathrm{D}(\mathbf{p} \,||\, \mathbf{q}) \leq \infty$, and $\mathrm{D}(\mathbf{p} \,||\, \mathbf{q}) = 0$ iff $\mathbf{p} = \mathbf{q}$.

- Established measure in statistics for comparing distributions.
- Information-theoretic: the average amount of information discriminating $\mathbf{p}$ from $\mathbf{q}$.

Cross entropy

An information-theoretic “distance” on distributions

Cross entropy of distributions $\mathbf{p}, \mathbf{q} : \{o_1, \ldots, o_m\} \to [0, 1]$.

$$\mathrm{D}(\mathbf{p} \mid\mid \mathbf{q}) = \sum_{i=1}^{m} \mathbf{p}(o_i) \cdot \log\big(\mathbf{p}(o_i)/\mathbf{q}(o_i)\big)$$

It holds $0 \leq \mathrm{D}(\mathbf{p} \mid\mid \mathbf{q}) \leq \infty$, and $\mathrm{D}(\mathbf{p} \mid\mid \mathbf{q}) = 0$ iff $\mathbf{p} = \mathbf{q}$.

- Established measure in statistics for comparing distributions.
- Information-theoretic: the average amount of information discriminating $\mathbf{p}$ from $\mathbf{q}$.

Expected cross entropy

A measure on probabilistic trust algorithms

- Goal of a probabilistic trust algorithm $\mathcal{A}$: given a history $\mathbf{X}$, approximate a distribution on the outcomes $O = \{o_1, \ldots, o_m\}$.
- Different histories $\mathbf{X}$ result in different output distributions $\mathcal{A}(\cdot \mid \mathbf{X})$.

Expected cross entropy from λ to $\mathcal{A}$

$$\mathrm{ED}^n(\lambda \,||\, \mathcal{A}) = \sum_{\mathbf{X} \in O^n} \mathit{Prob}(\mathbf{X} \mid \lambda) \cdot \mathrm{D}(\mathit{Prob}(\cdot \mid \mathbf{X}\, \lambda) \,||\, \mathcal{A}(\cdot \mid \mathbf{X}))$$

Expected cross entropy

A measure on probabilistic trust algorithms

- Goal of a probabilistic trust algorithm $\mathcal{A}$: given a history $\mathbf{X}$, approximate a distribution on the outcomes $O = \{o_1, \ldots, o_m\}$.
- Different histories $\mathbf{X}$ result in different output distributions $\mathcal{A}(\cdot \mid \mathbf{X})$.

Expected cross entropy from $\boldsymbol{\lambda}$ to $\mathcal{A}$

$$\mathrm{ED}^n(\boldsymbol{\lambda} \,||\, \mathcal{A}) = \sum_{\mathbf{X} \in O^n} Prob(\mathbf{X} \mid \boldsymbol{\lambda}) \cdot \mathrm{D}(Prob(\cdot \mid \mathbf{X}\,\boldsymbol{\lambda}) \,||\, \mathcal{A}(\cdot \mid \mathbf{X}))$$

An application of cross entropy (1/2)

Consider the beta model $\boldsymbol{\lambda}_\beta$ and the algorithms $\mathcal{A}_0$ of maximum likelihood (Despotovic et al.) and $\mathcal{A}_1$ beta (Mui et al.).

Theorem

If $\theta = 0$ or $\theta = 1$ then $\mathcal{A}_0$ computes the exact distribution, whereas $\mathcal{A}_1$ does not. That is, for all $n > 0$ we have:

$$\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_0) = 0 < \mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_1)$$

If $0 < \theta < 1$, then $\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_0) = \infty$, and $\mathcal{A}_1$ is always better.

An application of cross entropy (1/2)

Consider the beta model $\boldsymbol{\lambda}_\beta$ and the algorithms $\mathcal{A}_0$ of maximum likelihood (Despotovic et al.) and $\mathcal{A}_1$ beta (Mui et al.).

