

## Statistical relational learning of trust

Achim Rettinger · Matthias Nickles · Volker Tresp

Received: 1 June 2009 / Accepted: 1 May 2010 / Published online: 25 September 2010  
© The Author(s) 2010

**Abstract** The learning of trust and distrust is a crucial aspect of social interaction among autonomous, mentally-opaque agents. In this work, we address the learning of trust based on past observations and context information. We argue that from the truster's point of view trust is best expressed as one of several relations that exist between the agent to be trusted (trustee) and the state of the environment. Besides attributes expressing trustworthiness, additional relations might describe commitments made by the trustee with regard to the current situation, like: a seller offers a certain price for a specific product. We show how to implement and learn context-sensitive trust using statistical relational learning in form of a Dirichlet process mixture model called *Infinite Hidden Relational Trust Model* (IHRTM). The practicability and effectiveness of our approach is evaluated empirically on user-ratings gathered from eBay. Our results suggest that (i) the inherent clustering achieved in the algorithm allows the truster to characterize the structure of a trust-situation and provides meaningful trust assessments; (ii) utilizing the collaborative filtering effect associated with relational data does improve trust assessment performance; (iii) by learning faster and transferring knowledge more effectively we improve cold start performance and can cope better with dynamic behavior in open multiagent systems. The later is demonstrated with interactions recorded from a strategic two-player negotiation scenario.

---

Editors: S.V.N. Vishwanathan, Samuel Kaski, Jennifer Neville, and Stefan Wrobel.

A. Rettinger (✉)

Institute AIFB, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany  
e-mail: [rettinger@kit.edu](mailto:rettinger@kit.edu)

M. Nickles

Department of Computer Science, Technical University of Munich, 85748 Garching, Germany  
e-mail: [nickles@cs.tum.edu](mailto:nickles@cs.tum.edu)

V. Tresp

Corporate Technology, Siemens AG, 81739 Munich, Germany  
e-mail: [volker.tresp@siemens.com](mailto:volker.tresp@siemens.com)

**Keywords** Relational learning · Computational trust · Social computing · Infinite hidden relational models · Initial trust

## 1 Introduction

In open and dynamic distributed systems on the Internet, such as open auction platforms, tools and algorithms for the prediction of an actor's future behavior are of highest importance, since in many situations, the actors in such systems (humans as well as artificial agents and services) show a highly contingent behavior, and often it is not feasible to implement effective rules to enforce socially fair behavior as pursued using norms, mechanism design or preference aggregation. A potential solution to these problems is the transfer of the human notion of trust to a machine-computable model, realizing *computational trust*.

However, most current models for computational trust (see, e.g., Ramchurn et al. 2004 for a survey. More recent approaches are discussed in Sect. 5) lack the ability to take context sufficiently into account when trying to predict future behavior of interacting agents. Moreover, they are not able to transfer knowledge gained in a specific context to a related context. In other words, most existing approaches to trust in Computer Science are far away from the power of “human trust” and its main objective as a mechanism for contingency reduction in uncertain and unknown situations. It is even the essential property of most “trust situations” that the human (respectively the agent) does *not* have sufficient information that could directly be applied to assess a trust value. Instead, trust must be inferred from related contextual factors. We think that the lack of a certain basis for decision-making is *the* defining characteristic of trust and is not sufficiently taken into account in most current trust models. Consequently, approaches where trust can exclusively and more or less directly be based on the reputation calculated from recommendations or social *trust networks* (Golbeck and Hendler 2003) do not fully exploit the much richer notion of human trust (see Sect. 5). The same holds for cognitive and game theoretic models of trust based on computable incentives of the trustee or statistical models dependent on repetitive interactions in a restricted context-*independent* environment.

Based on those observations our objective is to relax existing restrictions of computational trust by trying to learn trust in a rich context-dependent relational environment: Modeling the environment *from the perspective of the truster*, two entities, both described by their respective attributes, constitute a trust situation: (i) the trustee and (ii) the state of the environment. Most importantly, both entities are interconnected by relational dependencies.

If the trustworthiness depends not only on the trustee but also on the state of the environment in which one needs to trust, the truster can make more precise decisions and can apply learned knowledge to a wider range of situation. For instance, a seller might be trustworthy if offering a specific product, but not when offering another product. Furthermore, in such a situation a relation like the price might help to assess trustworthiness while depending on a particular product and the seller at the same time. By taking all this into account, we can improve predictions, give more meaning to trust and at the same time—by generalizing from different contexts—increase learning efficiency.

In particular, the general requirements that we wish to meet are: Context sensitivity and trust transfer:

**Contextual information** that might be related to the trust decision to be made needs to be incorporated. This shall include attributes of the person one needs to trust, attributes of the external circumstances under which the trust decision is made, and actions and promises the person has given to seek one's confidence. Furthermore, specific trust values gained in a certain context need to be transferrable to new, unknown “trigger” situations.

**Multi-dimensionality** Most trust models assign a single trust value per agent. This ignores the fact that human trust decisions are made in relation to a whole spectrum of aspects (e.g., what a person is likely to do, such as the expected outcome of some information trading, even in the same context). For instance a certain information supplier agent might be trustworthy in terms of delivery date, but not in terms of information quality (e.g., precision, topicality, credibility . . .). Combining several trust related measures as in our approach is considerably much more flexible. In contrast, most existing approaches to trust still relate trust to “whole persons” only instead of their contextualized behavior.

At this, we focus on interaction-trust (i.e., (dis-)trust formed by agents during the course of an interaction regarding their opponents’ behavior) in order to tailor our model to the specifics of the probably most relevant application field for empirical trustability assessment.

In the following we show how to implement and learn context-sensitive relational trust using one specific statistical relational model. Our Infinite Hidden Relational Trust Model (IHRTM) is based on recently introduced *infinite relational models* (see Xu et al. 2006 and Kemp et al. 2006). The practicability and effectiveness of this approach is evaluated empirically on user-ratings gathered from eBay. Our results suggest that (i) the inherent clustering achieved in the algorithm allows the truster to characterize the structure of a trust-situation and provides meaningful trust assessments (see Sect. 4.2); (ii) utilizing the collaborative filtering effect associated with relational data does improve trust assessment performance (see Sect. 4.3); (iii) by learning faster and transferring knowledge more effectively we improve cold start performance and can cope better with dynamic behavior in open multiagent systems. The later is demonstrated with interactions recorded from a strategic two-player negotiation scenario (see Sect. 4.4).

The next section introduces the statistical relational representation used for context-dependent trust modeling accompanied by an intuitive illustration of modeling transactions and feedback on eBay. Section 3 describes the technical details and the inference algorithm used to calculate cluster assignments and trust values in the special case of using our IHRTM. Section 4 presents the experimental analysis on three different levels: First the clustering effect on the eBay data is evaluated, then the predictive performance is compared to propositional learning algorithms, and finally the learning efficiency is demonstrated on data from an automated negotiations scenario. Section 5 discusses related work, and Sect. 6 outlines future research directions and provides conclusions.

The research described in this paper builds upon and extends the work reported in a paper published in the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008).

