

Learning Valued Relations from Data

Willem Waegeman
Tapio Pahikkala
Antti Airola
Tapio Salakoski
Bernard De Baets

Abstract Driven by a large number of potential applications in areas like bioinformatics, information retrieval and social network analysis, the problem setting of inferring relations between pairs of data objects has recently been investigated quite intensively in the machine learning community. To this end, current approaches typically consider datasets containing crisp relations, so that standard classification methods can be adopted. However, relations between objects like similarities and preferences are in many real-world applications often expressed in a graded manner. A general kernel-based framework for learning relations from data is introduced here. It extends existing approaches because both crisp and valued relations are considered, and it unifies existing approaches because different types of valued relations can be modeled, including symmetric and reciprocal relations. This framework establishes in this way important links between recent developments in fuzzy set theory and machine learning. Its usefulness is demonstrated on a case study in document retrieval.

1 Introduction

Relational data can be observed in many predictive modeling tasks, such as forecasting the winner in two-player computer games [1], predicting proteins that interact with other proteins in bioinformatics [2], retrieving documents that are similar to a target document in text mining [3], investigating the persons that are friends of each other on social network sites [4], etc. All these examples represent fields of applica-

Willem Waegeman, Bernard De Baets
Ghent University, KERMIT, Department of Applied Mathematics, Biometrics and Process Control,
Coupure links 653, B-9000 Ghent, e-mail: forname.surname@ugent.be

Tapio Pahikkala, Antti Airola, Tapio Salakoski
University of Turku, Department of Information Technology and the Turku Centre for Computer
Science, Joukahaisenkatu 3-5 B 20520 Turku, e-mail: forname.surname@utu.fi

tion in which specific machine learning and data mining algorithms are successfully developed to infer relations from data; pairwise relations, to be more specific.

The typical learning scenario in such situations can be summarized as follows. Given a dataset of known relations between pairs of objects and a feature representation of these objects in terms of variables that might characterize the relations, the goal usually consists of inferring a statistical model that takes two objects as input and predicts whether the relation of interest occurs for these two objects. Moreover, since one aims to discover unknown relations, a good learning algorithm should be able to construct a predictive model that can generalize towards unseen data, i.e., pairs of objects for which at least one of the two objects was not used to construct the model. As a result of the transition from predictive models for single objects to pairs of objects, new advanced learning algorithms need to be developed, resulting in new challenges with regard to model construction, computational tractability and model assessment.

As relations between objects can be observed in many different forms, this general problem setting provides links to several subfields of machine learning, like statistical relational learning [5], graph mining [6], metric learning [7] and preference learning [8]. More specifically, from a graph-theoretic perspective, learning a relation can be formulated as learning edges in a graph where the nodes represent information of the data objects; from a metric learning perspective, the relation that we aim to learn should satisfy some well-defined properties like positive definiteness, transitivity or the triangle inequality; and from a preference learning perspective, the relation expresses a (degree of) preference in a pairwise comparison of data objects.

The topic of learning relations between objects is also closely related to recent developments in fuzzy set theory. This article will elaborate on these connections via two important contributions: (1) the extension of the typical setting of learning crisp relations to valued relations and (2) the inclusion of domain knowledge about relations into the inference process by explicit modeling of mathematical properties of these relations. For algorithmic simplicity, one can observe that many approaches only learn crisp relations, that is relations with only 0 and 1 as possible values, so that standard binary classifiers can be modified. Think in this context for example at inferring protein-protein interaction networks or metabolic networks in bioinformatics [2, 9].

However, not crisp but graded relations are observed in many real-world applications [10], resulting in a need for new algorithms that take graded relational information into account. Furthermore, the properties of valued relations have been investigated intensively in the recent fuzzy logic literature¹, and these properties are very useful to analyze and improve current algorithms. Using mathematical properties of valued relations, constraints can be imposed for incorporating domain knowledge in the learning process, to improve predictive performance or simply to guarantee that a relation with the right properties is learned. This is definitely the case for proper-

¹ Often the term fuzzy relation is used in the fuzzy set literature to refer to valued relations. However, fuzzy relations should be seen as a subclass of valued relations. For example, reciprocal relations should not be considered as fuzzy relations, because they often exhibit a probabilistic semantics rather than a fuzzy semantics.

ties like transitivity when learning similarity relations and preference relations – see e.g. [11, 12], but even very basic properties like symmetry, antisymmetry or reciprocity already provide domain knowledge that can steer the learning process. For example, in social network analysis, the notion “person A being a friend of person B” should be considered as a symmetric relation, while the notion “person A wins from person B in a chess game” will be antisymmetric (or, equivalently, reciprocal). Nevertheless, many examples exist too where neither symmetry nor antisymmetry necessarily hold, like the notion “person A trusts person B”.