Theorem

If $\theta = 0$ or $\theta = 1$ then $\mathcal{A}_0$ computes the exact distribution, whereas $\mathcal{A}_1$ does not. That is, for all $n > 0$ we have:

$$\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_0) = 0 < \mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_1)$$

If $0 < \theta < 1$, then $\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_0) = \infty$, and $\mathcal{A}_1$ is always better.

An application of cross entropy (2/2)

A parametric algorithm $\mathcal{A}_\epsilon$

$$\mathcal{A}_\epsilon(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + \epsilon}{|h| + 2\epsilon}, \qquad \mathcal{A}_\epsilon(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + \epsilon}{|h| + 2\epsilon}$$

Theorem

For any $\theta \in [0,1]$, $\theta \neq 1/2$ there exists $\bar{\epsilon} \in [0,\infty)$ that minimises $\mathrm{ED}^n(\lambda_\beta \,||\, \mathcal{A}_\epsilon)$, simultaneously for all n.

An application of cross entropy (2/2)

A parametric algorithm $\mathcal{A}_\epsilon$

$$\mathcal{A}_\epsilon(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + \epsilon}{|h| + 2\epsilon}, \qquad \mathcal{A}_\epsilon(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + \epsilon}{|h| + 2\epsilon}$$

Theorem

For any $\theta \in [0,1]$, $\theta \neq 1/2$ there exists $\bar{\epsilon} \in [0,\infty)$ that minimises $\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_\epsilon)$, simultaneously for all n.

An application of cross entropy (2/2)

A parametric algorithm $\mathcal{A}_\epsilon$

$$\mathcal{A}_\epsilon(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + \epsilon}{|h| + 2\epsilon}, \qquad \mathcal{A}_\epsilon(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + \epsilon}{|h| + 2\epsilon}$$

Theorem

For any $\theta \in [0, 1]$, $\theta \neq 1/2$ there exists $\bar{\epsilon} \in [0, \infty)$ that minimises $\mathrm{ED}^n(\boldsymbol{\lambda}_\beta \mid\mid \mathcal{A}_\epsilon)$, simultaneously for all n.

That is, unless behaviour is completely unbiased, there exists a unique best $\mathcal{A}_\epsilon$ algorithm that for all n outperforms all the others.
If $\theta = 1/2$, the larger the ϵ, the better.

An application of cross entropy (2/2)

A parametric algorithm $\mathcal{A}_\epsilon$

$$\mathcal{A}_\epsilon(\mathbf{s} \mid h) = \frac{\mathrm{N}_{\mathbf{s}}(h) + \epsilon}{|h| + 2\epsilon}, \qquad \mathcal{A}_\epsilon(\mathbf{f} \mid h) = \frac{\mathrm{N}_{\mathbf{f}}(h) + \epsilon}{|h| + 2\epsilon}$$

Theorem

For any $\theta \in [0,1]$, $\theta \neq 1/2$ there exists $\bar{\epsilon} \in [0, \infty)$ that minimises $\mathrm{ED}^n(\boldsymbol{\lambda}_{\boldsymbol{\beta}} \mid\mid \mathcal{A}_\epsilon)$, simultaneously for all n.

- Algorithm $\mathcal{A}_0$ is optimal for $\theta = 0$ and for $\theta = 1$.
- Algorithm $\mathcal{A}_1$ is optimal for $\theta = \frac{1}{2} \pm \frac{1}{\sqrt{12}}$.

A trust model based on event structures

Move from $O = \{\mathbf{s}, \mathbf{f}\}$ to complex outcomes

Interactions and protocols

- At an abstract level, entities in a distributed system interact according to protocols;
- Information about an external entity is just information about (the outcome of) a number of (past) protocol runs with that entity.