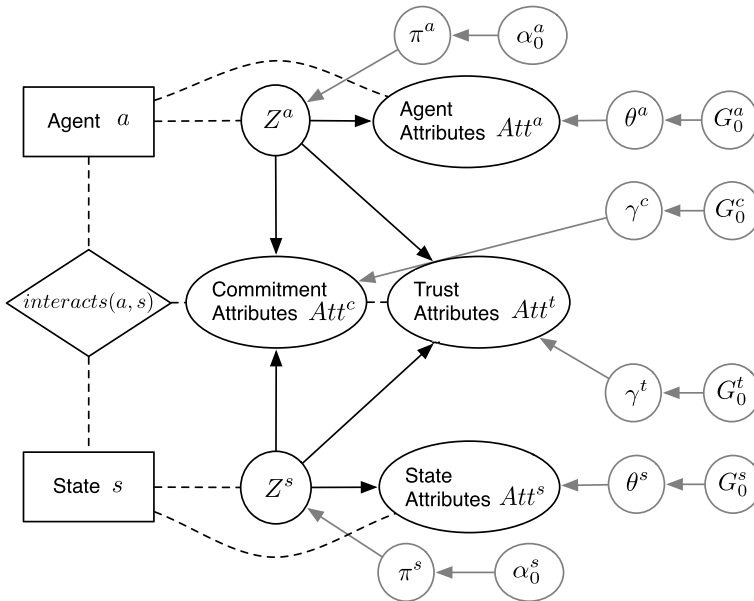
## 2 Model description

Relational models are an obvious formalization of requirements arising from the relational dependencies of entities in social, biological, physical and many other systems.

Our Infinite Hidden Relational Trust Model (IHRTM) consists of two entity classes: On the one hand the trustee-agent  $a$  and on the other hand specific elements of the state  $s$  of the environment. Both entities can be equipped with attributes  $Att^a$  and  $Att^s$ , respectively. The interdependencies are expressed as relation  $interacts(a, s)$  with attributes  $Att^c$  (commitment) and  $Att^t$  (trust).

The basic interaction-trust scenario thus consists of:

- An agent  $a$  (trustee) which is potentially willing to interact with the truster, and who is characterized by a set of observable attributes  $Att^a$ . An agent can be considered as a person



**Fig. 1** Graphical representation 1: DAPER model

or more general any instance that can be trusted, like a web service, an information source, a company, a brand, or an authority.

- A set of conditions of the environmental state  $s$  with corresponding attributes  $Att^s$ . An apparent condition would e.g. be the type of service provided by the trustee, for instance a specific merchandize or an information supply in case of information trading agents. Moreover this implies all external facts comprising this particular state like the truster’s own resources or the current market value of the merchandize in question.
- A relation  $interacts(a, s)$  with a set of relationship attributes  $Att^c$  and  $Att^t$ , capturing all negotiable interaction issues depending on specific conditions  $c \in C$ . Those attributes are separated into two different sets:

(a) Mutual commitments of the trustee and the truster: attributes  $Att^c$  of these commitments are (in general) observable before the trust act has been carried out.

A typical attribute of this category is for example the price for the merchandize or the scope of the services offered, such as the amount and precision of information in case of a negotiation among agents regarding the delivery of information.

(b) Trust attributes  $Att^t$ : these attributes can include all dimensions of trust wrt. the objectives or interests of the truster, e.g., whether after a merchandize the bought item is as described by the seller, or the actual delivery time (see example below). These attributes are not observable until the trust act has been carried out.

Figure 1 illustrates the IHRTM as a DAPER model which is commonly used for relational graphical models (c.f. Heckerman et al. 2004). Entity classes are depicted as rectangles and the relationship class as a rhombus. Observable evidence  $Att$  is modeled in attribute classes of entities and relationships (ovals). As in a classical nonrelational Bayesian network, direct statistical dependencies are modeled as directed arcs. The DAPER model should be thought of as a template which, based on the actual objects in the domain, is expanded into the ground Bayesian network.

To illustrate the abstract model we will use the *eBay feedback-system* as a concrete example throughout this paper. Being the most popular online auction and shopping website, fraud on trading and auction platforms is a serious and well-known problem. An attempt to deal with fraud is the eBay feedback-system where users leave feedback about their past transactions with other eBay-users.

Suppose the truster-agent is a buyer who wants to build a context-sensitive relational trust model to analyze the trust situation on eBay in general and assess trust values for purchases from eBay in particular. In this scenario, the truster itself does not need to be modeled explicitly because he learns a *personalized* model based on its own viewpoint and experience. The trustee  $a$  however represents sellers on eBay and the state  $s$  represents items that are for sale. The relation  $interacts(a, s)$  would best be specified as  $offers(a, s)$  in this context.

The attributes  $Att$  specify the observable features of the trust situation.  $Att^a$  describes properties of the seller like the feedback score, the percentage of positive feedback and his length of membership.<sup>1</sup>  $Att^s$  specifies features that are associated with the product, for instance its category and its condition (new or used). The price however is represented as a relational attribute  $Att^c$  because a different seller could offer the same product for a different price. Thus,  $Att^c$  stands for all *commitments* seller and buyer make in the negotiation process. Besides the price or winning bid this can e.g. be shipping costs, bidding history, extent of warranty, payment details and shipping rates. Finally  $Att^t$  can include all dimensions of trust that are important for the truster when he finally gives feedback about his purchases. Relevant dimensions might be: actual shipping time, whether the item was as described, if the communication with the seller was as expected and so on.

As an example, one could now express the trustworthiness of an offer concerning product quality  $Att^t$ , given the seller  $a$  offers item  $s$  for price  $Att^c$ . Note that more than one attribute per entity or relation can be considered as well.

### 3 Technical details

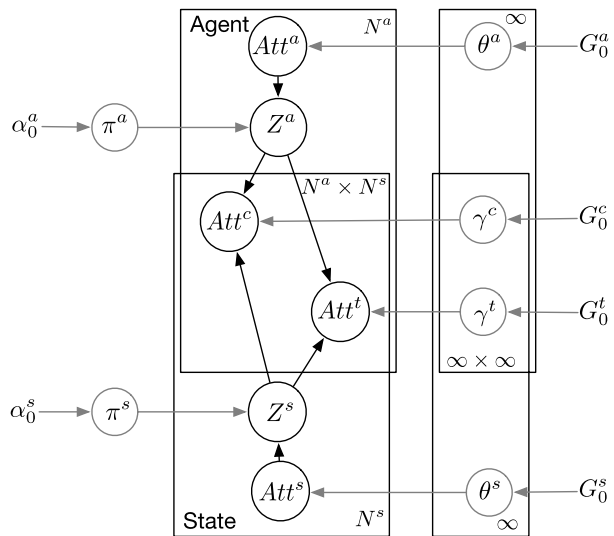
To complete the technical details of our specific relational trust model we now introduce the remaining elements of the IHRTM. Following the ideas of Xu et al. (2006) and Kemp et al. (2006) we assign to each entity a hidden variable, denoted as  $Z^a$  and  $Z^s$  and depicted as circles in Fig. 1. Related to the hidden states in hidden Markov models, they can be thought of as unknown attributes of the entities and are the parents of both the entity attributes and the relationship attributes. The underlying assumption is that if the hidden variables were known, both entity attributes and relationship attributes can be well predicted. A very important result of introducing the hidden variables is that now information can propagate in the ground network, which consists here of attribute variables exchanging information via a network of hidden variables.

Given that the hidden variables  $Z$  have discrete probability distributions they intuitively can be interpreted as cluster variables where similar entities (similar sellers or similar items) are grouped together. The cluster assignments (or hidden states) of the entities are decided not only by their attributes, but also by their relations. If both the associated seller and item

---

<sup>1</sup>Note that this refers to assessments of the trustee provided by other trusters (agents who provided feedback on eBay), and these are not necessarily trustworthy themselves. However, eBay provides a range of measures in order to increase the trustworthiness of feedback (e.g., to prevent dummy accounts). Furthermore, as we will describe next, eBay feedback is just one (optional) type of information about the trustee in our model. Other information might be more influential.

**Fig. 2** Graphical representation 2: Plate model



have strong known attributes  $Att^a$  and  $Att^s$ , those will determine the cluster assignments and the prediction for  $Att^t$ . In terms of a recommender-system terminology we would obtain a content-based recommendation system.