In this paper we present a general kernel-based approach that unifies all the above cases into one general framework where domain knowledge can be easily specified by choosing a proper kernel and model structure, while different learning settings are distinguished by means of the loss function. From this perspective, one can make a subdivision between learning crisp relations, ordinal relations and $[0, 1]$ -valued relations. Furthermore, one can integrate in our framework different types of domain knowledge, by guaranteeing that certain properties are satisfied. Apart from the general case of arbitrary binary relations, we will specifically emphasize the prediction of reciprocal and symmetric relations.

2 General framework

The framework that we propose strongly relies on graphs, where nodes represent the data objects that are studied and the edges represent the relations present in the training set. The weights on the edges characterize the values of known relations, while unconnected nodes indicate pairs of objects for which the unknown relation needs to be predicted.

Let us start with introducing some notations. We assume that the data is structured as a graph $G = (\mathcal{V}, \mathcal{E}, Q)$, where \mathcal{V} corresponds to the set of nodes v and $\mathcal{E} \subseteq \mathcal{V}^2$ represents the set of edges e , for which training labels are provided in terms of relations. Moreover, these relations are represented by training weights y_e on the edges, generated from an unknown underlying relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$. Relations are required to take values in the interval $[0, 1]$ because some properties that we need are historically defined for such relations, but an extension to real-valued relations $h : \mathcal{V}^2 \rightarrow \mathbb{R}$ can always be realized with a simple increasing mapping $\sigma : \mathbb{R} \rightarrow [0, 1]$ such that

$$Q(v, v') = \sigma(h(v, v')), \quad \forall (v, v') \in \mathcal{V}^2. \quad (1)$$

Following the standard notations for kernel methods, we formulate our learning problem as the selection of a suitable function $h \in \mathcal{H}$, with \mathcal{H} a certain hypothesis space, in particular a reproducing kernel Hilbert space (RKHS). More specifically, the RKHS supports in our case hypotheses $h : \mathcal{V}^2 \rightarrow \mathbb{R}$ denoted as

$$h(e) = \mathbf{w}^T \Phi(e),$$

with \mathbf{w} a vector of parameters that needs to be estimated from training data, Φ a joint feature mapping for edges in the graph (see below) and \mathbf{a}^T the transpose of a vector \mathbf{a} . Let us denote a training dataset of cardinality $q = |\mathcal{E}|$ as a set $T = \{(e, y_e) \mid e \in \mathcal{E}\}$ of input-label pairs, then we formally consider the following optimization problem, in which we select an appropriate hypothesis h from \mathcal{H} for training data T :

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{q} \sum_{e \in \mathcal{E}} \mathcal{L}(h(e), y_e) + \lambda \|h\|_{\mathcal{H}}^2 \quad (2)$$

with \mathcal{L} a given loss function, $\|\cdot\|_{\mathcal{H}}^2$ the traditional quadratic regularizer on the RKHS and $\lambda > 0$ a regularization parameter. According to the representer theorem [13], any minimizer $h \in \mathcal{H}$ of (2) admits a dual representation of the following form:

$$h(\bar{e}) = \mathbf{w}^T \Phi(\bar{e}) = \sum_{e \in \mathcal{E}} a_e K^\Phi(e, \bar{e}), \quad (3)$$

with $a_e \in \mathbb{R}$ dual parameters, K^Φ the kernel function associated with the RKHS and Φ the feature mapping corresponding to K^Φ and

$$\mathbf{w} = \sum_{e \in \mathcal{E}} a_e \Phi(e).$$

We will alternate several times between the primal and dual representation for h in the remainder of this article.

The primal representation as defined in (2) and its dual equivalent (3) yield an RKHS defined on edges in the graph. In addition, we will establish an RKHS defined on nodes, as every edge consists of a couple of nodes. Given an input space \mathcal{V} and a kernel $K : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$, the RKHS associated with K can be considered as the completion of

$$\left\{ f \in \mathbb{R}^{\mathcal{V}} \mid f(v) = \sum_{i=1}^m \beta_i K(v, v_i) \right\},$$

in the norm

$$\|f\|_K = \sqrt{\sum_{i,j} \beta_i \beta_j K(v_i, v_j)},$$

where $\beta_i \in \mathbb{R}, m \in \mathbb{N}, v_i \in \mathcal{V}$.