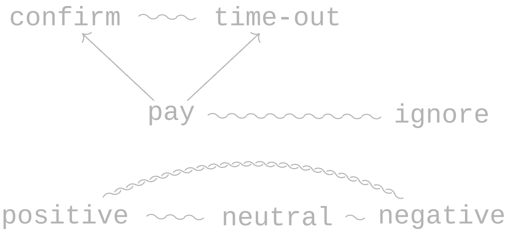
Events as model of information

- A protocol can be specified as a concurrent process, at different levels of abstractions.
- Event structures were invented to give formal semantics to truely concurrent processes, expressing “causation” and “conflict.”

A model for behavioural information

- $ES = (E, \leq, \#)$, with E a set of events, $\leq$ and $\#$ relations on E.
- Information about a session is a finite set of events $x \subseteq E$, called a configuration (which is ‘conflict-free’ and ‘causally-closed’).
- Information about several interactions is a sequence of outcomes $h = x_1 x_2 \cdots x_n \in \mathcal{C}^*_{ES}$, called a history.

eBay (simplified) example:

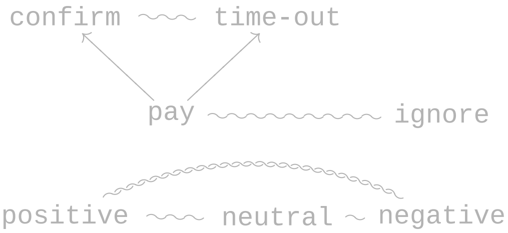


e.g., $h = \{\mathtt{pay}, \mathtt{confirm}, \mathtt{pos}\}\ \{\mathtt{pay}, \mathtt{confirm}, \mathtt{neu}\}\ \{\mathtt{pay}\}$

A model for behavioural information

- $ES = (E, \leq, \#)$, with E a set of events, $\leq$ and $\#$ relations on E.
- Information about a session is a finite set of events $x \subseteq E$, called a configuration (which is 'conflict-free' and 'causally-closed').
- Information about several interactions is a sequence of outcomes $h = x_1 x_2 \cdots x_n \in \mathcal{C}_{ES}^*$, called a history.

eBay (simplified) example:

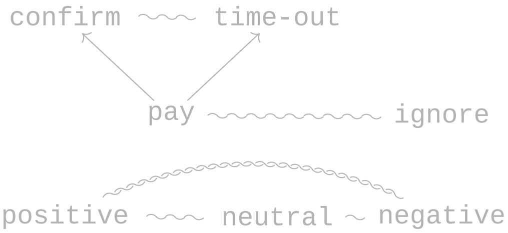


e.g., $h = \{\mathtt{pay}, \mathtt{confirm}, \mathtt{pos}\}\ \{\mathtt{pay}, \mathtt{confirm}, \mathtt{neu}\}\ \{\mathtt{pay}\}$

A model for behavioural information

- $ES = (E, \leq, \#)$, with E a set of events, $\leq$ and $\#$ relations on E.
- Information about a session is a finite set of events $x \subseteq E$, called a configuration (which is 'conflict-free' and 'causally-closed').
- Information about several interactions is a sequence of outcomes $h = x_1 x_2 \cdots x_n \in \mathcal{C}_{ES}^*$, called a history.

eBay (simplified) example:

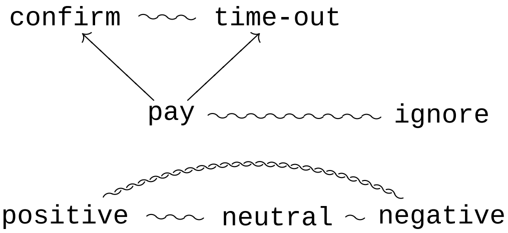


e.g., $h = \{\texttt{pay}, \texttt{confirm}, \texttt{pos}\}\ \{\texttt{pay}, \texttt{confirm}, \texttt{neu}\}\ \{\texttt{pay}\}$