Conversely, if the known attributes  $Att^a$  are weak, then the cluster assignments  $Z^a$  for the seller  $a$  might be determined by the relations to items  $s$  and cluster assignments of those items cluster assignments  $Z^s$ . Accordingly, this applies to items  $s$  and its cluster assignment  $Z^s$ . In terms of a recommender-system terminology we would obtain a collaborative-filtering system because IHRTM can exploit regularities in the relations of sellers to items. Consequently, IHRTM provides an elegant way to combine content-based predictions with collaborative-filtering prediction.

In the IHRTM,  $Z$  has a potentially infinite number of states, which means that the number of actually occupied components is determined automatically in the inference process. The fundamental idea is, that depending on the complexity of the problem, the model can “decide” for itself about the optimal number of states for the hidden variables; thus, a time consuming optimization of the number of clusters can be avoided.

After sketching the functioning of the infinite hidden variables, we can complete the model by describing the local distribution classes denoting the parameters and hyperparameters of the probability distributions. They are shown as small gray circles in the DAPER model (Fig. 1). As an alternative to the DAPER model, we can display the structure of the IHRTM as a plate model, another commonly used graphical representation for Bayesian models (Fig. 2).

Often, the hardest modeling choice is to decide which likelihood distribution to use, especially to decide on the number of parameters for the model. To meet this challenge, *non-parametric Bayesian models* are used to learn the functions of interest directly from the data. The term non-parametric does not mean that there are no parameters in the models, but that the number and properties of the parameters are flexible and not fixed in advance.

In statistical machine learning, the most common non-parametric Bayesian models use *Dirichlet processes* (DP) or Gaussian processes. Dirichlet processes are used in discrete settings for density estimation and clustering, like in our application. The term process—as opposed to distribution—means that the degrees of freedom of the model are infinite.

A popular pattern for modeling relational settings are *mixture models* which also rely on DPs. Mixture models are well suited in situations where the samples are potentially generated under different conditions and are widely used for clustering and classification problems. The idea is to model the data as separate distributions, rather than building a single distribution

One problem in applying finite mixture models is the difficulty to decide the number of mixture components in advance. A possible solution for the problem is to embed the finite mixture model in a non-parametric Bayesian framework using a DP resulting in a *infinite mixture model*. Then, the number of mixture components is not restricted and will be optimized with respect to the data in a self-organized way. This means that only the hyperparameters  $\alpha_0$  of the DP need to be specified by hand.

Like the term non-parametric the term *infinite* does not mean the number of mixture components are infinite, but the number is flexible and not fixed in advance. Due to the combination with DPs, the infinite mixture model is also referred to as Dirichlet process mixture model. For additional technical details on Dirichlet process mixture models, consult for example Tresp (2006).

Let's consider the variables for the seller entity. For each specific seller  $i$  there is a hidden variable  $Z_i^a$  with the flexible and potentially infinite number of states  $K^a$ . The clustering  $Z_i^a = k$  specifies the assignment of seller  $i$  to the specific cluster  $k$ . The weights  $\pi^a = (\pi_1^a, \dots, \pi_{K^a}^a)$  are multinomial parameters with  $P(Z^a = k) = \pi^a$  and are drawn from a conjugated Dirichlet prior,  $\pi^a \propto \text{Dir}(\cdot | \alpha_0^a, \alpha^a)$ .  $\alpha^a = (\alpha_1^a, \dots, \alpha_{K^a}^a)$ .  $\alpha_k^a$  represents our prior expectation about the probability of a seller being in cluster  $k$ .  $\alpha_0^a > 0$  determines the tendency of the model to either use a large number (large  $\alpha_0^a$ ) or a small number of clusters in  $Z$  (small  $\alpha_0^a$ ).

Since we only consider discrete attributes in our eBay example, a particular attribute  $Att^a$  is a sample from a multinomial distribution with multinomial parameters  $\theta^a = (\theta_1^a, \dots, \theta_{K^a}^a)$ . The base distributions  $G_0^a$  and  $G_0^s$  are the associated conjugate priors. So,  $\theta^a \propto G_0^a$ . The same applies to the multinomial parameter  $\gamma$  for each of the  $K^a \times K^s$  configurations related to each relational attribute  $Att^c$  and  $Att^t$ . Again, a Dirichlet process prior is employed, so that  $\gamma^c \propto G_0^c$ .

Now we briefly describe the generative models for the IHRTM. The method we use to generate samples from a Dirichlet Process mixture model is the Chinese restaurant process (CRP, see Tresp 2006). The clustering of data points in a DP can be explained by the following analogy: Imagine a restaurant with an infinite number of tables. Now customers enter the restaurant one by one and choose a table to sit down. Each customer either chooses to sit down at an unoccupied table or join some other customers at an already occupied table, where the table selection probability is proportional to the number of persons already sitting at a table. Applying this scenario to the Dirichlet process, the tables are clusters and the customers are data-points. After  $N$  data-points are sampled the  $N + 1$ th sample is generated as follows.

- The  $N + 1$ th agent is assigned to an existing agent cluster  $i$  with probability  $\frac{N_i}{N + \alpha_0}$  and inherits parameters  $\theta_i$  and  $\gamma$ .
- With probability  $\frac{\alpha_0}{N + \alpha_0}$  the agent is assigned to a new cluster  $K + 1$ .

For the new user cluster, new parameters  $\theta_i$  and  $\gamma$  are generated as described above.

The procedure is repeatedly applied to all hidden variables in the ground network.

### 3.1 Inference

Based on the generative model presented in the previous section we can now generate samples from the IHRTM. In particular, we are interested in how to generate samples from the unknown states and parameters, given observed data. The most important goal is to infer the conditional distribution of the hidden variables  $Z^a$ ,  $Z^s$  given all known attributes entity attributes  $Att^a$  and  $Att^s$  as well as relationship attributes  $Att^c$  and  $Att^t$ . This eventually allows us to make predictions about unknown attributes, like target value  $Att^t$ .

A way to approximate this posterior distribution of the hidden variables is by means of Gibbs sampling (GS), an MCMC-method. In our model, it is possible to formulate a GS in which only samples from the hidden variables are generated by integrating out model parameters (see Xu et al. 2006). The Markov chain is thus defined only for the hidden variables of all entities in the given domain. The GS iteratively samples the hidden variable  $Z^a$ , conditioned on the other hidden variables  $Z^s$  until the procedure converges. In particular,  $Z$  is updated as:

1. For  $Z^a$ : Pick a random agent  $i$ . Assume that for  $N_k^a$  agents,  $Z^a = k$  without counting user  $i$ .

Either assign agent  $i$  to cluster  $k$  with probability proportional to

$$P(Z_i^a = k | Z_{j \neq i}^a, Att_i^a, \theta^a, \gamma^c, \gamma^t, Z^s) \propto k P(Att_i^a | \theta_k^a, \gamma_{k,*}^c, \gamma_{k,*}^t)$$

where  $N_k$  is the number of agents already assigned to cluster  $k$  and  $\gamma_{k,*}$  notes the relation parameters of agent cluster  $k$  and all state clusters.

Or generate a new cluster  $K + 1$  with probability proportional to

$$P(Z_i^a = K + 1 | Z_{j \neq i}^a, Att_i^a, \theta^a, \gamma^c, \gamma^t, Z^s) \propto \alpha_0^a P(Att_i^a | \theta_{K+1}^a, \gamma_{K+1,*}^c, \gamma_{K+1,*}^t).$$

2. For  $Z^a$ : Pick a random state  $j$  and update its cluster assignment  $Z^s$ , accordingly.
3. If during sampling a state becomes unoccupied, remove that state from the model and re-assigned indices.