As mentioned in the introduction, both crisp and valued relations can be handled by our framework. To make a subdivision between different cases, a loss function needs to be specified. For crisp relations, one can typically use the hinge loss, which is given by:

$$\mathcal{L}(h(e), y) = [1 - yh(e)]_+,$$

with $[\cdot]_+$ the positive part of the argument. Alternatively, one can opt to optimize a probabilistic loss function like the logistic loss:

$$\mathcal{L}(h(e), y) = \ln(1 + \exp(-yh(e))).$$

Conversely, if the observed relations in a given application are valued instead of crisp, other loss functions have to be considered. Further below, we will run experiments with a least-squares loss function:

$$\mathcal{L}(h(e), y) = (y_e - h(e))^2, \quad (4)$$

resulting in a regression type of learning setting. Alternatively, one could prefer to optimize a more robust regression loss like the ε -insensitive loss, in case outliers are expected in the training dataset.

So far, our framework does not differ from standard classification and regression algorithms. However, the specification of a more precise model structure for (2) offers a couple of new challenges. In the most general case, when no further restrictions on the underlying relation can be specified, the following Kronecker product feature mapping is proposed to express pairwise interactions between features of nodes:

$$\Phi(e) = \Phi(v, v') = \phi(v) \otimes \phi(v'),$$

where ϕ represents the feature mapping for individual nodes. Remark that in general the Kronecker product of two matrices \mathbf{M} and \mathbf{N} is defined as

$$\mathbf{M} \otimes \mathbf{N} = \begin{pmatrix} \mathbf{M}_{1,1}\mathbf{N} & \cdots & \mathbf{M}_{1,n}\mathbf{N} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m,1}\mathbf{N} & \cdots & \mathbf{M}_{m,n}\mathbf{N} \end{pmatrix}.$$

As first shown in [14], the Kronecker product pairwise feature mapping yields the Kronecker product edge kernel (a.k.a. the tensor product pairwise kernel) in the dual representation:

$$K_{\otimes}^{\Phi}(e, \bar{e}) = K_{\otimes}^{\Phi}(v, v', \bar{v}, \bar{v}') = K^{\phi}(v, \bar{v})K^{\phi}(v', \bar{v}'), \quad (5)$$

with K^{ϕ} the kernel corresponding to ϕ . With an appropriate choice for K^{ϕ} , such as the Gaussian RBF kernel, the kernel K^{Φ} generates a class \mathcal{H} of universally approximating functions for learning any type of relation (formal proof omitted).

3 Special relations

If no further information is available about the relation that underlies the data, one should definitely use the Kronecker product edge kernel. In this most general case, we allow that for any pair of nodes in the graph several edges can exist, in which an edge in one direction not necessarily imposes constraints on the edge in the opposite direction and multiple edges in the same direction can connect two nodes. This

construction is required to allow repeated measurements. However, two important subclasses of relations deserve further attention: reciprocal relations and symmetric relations. Let us start with the former.

Definition 1. A binary relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$ is called a reciprocal relation if for all $(v, v') \in \mathcal{V}^2$ it holds that $Q(v, v') = 1 - Q(v', v)$.

Given the increasing transformation (1), every reciprocal relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$ can be rewritten as an antisymmetric relation $h : \mathcal{V}^2 \rightarrow \mathbb{R}$, formally defined as follows.

Definition 2. A binary relation $h : \mathcal{V}^2 \rightarrow \mathbb{R}$ is called an antisymmetric relation if for all $(v, v') \in \mathcal{V}^2$ it holds that $h(v, v') = -h(v', v)$.

For reciprocal and antisymmetric relations, every edge $e = (v, v')$ induces in the multi-graph that was defined above an unobserved invisible edge $e_R = (v', v)$ with appropriate weight in the opposite direction. Applications arise here in domains such as preference learning, game theory and bioinformatics for representing preference relations, choice probabilities, winning probabilities, gene regulation, etc. The weight on the edge defines the real direction of such an edge. If the weight on the edge $e = (v, v')$ is higher than 0.5, then the direction is from v to v' , but when the weight is lower than 0.5, then the direction should be interpreted as inverted. If the relation is 3-valued as $Q : \mathcal{V}^2 \rightarrow \{0, 1/2, 1\}$, then we end up with a three-class ordinal regression setting instead of an ordinary regression setting. Interestingly, reciprocity can be easily incorporated in our framework.