A model for behavioural information

- $ES = (E, \leq, \#)$, with E a set of events, $\leq$ and $\#$ relations on E.
- Information about a session is a finite set of events $x \subseteq E$, called a configuration (which is 'conflict-free' and 'causally-closed').
- Information about several interactions is a sequence of outcomes $h = x_1 x_2 \cdots x_n \in \mathcal{C}_{ES}^*$, called a history.

eBay (simplified) example:

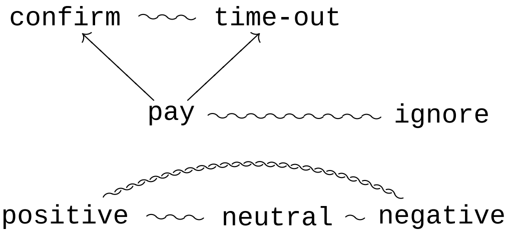


e.g., $h = \{\mathtt{pay}, \mathtt{confirm}, \mathtt{pos}\}\ \{\mathtt{pay}, \mathtt{confirm}, \mathtt{neu}\}\ \{\mathtt{pay}\}$

A model for behavioural information

- $ES = (E, \leq, \#)$, with E a set of events, $\leq$ and $\#$ relations on E.
- Information about a session is a finite set of events $x \subseteq E$, called a configuration (which is 'conflict-free' and 'causally-closed').
- Information about several interactions is a sequence of outcomes $h = x_1 x_2 \cdots x_n \in \mathcal{C}_{ES}^*$, called a history.

eBay (simplified) example:



e.g., $h = \{\texttt{pay}, \texttt{confirm}, \texttt{pos}\}\ \{\texttt{pay}, \texttt{confirm}, \texttt{neu}\}\ \{\texttt{pay}\}$

A bit of magic: the Dirichlet probability distribution

The Dirichlet family $\mathcal{D}(\boldsymbol{\Theta} \mid \boldsymbol{\alpha}) \propto \prod \Theta_1^{\alpha_1 - 1} \cdots \Theta_K^{\alpha_K - 1}$

Theorem

The Dirichlet family is a conjugate prior for multinomial trials. That is, if

- *$Prob[\boldsymbol{\Theta} \mid \lambda]$ is $\mathcal{D}(\boldsymbol{\Theta} \mid \alpha_1, ..., \alpha_K)$ and*
- *$Prob[\mathbf{X} \mid \boldsymbol{\Theta}\, \lambda]$ follows the law of multinomial trials $\Theta_1^{n_1} \cdots \Theta_K^{n_K}$,*

then $Prob[\boldsymbol{\Theta} \mid \mathbf{X}\, \lambda]$ is $\mathcal{D}(\boldsymbol{\Theta} \mid \alpha_1 + n_1, ..., \alpha_K + n_K)$ according to Bayes.

A bit of magic: the Dirichlet probability distribution

The Dirichlet family $\mathcal{D}(\mathbf{\Theta} \mid \boldsymbol{\alpha}) \propto \prod \Theta_1^{\alpha_1 - 1} \cdots \Theta_K^{\alpha_K - 1}$

Theorem

The Dirichlet family is a conjugate prior for multinomial trials. That is, if

- *$Prob[\mathbf{\Theta} \mid \boldsymbol{\lambda}]$ is $\mathcal{D}(\mathbf{\Theta} \mid \alpha_1, ..., \alpha_K)$ and*
- *$Prob[\mathbf{X} \mid \mathbf{\Theta}\, \boldsymbol{\lambda}]$ follows the law of multinomial trials $\Theta_1^{n_1} \cdots \Theta_K^{n_K}$,*

then $Prob[\mathbf{\Theta} \mid \mathbf{X}\, \boldsymbol{\lambda}]$ is $\mathcal{D}(\mathbf{\Theta} \mid \alpha_1 + n_1, ..., \alpha_K + n_K)$ according to Bayes.

Interpretation of results

As a result, we can lift the trust computational algorithms based on λ_β to our event-base models by replacing

Binomials (Bernoulli) trials	$\mapsto$	multinomial trials;
β-distribution	$\mapsto$	Dirichlet distribution.