After the Markov chain has converged, standard statistical parameter estimation techniques can be used for estimating the parameters  $\gamma_{k^a, k^s}^t$  of  $Att^t$  from given cluster assignments. For a detailed description of the algorithm we refer to Xu et al. (2006). We extended the algorithm, as just described, to enable the handling of more than one relationship attribute. Being able to use an arbitrary number of relationships is essential to enable a rich representation of the interaction context as well as multidimensional trust values.

### 3.2 Implications

The ultimate goal of the model is to group entities into clusters. A good set of partitions allows to predict the parameters  $\gamma$  of attributes  $Att^c$  and  $Att^t$  by their mere cluster assignments. In the ground truth, our model assumes that each entity belongs to exactly one cluster. It simultaneously discovers clusters and the relationships in-between clusters that are best supported by the data, ignoring irrelevant attributes.

Although the value of attributes is determined entirely by the cluster assignment of associated entities, there is no need for direct dependencies between attributes. The cluster assessment of an entity is influenced by all corresponding attributes and cluster assessments of related entities. This way information can propagate through the whole network while the infinite hidden variables  $Z$  act as “hubs”. This allows for a collaborative filtering effect.



Cross-attribute and cross-entity dependencies can be learned, something which is not possible with a “flat” propositional approach that assumes independent and identical distributed (i.i.d.) data.

At the same time the number of clusters does *not* need to be fixed in advance. Thus, it can be guaranteed that the representational power is unrestricted.

## 4 Experimental analysis

To investigate the performance of the IHRTM we employ real world data from the eBay example used for illustration in the previous section. Before the empirical results of our experiments will be presented, we first describe the experimental setup. In the following sections three different aspects of the IHRTM’s performance were investigated in the course of our research: First the algorithm’s abilities to characterize a trust-situation by clustering are investigated in Sect. 4.2. Second the predictive performance concerning trust values is tested. Finally, the learning efficiency is analyzed in the context of dynamic behavior of non-stationary trustees. As the later cannot be analyzed within the eBay scenario we used interactions recorded from a negotiation game. The experimental setup and evaluation is covered in Sect. 4.4.

### 4.1 Experimental setup: eBay-user ratings

eBay feedback-profiles are a valuable source of easily accessible data that expresses human-trust assessment. Every eBay member has a public feedback profile where all items he has bought or sold in the last 90 days are listed with the associated feedback ratings he received. In addition the feedback profile includes statistics on all transactions of the user.

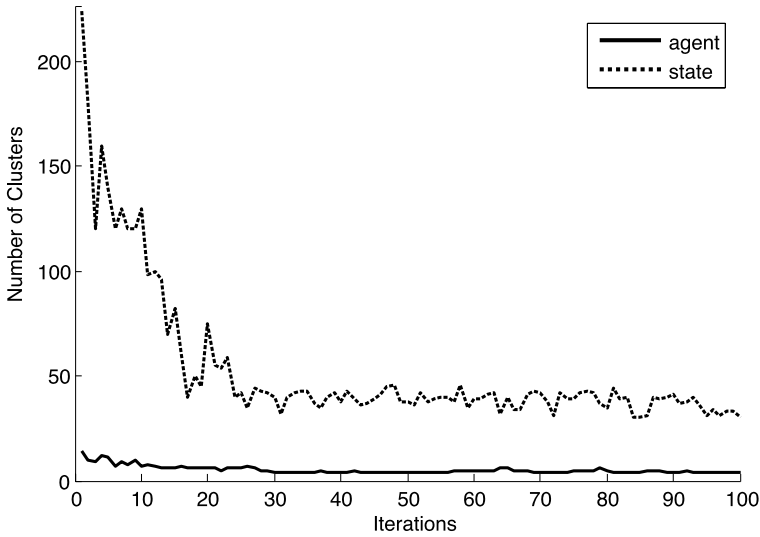
We gathered data from 47 sellers that on the one hand had at least 10 negative or neutral ratings and on the other hand sold items in at least one of 4 selected categories from the lowest level within the eBay-taxonomy. The former is important because negative or neutral user-ratings on eBay are rather rare. Neutral ratings were treated as negative ratings. To further balance the ratio of positive vs. non-positive ratings we only evaluated as many positive rated transactions as there were non-positive ones. In this way the data-set is stratified, meaning that there is an equal number of positive and non-positive ratings per seller.

Attributes  $Att^d$  of the seller were directly extracted from the feedback profile. We picked the positive feedback and the feedback score and discretized both in 2 and 5 classes, respectively. For the item attributes  $Att^s$  we chose the top level category in the eBay taxonomy on the one hand, resulting in 47 discrete states. On the other hand we collected the item condition which is a binary feature: either new or used.

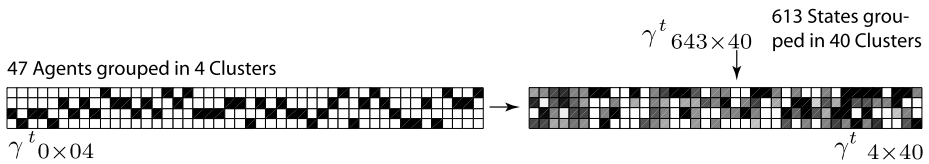
From those 47 hand-picked sellers we gathered a total of 1818 rated sales of 630 different items. Two items were assumed to be alike if they were in the same lowest level category and their attributes were identical. Relation attributes are always of size  $seller \times items$ , so  $Att^c$  and  $Att^f$  both are sparse matrices with  $47 \times 630$  possible entries. The non-zero entries indicate that this seller has sold this item.

The extracted data set is not very extensive, but constitutes a representative experimental setup for trust assessment. Results that focuses on the computational complexity of a comparable algorithm in a different setting are presented in Xu et al. (2009).

As we wanted to keep the formal and computational complexity low we only considered binary relational attributes  $Att^c$  resp.  $Att^f$ . For  $Att^c$  we chose the binarized final price of the auction and for  $Att^f$  the rating. Negative and neutral ratings were both treated as negatives.



**Fig. 3** Trace of the number of agent- and state-clusters up to 100 iterations



**Fig. 4** *Left:* Final clustering of trustees  $Z^a$ . *Top right:* Items  $Z^s$ . *Bottom right:*  $P(\gamma^t | Z^a, Z^s)$

After having extracted the data, the GS-process to train the IHRTM was started. In the beginning the sellers and items are re-clustered intensely and both cluster assignments and cluster sizes are unstable. Once the Markov chain starts to converge the cluster sizes tend to stabilize and eventually, the training can be stopped. The decrease of the number of clusters is exemplarily shown in Fig. 3 for one cross-validation run.

### 4.2 Clustering effect

After the clusters have stabilized we can visualize two interesting facts about the trust situation.

First, we can plot a matrix showing the assignments of each seller to a cluster. This can provide knowledge about how many different clusters exist, which are the most popular clusters and which elements are grouped together. After convergence, the 47 sellers were assigned to 4 clusters as shown on the left half of Fig. 4. The same assignment matrix can be generated for the items cluster assignment but since there are 613 items and 40 item clusters, we did not plot the matrix and simply show its symbol  $Z^s$  on top of the right matrix in Fig. 4.

Second, the posterior probability  $P(Att^c, Att^t | Z^a, Z^s)$  can be visualized. The matrix on the right side in Fig. 4 illustrates the probability of getting a positive rating given the cluster assignments of a seller and an item. A darker value indicates a higher probability of being trustworthy in a given interaction. Now, picking a row (representing an agent cluster) or

a column (representing a state cluster) we can identify clusters that are in general more trustworthy than others.