Proposition 1. Let Ψ be a feature mapping on \mathcal{V}^2 , let $\sigma : \mathbb{R} \rightarrow [0, 1]$ be an increasing mapping and let h be a hypothesis defined by (2), then the relation Q of type (1) is reciprocal if Φ is given by

$$\Phi_R(e) = \Phi_R(v, v') = \Psi(v, v') - \Psi(v', v),$$

while σ satisfies $\sigma(1/2) = 0$ and $\sigma(x) = 1 - \sigma(-x)$ for all $x \in \mathbb{R}$.

The proof is immediate. In addition, one can easily show that reciprocity as domain knowledge can be enforced in the dual formulation. Let us in the least restrictive form now consider the Kronecker product for Ψ , then one obtains for Φ_R the kernel $K_{\otimes R}^\Phi$ given by

$$K_{\otimes R}^\Phi(e, \bar{e}) = 2(K^\Phi(v, \bar{v})K^\Phi(v', \bar{v}') - K^\Phi(v, \bar{v}')K^\Phi(v', \bar{v})). \quad (6)$$

Similar to the general case, one can show that this kernel can represent any type of reciprocal relation by means of universal approximation.

Symmetric relations form another important subclass of relations in our framework. As a specific type of symmetric relations, similarity relations constitute the underlying relation in many application domains where relations between objects need to be learned. Symmetric relations are formally defined as follows.

Definition 3. A binary relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$ is called a symmetric relation if for all $(v, v') \in \mathcal{V}^2$ it holds that $Q(v, v') = Q(v', v)$.

Definition 4. A binary relation $h : \mathcal{V}^2 \rightarrow \mathbb{R}$ is called a symmetric relation if for all $(v, v') \in \mathcal{V}^2$ it holds that $h(v, v') = h(v', v)$.

For symmetric relations, edges in the multi-graph introduced above become undirected. Applications arise in many domains and metric learning or learning similarity measures can be seen as special cases. If the relation is 2-valued as $Q : \mathcal{V}^2 \rightarrow \{0, 1\}$, then we end up with a classification setting instead of a regression setting. Just like reciprocal relations, it turns out that symmetry can be easily incorporated in our framework.

Proposition 2. *Let Ψ be a feature mapping on \mathcal{V}^2 , let $\sigma : \mathbb{R} \rightarrow [0, 1]$ be an increasing mapping and let h be a hypothesis defined by (2), then the relation Q of type (1) is symmetric if Φ is given by*

$$\Phi_S(e) = \Phi_S(v, v') = \Psi(v, v') + \Psi(v', v).$$

In addition, by using the mathematical properties of the Kronecker product, one obtains in the dual formulation an edge kernel that looks very similar to the one derived for reciprocal relations. Let us again consider the Kronecker product for Ψ , then one obtains for Φ_S the kernel $K_{\otimes_S}^\Phi$ given by $K_{\otimes_S}^\Phi(e, \bar{e}) =$

$$2(K^\phi(v, \bar{v})K^\phi(v', \bar{v}') + K^\phi(v, \bar{v}')K^\phi(v', \bar{v})).$$

Thus, the subtraction of kernels in the reciprocal case becomes an addition of kernels in the symmetric case. The above kernel has been used for predicting protein-protein interactions in bioinformatics [14]. Unlike many existing kernel-based methods for pairwise data, the models obtained with these kernels are able to represent any reciprocal or symmetric relation respectively, without imposing additional transitivity properties of the relations.

We also remark that for symmetry as well, one can prove that the Kronecker product edge kernel yields a model that is flexible enough to represent any type of underlying symmetric relation.

4 Relationships with fuzzy set theory

The previous section revealed that specific Kronecker product edge kernels can be constructed for modeling reciprocal and symmetric relations, without requiring any further background about these relations. In this section we demonstrate that the Kronecker product edge kernels K_{\otimes}^Φ , $K_{\otimes_R}^\Phi$ and $K_{\otimes_S}^\Phi$ are particularly useful for modeling intransitive relations, which occur in a lot of real-world scenarios, like game playing [15, 16], competition between bacteria [17, 18, 19, 20, 21, 22] and fungi [23], mating choice of lizards [24] and food choice of birds [25], to name just a few.