### 4.3 Predictive performance

In order to judge the performance of predicting the trust value  $Att^t$  we compared the results of IHRTM with two other standard machine learning algorithms, namely a support vector machine (SVM) using a PolyKernel and a Decision Tree (*DecTree*) both from the Weka toolbox (Witten and Frank 2002). Since those algorithms are both propositional learners, meaning they cannot handle a relational data representation but only a vector of independent and identically distributed features plus a label, we had to “flatten” the data first. By transforming the data into a flat representation, also known as “propositionalization”, the structural information can get lost. In general there is no standard propositionalization procedure (see Krogel 2005). The potential low quality of propositional features is not crucial in our simple scenario but becomes increasingly problematic in more complex relational models.

We propositionalized the data in three different ways: First, we only considered the target trust variable  $Att^t$  and tried to predict trustworthiness by the mere rate of positive feedback as it is done in most existing statistical trust models (see *Ratio* in Table 1). Clearly, the result cannot be better than random guessing as the data-set is stratified. However, this demonstrates that the assumption of context independency made by many trust models is fatal when trust observations are uniformly distributed. Second, we tested the performance of the propositional algorithms with all features—namely  $Att^a$ ,  $Att^s$ ,  $Att^c$  and again  $Att^t$ —as the label. As a result we extracted 1818 samples with 5 features and one label, each. This way, the same features are available to the propositional learners as they are to the IHRTM. Third, we accounted for the missing relational information (which seller sold which product) by introducing two further features: An ID-number for the seller and the item, respectively. So the input to the propositional learners was a  $1818 \times 8$  matrix in this setup.

In addition, we compared a standard single relational algorithm namely a Singular Value Decomposition (SVD) to test the performance when taking advantage of the collaborative effect. Note, that SVD can only use  $Att^t$ ,  $Att^a$  and  $Att^s$  as inputs, because it is not a multi-relational algorithm.

The result of all 3 setups is shown in Table 1. We report the accuracy of predicting positive ratings as well as the AUC (also called ROC area). This measure represents the area under the *receiver operating characteristic curve* which is used for evaluating binary classifiers that can output probabilities instead of binary decisions. In all our experiments we averaged our results using 5-fold cross-validation. The accompanying 95%-confidence intervals are reported as well. Finally the prediction performance is also evaluated for the IHRTM and compared to the previous attempts (see Table 1).

In general, the task of predicting eBay-user ratings seems to be difficult, which can be explained when reading the comments assigned to the ratings. The reasons for a positive or a negative evaluation are most of the time not related to specific properties of sellers or items but a unique incident. Besides that, the high incentives to give positive ratings despite having had negative experience are a general and well known flaw in the eBay-feedback mechanism: sellers usually wait for the buyer’s rating before they rate the buyer. Thus, buyers often give positive rating just to receive a positive rating from the seller as well. As a response to this problem, eBay has introduced a new feedback mechanism in 2008.

Still, the IHSTM’s performance clearly outperforms random guessing and could verifiably outperform the propositional and single-relational learners. This is most likely due to

**Table 1** Predictive performance on eBay-user ratings

	Accuracy	ROC area
Ratio	48.5334 ( $\pm 3.2407$ )	–
SVM	54.1689 ( $\pm 3.5047$ )	0.512 ( $\pm 0.0372$ )
DecTree	54.6804 ( $\pm 5.3826$ )	0.539 ( $\pm 0.0502$ )
SVM+ID	56.1998 ( $\pm 3.5671$ )	0.5610 ( $\pm 0.0362$ )
DecTree+ID	60.7901 ( $\pm 4.9936$ )	0.6066 ( $\pm 0.0473$ )
SVD	65.4728 ( $\pm 6.0375$ )	0.6843 ( $\pm 0.06421$ )
IHRTM	<b>71.4196</b> ( $\pm 5.5063$ )	<b>0.7996</b> ( $\pm 0.0526$ )

the collaborative filtering effect, that can only be utilized by the IHSTM and partly by the SVD. Thus, there seems to be a gain if learning with the assumption that e.g., when two sellers sell similar items they might be comparable in their trust-ratings. More precisely, if two sellers both got positive ratings after selling one specific item their ratings might be comparable when selling a different item as well. Or the other way round, if two items both got positive ratings after sold by one specific seller their ratings might be comparable when sold by a different seller as well. However, the performance of SVD shows, that modeling the trust-relation alone also gives inferior results.

#### 4.4 Learning efficiency

As mentioned in the introduction, the learning efficiency<sup>2</sup> and the ability to rapidly adapt is crucial, especially in so called initial-trust situations or in situations where the trustee does learn and adapt as well. To evaluate the performance concerning learning efficiency, we had to use a different, more controlled experimental setup as in the previous eBay example. Only if we know about the stationarity of agents we can compare the performance of an adapting agent to a stationary agent. For this purpose, we recorded interactions in a simulated strategic two-player negotiation scenario.

##### *Experimental setup: Negotiation game*

Finding an agreement amongst a group of conflicting interests is one of the core issues of distributed artificial intelligence. Auctions, information markets, preference and judgement aggregation, game theory and automated negotiations are all research areas that deal with those kind of problems. However, most of the approaches neglect the fact that finding the best agreeable solution is not sufficient if commitments can not be enforced by the interaction mechanism or the incentives of the opponents can not be inferred. In order to investigate this issue we extended the implementation of a multiagent trading framework by an additional negotiation step.

In the chosen scenario, players try to collect a certain number of resources for selling. Only one type of resource can be collected at a time. In each round, every player gets new random resources from the deck and some resources must be added to the stack of collected resources. If the types of the resources previously held in the stack and the types of new resources are not identical, all resources collected so far are wasted. To avoid this, players

<sup>2</sup>By *learning efficiency* we do not mean computational complexity of the learning algorithm, but numbers of observations needed to make effective predictions.

can trade with other players and exchange some of their resources. Resources received from fellow players are pushed onto the stack.

As defined before, let  $c$  be the commitments that the agents are negotiating over. The outcome of this negotiation is specified by a set of binary features  $Att^c$ . Now, given a set of commitments  $c$  that two agents have agreed on and promised to fulfill, the agents enter an additional trading step in which each of them is free to decide which action to take. This way, the agent can decide whether to stick to a commitment or break it at will.

One interaction-round consists of three phases:

**Negotiation:** Each agent  $a$  follows a predefined strategy that proposes a potential set of actions  $c$  both parties might agree on (e.g., an exchange of goods). In doing so, agents have neither knowledge of the actions available to the other agents nor their reward function. Thus, agents can propose an infeasible action to convince the opponent to act more to their favor. Received proposals can be rejected and counter-proposals can be made resulting in an iterative approximation of a solution. The negotiated result is considered as a commitment attribute  $Att^c$ .

**Trading:** This is the final decision made by every agent whether to stick to a commitment or break it. Finally, the action  $t$  chosen by agent  $a$  is executed accordingly.

**Evaluation:** The agents can review the effective actions  $t$  of the opponent by observing the received goods and draw conclusions for future interactions. The next stage game is sampled according to a stochastic transition function.

This procedure is repeated over a potential infinite number of rounds with different types of agents playing against each other.

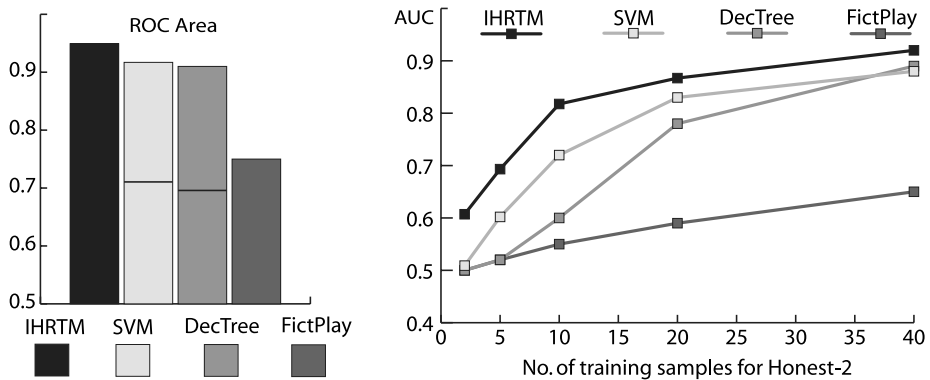
### Evaluation

Three different agent types with two different negotiation strategies and three different trading strategies were used as opponents in the negotiation game.

The two negotiation strategies are both stationary and are based on a monotonic concession protocol (cf. Endriss 2006). The agents denoted *Honest* and *Fictitious* only propose actions that they actually could perform, while agent *Greedy* also offers and accepts infeasible actions with the intend to achieve an opponent action with higher payoffs. Both strategies iteratively propose a larger set of actions by lowering their expected utility and offering less favorable outcomes.

Each agent type plays a different trading strategy where *Honest* and *Greedy* are stationary and *Fictitious* is adaptive. *Greedy* always maximizes its utility regardless of  $c$ , while *Honest*-agent always sticks to  $c$ . At last, *Fictitious* plays according to the *fictitious play* algorithm. It's a traditional learning algorithm from game theory for repeated games, where the opponent is assumed to be playing a stationary strategy. The opponent's past actions are observed, a mixed strategy is calculated according to the frequency of each action and then the best response is played, accordingly.

In every round that was played the commitment  $c$  and the effective outcome  $t$  were recorded and features  $Att^s$ ,  $Att^c$  and  $Att^t$  were extracted. No specific attributes for  $Att^a$  were available except for the identity of the agent. Three discrete features  $Att^s$  from  $s$  were calculated describing the average payoff over all possible opponent actions, the maximum possible payoff and the number of feasible actions.  $Att^c$  describes a single binary feature stating whether there is a feasible action that could be carried out and would result in a positive reward if the negotiated commitment was carried out by the opponent. The same feature was recorded for  $Att^t$  after the actual action took place.



**Fig. 5** Results for play against *Honest*. Bar graph on the left: AUC for classifying  $Att^f$ . Graph on the right: learning curve for increasing number of training data for the additional *Honest-2*

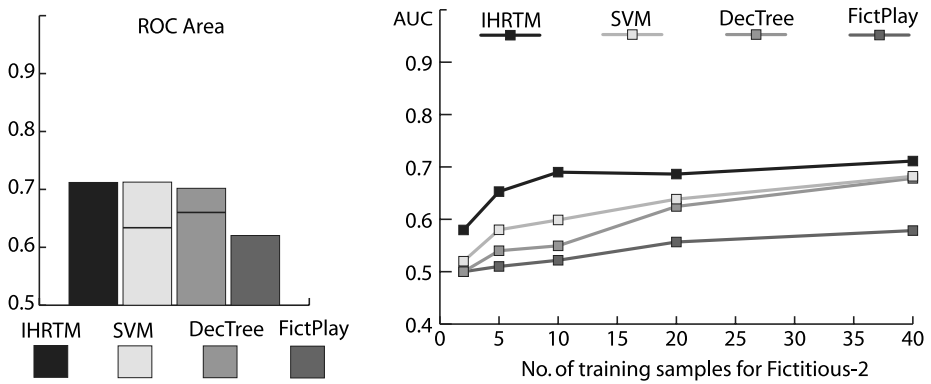
In this way a total of 600 interactions, 200 per agent type, containing a total of 289 different stage games were recorded. The input for the IHRTM consisted of three  $Att^s$  vectors with 289 elements, and two  $289 \times 3$  matrices for  $Att^c$  and  $Att^f$ . Again, for a comparison with propositional machine learning algorithms the data was propositionalized, resulting in 600 feature vectors with  $3 \times Att^s + 1 \times Att^c$  elements and 600 corresponding labels. As before, the content based algorithms were also evaluated with an agent- and state-ID as an additional feature. The evaluation procedure is the same as in the eBay experiments.

The overall performance according to AUC is depicted in the bar graph on the left of Fig. 5. IHRTM shows a slightly better performance in classifying  $Att^f$  than SVM and DecTree. Without the agent-ID as an additional feature the performance of DecTree and SVM drops considerably (black line at around 0.7). Again, we explain the superior performance by IHRTM's ability to exploit cross-entity dependencies. *Fictitious*, as expected, performs much worse as it is not able to generalize over different interactions and can't make use of the context provided by  $Att^s$  and  $Att^c$ .

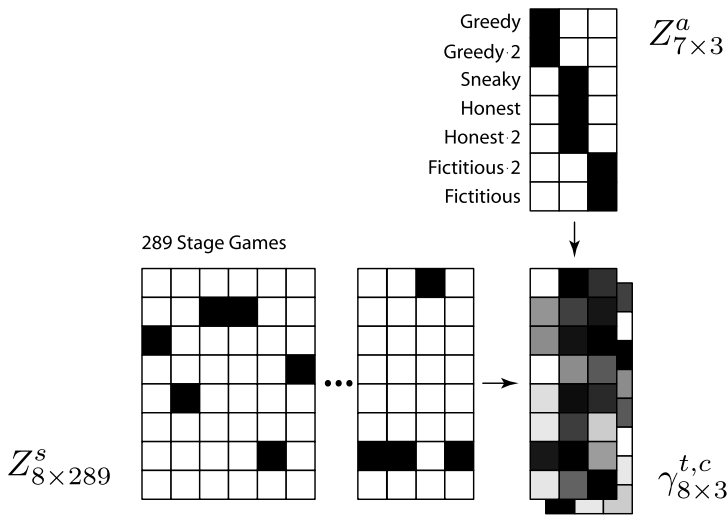
The inherent clustering ability of IHRTM suggests that it is especially well suited for rapid adaptation when unknown but related agents and conditions are observed. Actually, entities can be correctly assigned to a cluster without having seen a single effective  $Att^f$  related to this entity just by the other attributes. To check this assumption we gathered data from interactions with another *Honest* type agent and evaluated the performance for different numbers of training samples. On the right of Fig. 5 the learning rates for agent *Honest-2* are plotted. The results confirm that especially for a small sample size  $\leq 20$  the performance of IHRTM is clearly better compared to the content based approaches.

In contrast, the performance in the task of trying to predict *Fictitious* is clearly worse for all of the techniques (see Fig. 6). Expectedly, IHRTM, SVM and DecTree cannot handle dynamic opponents. Again, the IHRTM is most competitive in terms of efficient learning.

In addition, the IHRTM offers another advantage over the other techniques. The predictions are based on an inherent construction of clusters of  $Z^a$  and  $Z^s$ . The fast learning rate indicates that a previously unknown trustee is correctly assigned to an existing cluster if this type of agent has been observed before. Consequently, once *Fictitious-2* is assigned to the "*Fictitious-cluster*" IHRTM could assess its performance on this cluster and eventually suggest a different learning scheme for agents in this cluster. In other words it can identify non-stationary behaving agents.