Despite the occurrence of intransitive relations in many domains, one has to admit that most applications are still characterized by relations that fulfill relatively

strong transitivity requirements. For example, in decision making, preference modeling and social choice theory, one can argue that reciprocal relations like choice probabilities and preference judgments should satisfy certain transitivity properties, if they represent rational human decisions made after well-reasoned comparisons on objects [26, 27, 28]. For symmetric relations as well, transitivity plays an important role [29, 30], when modeling similarity relations, metrics, kernels, etc.

It is for this reason that transitivity properties have been studied extensively in fuzzy set theory and related fields. For reciprocal relations, one can distinguish the notions of stochastic transitivity [26], FG-transitivity [31] and the more general recent framework of cycle transitivity [32, 12]. For valued symmetric relations, the notion of T -transitivity has been put forward [33, 34]. In addition, several authors have shown that various forms of transitivity give rise to utility representable or numerically representable relations, also called fuzzy weak orders – see e.g. [26, 35, 36, 37, 38]. We will use the term ranking representability to establish a link with machine learning. We give a slightly specific definition that unifies reciprocal and symmetric relations.

Definition 5. A reciprocal or symmetric relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$ is called ranking representable if there exists a ranking function $f : \mathcal{V} \rightarrow \mathbb{R}$ and an increasing mapping $\sigma : \mathbb{R} \rightarrow [0, 1]$ such that for all pairs $(v, v') \in \mathcal{V}^2$ it respectively holds that

1. $Q(v, v') = \sigma(f(v) - f(v'))$ (reciprocal case) ;
2. $Q(v, v') = \sigma(f(v) + f(v'))$ (symmetric case) .

The main idea is that ranking representable relations can be constructed from a utility function f . Ranking representable reciprocal relations correspond to directed acyclic graphs, and a unique ranking of the nodes in such graphs can be obtained with topological sorting algorithms.

Interestingly, ranking representability of reciprocal relations and symmetric relations can be easily achieved in our framework by simplifying the joint feature mapping Ψ . Let $\Psi(v, v') = \phi(v)$ such that K^Φ simplifies to

$$\begin{aligned} K_{fR}^\Phi(e, \bar{e}) &= K^\Phi(v, \bar{v}) + K^\Phi(v', \bar{v}') - K^\Phi(v, \bar{v}') - K^\Phi(v', \bar{v}), \\ K_{fS}^\Phi(e, \bar{e}) &= K^\Phi(v, \bar{v}) + K^\Phi(v', \bar{v}') + K^\Phi(v, \bar{v}') + K^\Phi(v', \bar{v}), \end{aligned}$$

when $\Phi(v, v') = \Phi_R(v, v')$ or $\Phi(v, v') = \Phi_S(v, v')$, respectively, then the following proposition holds.

Proposition 3. *The relation $Q : \mathcal{V}^2 \rightarrow [0, 1]$ given by (1) and h defined by (2) with $K^\Phi = K_{fR}^\Phi$ (respectively $K^\Phi = K_{fS}^\Phi$) is a ranking representable reciprocal (respectively symmetric) relation.*

The proof directly follows from the fact that for this specific kernel, $h(v, v')$ can be respectively written as $f(v) - f(v')$ and $f(v) + f(v')$. The kernel K_{fR}^Φ has been initially introduced in [39] for ordinal regression and during the last decade it has been extensively used as main building block in many kernel-based ranking algorithms. Since ranking representability of reciprocal relations implies strong stochastic transitivity of reciprocal relations, K_{fR}^Φ can represent this type of domain knowledge.

The notion of ranking representability is powerful for reciprocal relations, because the majority of reciprocal relations satisfy this property, but for symmetric relations it has a rather limited applicability. Ranking representability as defined above cannot represent relations that originate from an underlying metric or similarity measure. For such relations, one needs another connection with its roots in Euclidean metric spaces [29].