**Fig. 6** Results for play against *Fictitious*. Bar graph on the left: AUC for classifying  $Att^l$ . Graph on the right: learning curve for increasing number of training data for the additional *Fictitious-2*



**Fig. 7** Final clustering of agent types and states ( $Z^s$  and  $Z^a$ ). Bottom right:  $P(\gamma^l, \gamma^c | Z^s, Z^a)$

Figure 7 visualizes the final cluster sizes and cluster assignments. The top right matrix shows the assignment of seven different agents to  $Z^a$ . All three agent types were clustered correctly into three groups (columns). To evaluate this further we generated data from another stationary opponent with a different trading strategy that is very similar to *Honest*: *Sneaky*-agent only deviates from  $c$  if it can increase its utility by a large margin. Interestingly, the assignment of *Sneaky*- and *Honest*-agent to the same cluster suggests that this strategy might effectively build trust. The matrix in the lower left corner of Fig. 7 visualizes  $Z^s$ . From 289 stage games (columns) 8 different clusters (rows) emerged. This is an impressive reduction in complexity while still having good classification results. The two stacked matrices in the bottom right corner represent  $Att^l$  and  $Att^c$  (below). Each row indicates one state cluster, each column an agent cluster. Brighter rectangles indicate a lower probability for a positive reward. As expected, the first column (*Greedy* cluster) is on average brighter

than the second and third column (*Honest* and *Fictitious* cluster). All those observations, including the misclassification of *Sneaky*, correspond well to human intuition.

## 5 Related work

A very large number of works on trust exist in Artificial Intelligence. However, most of them are only remotely related to our approach, since they do not attempt to learn trust from data using machine learning approaches, but instead propose trust management frameworks, or trust models or trust metrics which determine how trust and reputation of different dimensions and/or from various sources are represented and aggregated to form a single notion of trust. See Sabater and Sierra (2002), Huynh et al. (2004), Ramchurn et al. (2004), Maximilien and Singh (2005), Teacy et al. (2005), Richardson et al. (2003), Golbeck (2008), Kamvar et al. (2003), and Mauá and Cozman (2008) for a few examples from the fields of intelligent agents, the Web and the Semantic Web, Social Networking, and P2P networks, respectively.

The predominant fraction of existing trust models also assigns only a single trust value to each agent as a whole. They do not take into account that trust is a result of past behavior which is a sequence of (re)actions. The (re)action of an agent in a trust situation determines the assessment of trust in future actions of this agent (cmp. Moranz 2004). It does not directly allow to assign a trust value to the agent as a whole.

In addition, most existing models do not allow trust decisions to be situated in a specific context. This applies to implemented trust systems like (see Zanero 2005; Kinateder et al. 2005; Fullam et al. 2005) and to extended trust models. We will discuss related works w.r.t. context later in this section.

Examples for extensions or variations of “plain trust” are e.g., risk (see Ruohomaa and Kutvonen 2005), distrust, mistrust and untrust (see Marsh and Dibben 2005), context (see Jøsang et al. 2005), forgiveness (see Vasalou and Pitt 2005), recommendations (see Chadwick 2005) and affects (see Hassell 2005). However, we aim at including a wide range of contextual relations to be considered during trust assessment.

Given a trust model trust values are in general estimated by counting the success of previous interactions concerning a *trust-metric*. This trust-metric is improved over time (see Witkowski et al. 2001). The target function of the agents are the optimization of the outcome of interactions and the cooperation between agents (see Wu and Sun 2001). The evaluation is performed using various different approaches, including e.g., fuzzy set theory (see Rehak et al. 2005) and Bayesian estimation (see Shi and Adams 2005).

As already pointed out, connecting trust to the trusted agent alone without considering contextual and other aspects (dimensions) of trust is not sufficient in many scenarios. Whereas some research on trust concedes the importance of context information, most of them do not actually use such information for the calculation of trust degrees. Especially, most machine learning techniques used for trust learning have not been focused on this issue or do not offer a solution in a general and automated way (see Teacy 2006). Among those approaches which take context actually into account, only Rehak and Pechoucek (2007) makes use of rich context information as we do, including agent identity and environmental state, in order to assess trustworthiness. Besides that, using rich contextual information for initial trust assessment and the transfer of trust between contexts is novel to our knowledge.

Other approaches take into account restricted, specific fragments of an agent’s environment or properties. E.g., the NetProbe approach (Pandit et al. 2007) uses the position of trustees within certain kinds of social networks in order to detect fraudulence. In contrast,



the focus of our approach is the initial trust learning from a variety of data types and sources, not just from patterns in social network graphs. This possibly explains why we get on fine with a relatively small domain size in our eBay experiments.

Regarding its dimensionality, most work represents trust as a single discrete or continuous variable associated with one specific agent. Modeling trust in multiple dimensions is only considered by a few elaborate approaches such as Maximilien and Singh (2005). Our approach leaves it to the actual scenario how trust needs to be modeled in this respect. In principle, IHRTM can handle an arbitrary number of trust variables, each associated with one aspect of the trustor's expectations and represented with any probability distribution needed.

Analogously, we argue that a fine grained modeling of relations between agents and their environment is essential to capture the essence of trust, especially in initial trust situations. There exist a few approaches that can take relationships into account when modeling trust. But in most of this research such relationships are either only considered as reputation or recommendations (see Sabater and Sierra 2001), or as interactions between a group of agents (e.g., Ashri et al. 2005). The diverse kinds of relations that exist between two agents in a specific situational context are not modeled in detail. In addition, most learning techniques are optimized for one specific scenario only and do not make use of well funded techniques from probability theory.

Assessing initial trust values for unknown agents based on pre-specified membership to a certain group has been addressed by Sun et al. (2005). Here, a group-based reputation architecture is proposed where new agents are assessed according to their pre-specified membership to a certain group of agents. Likewise, the *TRAVOS-C* system proposed by Teacy (2006) includes rudimentary ideas from hierarchical Bayes modeling by assigning parameter distributions to groups of agents but does not come to the point to give a fully automated and intuitive way of how to learn infinite hidden variables.

Another interesting recent approach is presented in Hendrix et al. (2009). However, the focus is here on the effective and most trustable combination of information provided by a number of different information sources. It would be interesting to combine this approach with ours in order to integrate information from third-party trusters who provide reviews on the trustee (its offers, respectively) into our model.

## 6 Conclusions and future work

In this contribution, we presented a context-dependent way to build statistical relational trust models in general and our *Infinite Hidden Relational Trust Model* (IHRTM) in particular. We demonstrated how trust can be modeled and learned in theory and in two experimental setups: first, a real world data set from the *eBay feedback-system* and second a simulated negotiation game.

Our experimental results suggest that the IHRTM offers advantages in 3 different dimensions. First, the inherent clustering capabilities increase interpretability of trust situations. Second, the predictive performance can be improved compared to a “flat”, feature-based machine learning approach if trained with relational data that exhibit cross-attribute and cross-entity dependencies. Third, the IHRTM is especially well suited for rapid adaptation because of its ability to transfer knowledge between related contexts.

While the IHRTM cannot handle trustees with strategies that are non-stationary effectively, it can identify non-stationary agents. An adaptive learning strategy could be part of

future work. Furthermore, we plan to extend our framework to scenarios with arbitrary numbers of concurrently interacting trustees. While propositional machine learning algorithms cannot be easily applied in this case it can be realized by relational models.

We introduced statistical relational trust learning in general and presented the IHRTM in particular. The goal of our work is to advance research on computational trust by making modeling and learning of trust more applicable, efficient, intuitive and interpretable.

## References

- Ashri, R., Ramchurn, S. D., Sabater, J., Luck, M., & Jennings, N. R. (2005). Trust evaluation through relationship analysis. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (pp. 1005–1011). New York, NY, USA, 2005. New York: ACM Press.
- Chadwick, D. W. (2005). Operational models for reputation servers. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 108–115). Paris, France, May 23–26, 2005. Berlin: Springer.
- Endriss, U. (2006). Monotonic concession protocols for multilateral negotiation. In *AAMAS '06: Proceedings of the fifth international joint conference on autonomous agents and multiagent systems* (pp. 392–399). New York, NY, USA, 2006. New York: ACM Press.
- Fullam, K. K., Klos, T. B., Muller, G., Sabater, J., Topol, Z., Barber, K. S., Rosenschein, J. S., & Vercouter, L. (2005). The agent reputation and trust (art) testbed architecture. In *The eighth workshop on trust in agent societies (TRUST 2005), at the fourth international joint conference on autonomous agents and multiagent systems*. Utrecht, The Netherlands, July 2005.
- Golbeck, J. (2008). Computer science: Weaving a web of trust. *Science*, 321(5896), 1640–1641.
- Golbeck, B. P. J., & Hendler, J. (2003). Trust networks on the semantic web. In *Proceedings of cooperative intelligent agents*.
- Hassell, L. (2005). Affect and trust. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 131–145). Paris, France, May 23–26, 2005. Berlin: Springer.
- Heckerman, D., Meek, C., & Koller, D. (2004). *Probabilistic models for relational data* (Technical Report MSR-TR-2004-30). Microsoft Research.
- Hendrix, P., Gal, Y., & Pfeffer, A. (2009). Learning whom to trust: using graphical models for learning about information providers. In *AAMAS '09: Proceedings of the 8th international conference on autonomous agents and multiagent systems* (pp. 1261–1262). Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.
- Huynh, T. D., Jennings, N. R., & Shadbolt, N. (2004). Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the 7th international workshop on trust in agent societies*.
- Jøsang, A., Keser, C., & Dimitrakos, T. (2005). Can we manage trust? In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 93–107). Paris, France, May 23–26, 2005. Berlin: Springer.
- Kamvar, S. D., Schlosser, M. T., & Garcia-Molina, H. (2003). The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the twelfth international World Wide Web conference*.
- Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the national conference on artificial intelligence (AAAI)*.
- Kinatader, M., Baschny, E., & Rothermel, K. (2005). Towards a generic trust model—comparison of various trust update algorithms. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 177–192). Paris, France, May 23–26, 2005. Berlin: Springer.
- Krogl, M.-A. (2005). *On propositionalization for knowledge discovery in relational databases*. PhD thesis, die Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg.
- Marsh, S., & Dibben, M. R. (2005). Trust, untrust, distrust and mistrust—an exploration of the dark(er) side. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 17–33). Paris, France, May 23–26, 2005. Berlin: Springer.
- Mauá, D., & Cozman, F. (2008). Managing trust in virtual communities with Markov logic. In *Proceedings of WTDIA '08: The IV workshop on MSc dissertation and PhD thesis in artificial intelligence*.
- Maximilien, E. M., & Singh, M. P. (2005). Agent-based trust model involving multiple qualities. In *Proceedings of the 4th international joint conference on autonomous agents and multiagent systems (AAMAS)*.

- Moranz, C. (2004). *Initiales Vertrauen in der virtuellen Anbahnung geschäftlicher Kooperationen*. PhD thesis, Technische Universität München, Lehrstuhl für Psychologie.
- Pandit, S., Chau, D. H., Wang, S., & Faloutsos, C. (2007). Netprobe: a fast and scalable system for fraud detection in online auction networks. In *WWW '07: Proceedings of the 16th international conference on World Wide Web* (pp. 201–210). New York, NY, USA, 2007. New York: ACM Press.
- Ramchurn, S. D., Huynh, D., & Jennings, N. R. (2004). Trust in multi-agent systems. *The Knowledge Engineering Review*, *19*(1), 1–25.
- Ramchurn, S. D., Sierra, C., Godo, L., & Jennings, N. R. (2004). Devising a trust model for multi-agent interactions using confidence and reputation. *Journal of Applied Artificial Intelligence*, *18*(9–10), 833–852.
- Rehak, M., & Pechoucek, M. (2007). Trust modeling with context representation and generalized identities. In M. Klusch, K. V. Hindriks, M. P. Papazoglou, & L. Sterling (Eds.), *Lecture notes in computer science: Vol. 4676. CIA* (pp. 298–312). Berlin: Springer.
- Rehak, M., Pechoucek, M., Benda, P., & Foltyn, L. (2005). Trust in coalition environment: Fuzzy number approach. In *The eighth workshop on trust in agent societies (TRUST 2005), at the fourth international joint conference on autonomous agents and multiagent systems*. Utrecht, The Netherlands, July 2005.
- Richardson, M., Agrawal, R., & Domingos, P. (2003). Trust management for the semantic Web. In *Proceedings of the 2nd international semantic Web conference (ISWC)*. Sanibel Island, FL.
- Ruohomaa, S., & Kutvonen, L. (2005). Trust management survey. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 77–92). Paris, France, May 23–26, 2005. Berlin: Springer.
- Sabater, J., & Sierra, C. (2001). REGRET: reputation in gregarious societies. In J. P. Müller, E. Andre, S. Sen, & C. Frasson (Eds.), *Proceedings of the fifth international conference on autonomous agents* (pp. 194–195). Montreal, Canada, 2001. New York: ACM Press.
- Sabater, J., & Sierra, C. (2002). Reputation and social network analysis in multi-agent systems. In M. Gini, T. Ishida, C. Castelfranchi, & W. L. Johnson (Eds.), *Proceedings of the 1st international joint conference on autonomous agents and multiagent systems (AAMAS)* (pp. 475–482). New York: ACM Press.
- Shi, G. V. B. J., & Adams, C. (2005). Dealing with recommendations in a statistical trust model. In *Proceedings of the workshop on trust in agent societies (held in conjunction with the 4th international joint conference on autonomous agents and multi agent systems (AAMAS))*.
- Sun, L., Jiao, L., Wang, Y., Cheng, S., & Wang, W. (2005). An adaptive group-based reputation system in peer-to-peer networks. In X. Deng & Y. Ye (Eds.), *Lecture notes in computer science: Vol. 3828. WINE* (pp. 651–659). Berlin: Springer.
- Teacy, W. T. L. (2006). *Agent-based trust and reputation in the context of inaccurate information sources*. PhD thesis, Electronics and Computer Science, University of Southampton.
- Teacy, L., Patel, J., Jennings, N. R., & Luck, M. (2005). Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of the 4th international joint conference on autonomous agents and multiagent systems (AAMAS)*. Utrecht, The Netherlands.
- Tresp, V. (2006). *Dirichlet processes and nonparametric Bayesian modelling*. Tutorial at the Machine Learning Summer School 2006 in Canberra, Australia.
- Vasalou, A., & Pitt, J. (2005). Reinventing forgiveness: A formal investigation of moral facilitation. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 146–160). Paris, France, May 23–26, 2005. Berlin: Springer.
- Witkowski, M., Artikis, A., & Pitt, J. (2001). Experiments in building experiential trust in a society of objective-trust based agents. In *Proceedings of the workshop on deception, fraud, and trust in agent societies held during the autonomous agents conference* (pp. 111–132). London, UK, 2001. Berlin: Springer.
- Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with java implementations. *SIGMOD Rec.*, *31*(1), 76–77.
- Wu, D., & Sun, Y. (2001). The emergence of trust in multi-agent bidding: A computational approach. In *HICSS '01: Proceedings of the 34th annual Hawaii international conference on system sciences (HICSS-34)* (vol. 1, p. 1041). Washington, DC, USA, 2001. New York: IEEE Computer Society.
- Xu, Z., Tresp, V., Yu, K., & Kriegel, H.-P. (2006). Infinite hidden relational models. In *Uncertainty in artificial intelligence (UAI)*.
- Xu, Z., Tresp, V., Rettinger, A., & Kersting, K. (2009). Social network mining with nonparametric relational models. In M. S. H. Zhang, L. Giles, & J. Yen (Eds.), *LNCS advances in social network mining and analysis*. Berlin: Springer.
- Zanero, S. (2005). Security and trust in the Italian legal digital signature framework. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), *Lecture notes in computer science: Vol. 3477. Third international conference, iTrust 2005, Proceedings* (pp. 34–44). Paris, France, May 23–26, 2005. Berlin: Springer.