5 An illustration in document retrieval

In the experiments, we test the ability of the pairwise kernels to model different relations, and the effect of enforcing prior knowledge about the properties of the learned relations. To this end, we train the regularized least-squares (RLS) algorithm to regress the relation values [40]. Extensive empirical results have been reported for reciprocal relations in [41], as a consequence we focus in this article on symmetric relations. To this end, we compare the ordinary and symmetric Kronecker kernels on a real-world data set based on newsgroups documents². The data is sampled from 4 newsgroups: rec.autos, rec.sport.baseball, comp.sys.ibm.pc.hardware and comp.windows.x. The aim is to learn to predict the similarity of two documents as measured by the number of common words they share. The node features correspond to the number of occurrences of a word in a document. Unlike previously reported experiments, the feature representation is very high-dimensional and sparse, as there are more than 50000 possible features, the majority of which are zero for any given document. First, we sample separate training, validation and test sets each consisting of 1000 nodes. Second, we sample edges connecting the nodes in the training and validation set using exponentially growing sample sizes to measure the effect of sample size on the differences between the kernels. The sample size grid is [100, 200, 400, ..., 102400]. Again, we sample only edges with different starting and end nodes. When computing the test performance, we consider all the edges in the test set, except those starting and ending at the same node. We train the RLS algorithm using conjugate gradient optimization with early stopping [42], optimization is terminated once the MSE on the validation set has failed to decrease for 10 consecutive iterations. The mean predictor achieves around 145 MSE test performance on this data.

The results are presented in Figure 1. Even for 100 pairs the errors are for both kernels much lower than the mean predictor results, showing that the RLS algorithm succeeds with both kernels in learning the underlying relation. Increasing the training set size leads to a decrease in test error. Using the prior knowledge about the symmetry of the learned relation is clearly helpful. The symmetric kernel achieves for all sample sizes a lower error than the ordinary Kronecker kernel and the largest differences are observed for the smallest sample sizes. For 100 training instances, the error is almost halved by enforcing symmetry.

² Available at: <http://people.csail.mit.edu/jrennie/20Newsgroups/>

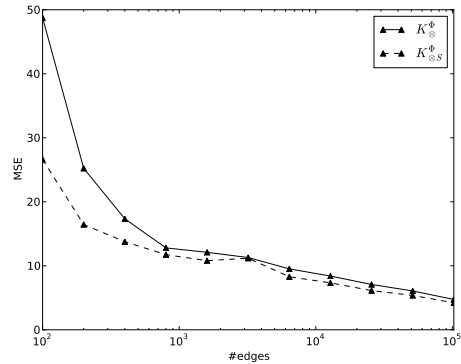


Fig. 1 The comparison of the ordinary Kronecker product pairwise kernel K_{\otimes}^{Φ} and the symmetric Kronecker product pairwise kernel $K_{\otimes S}^{\Phi}$ on the Newsgroups dataset. The mean squared error is shown as a function of the training set size.

6 Conclusion

A general kernel-based framework for learning various types of valued relations was presented in this article. This framework extends existing approaches for learning relations, because it can handle crisp and valued relations. A Kronecker product feature mapping was proposed for combining the features of pairs of objects that constitute a relation (edge level in a graph). In addition, we clarified that domain knowledge about the relation to be learned can be easily incorporated in our framework, such as reciprocity and symmetry properties. Experimental results on synthetic and real-world data clearly demonstrate that this domain knowledge really helps in improving the generalization performance. Moreover, important links with recent developments in fuzzy set theory and decision theory can be established, by looking at transitivity properties of relations.

Acknowledgments

W.W. is supported as a postdoc by the Research Foundation of Flanders (FWO Vlaanderen) and T.P. by the Academy of Finland (grant 134020).

References

1. M. Bowling, J. Fürnkranz, T. Graepel, and R. Musick, “Machine learning and games,” *Machine Learning*, vol. 63, no. 3, pp. 211–215, 2006.

2. Y. Yamanishi, J. Vert, and M. Kanehisa, "Protein network inference from multiple genomic data: a supervised approach," *Bioinformatics*, vol. 20, pp. 1363–1370, 2004.
3. Y. Yang, N. Bansal, W. Dakka, P. Ipeirotis, N. Koudas, and D. Papadias, "Query by document," in *Proceedings of the Second ACM International Conference on Web Search and Data Mining, Barcelona, Spain*, 2009, pp. 34–43.
4. B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," in *Advances in Neural Information Processing Systems*, 2004.
5. L. De Raedt, *Logical and Relational Learning*. Springer, 2009.
6. J.-P. Vert and Y. Yamanishi, "Supervised graph inference," in *Advances in Neural Information Processing Systems*, vol. 17, 2005.
7. E. Xing, A. Ng, M. Jordan, and S. Russell, "Distance metric learning with application to clustering with side information," in *Advances in Neural Information Processing Systems*, vol. 16, 2002, pp. 521–528.
8. E. Hüllermeier and J. Fürnkranz, *Preference Learning*. Springer, 2010.
9. P. Geurts, N. Touleimat, M. Dutreix, and F. d'Alché-Buc, "Inferring biological networks with output kernel trees," *BMC Bioinformatics*, vol. 8, no. 2, p. S4, 2007.
10. J. Doignon, B. Monjardet, M. Roubens, and P. Vincke, "Biororder families, valued relations and preference modelling," *Journal of Mathematical Psychology*, vol. 30, pp. 435–480, 1986.
11. Z. Switalski, "Transitivity of fuzzy preference relations - an empirical study," *Fuzzy Sets and Systems*, vol. 118, pp. 503–508, 2000.
12. B. De Baets, H. De Meyer, B. De Schuymer, and S. Jenei, "Cyclic evaluation of transitivity of reciprocal relations," *Social Choice and Welfare*, vol. 26, pp. 217–238, 2006.
13. B. Schölkopf and A. Smola, *Learning with Kernels, Support Vector Machines, Regularisation, Optimization and Beyond*. The MIT Press, 2002.
14. A. Ben-Hur and W. Noble, "Kernel methods for predicting protein-protein interactions," *Bioinformatics*, vol. 21 Suppl 1, pp. 38–46, 2005.
15. B. De Schuymer, H. De Meyer, B. De Baets, and S. Jenei, "On the cycle-transitivity of the dice model," *Theory and Decision*, vol. 54, pp. 261–285, 2003.
16. L. Fisher, *Rock, Paper, Scissors: Game Theory in Everyday Life*. Basic Books, 2008.
17. B. Kerr, M. Riley, M. Feldman, and B. Bohannan, "Local dispersal promotes biodiversity in a real-life game of rock paper scissors," *Nature*, vol. 418, pp. 171–174, 2002.
18. T. Czárán, R. Hoekstra, and L. Pagie, "Chemical warfare between microbes promotes biodiversity," *Proceedings of the National Academy of Sciences*, vol. 99, no. 2, pp. 786–790, 2002.
19. M. Nowak, "Biodiversity: Bacterial game dynamics," *Nature*, vol. 418, pp. 138–139, 2002.
20. B. Kirkup and M. Riley, "Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo," *Nature*, vol. 428, pp. 412–414, 2004.
21. G. Károlyi, Z. Neufeld, and I. Scheuring, "Rock-scissors-paper game in a chaotic flow: The effect of dispersion on the cyclic competition of microorganisms," *Journal of Theoretical Biology*, vol. 236, no. 1, pp. 12–20, 2005.
22. T. Reichenbach, M. Mobilia, and E. Frey, "Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games," *Nature*, vol. 448, pp. 1046–1049, 2007.
23. L. Boddy, "Interspecific combative interactions between wood-decaying basidiomycetes," *FEMS Microbiology Ecology*, vol. 31, pp. 185–194, 2000.
24. S. Sinervo and C. Lively, "The rock-paper-scissors game and the evolution of alternative mate strategies," *Nature*, vol. 340, pp. 240–246, 1996.
25. T. Waite, "Intransitive preferences in hoarding gray jays (*Perisoreus canadensis*)," *Journal of Behavioural Ecology and Sociobiology*, vol. 50, pp. 116–121, 2001.
26. R. Luce and P. Suppes, *Handbook of Mathematical Psychology*. Wiley, 1965, ch. Preference, Utility and Subjective Probability, pp. 249–410.
27. P. Fishburn, "Nontransitive preferences in decision theory," *Journal of Risk and Uncertainty*, vol. 4, pp. 113–134, 1991.
28. A. Tversky, *Preference, Belief and Similarity*, E. Shafir, Ed. MIT Press, 1998.
29. J. Gower and P. Legendre, "Metric and Euclidean properties of dissimilarity coefficients," *Journal of Classification*, vol. 3, pp. 5–48, 1986.

30. F. Jäkel, B. Schölkopf, and F. Wichmann, "Similarity, kernels, and the triangle inequality," *Journal of Mathematical Psychology*, vol. 52, no. 2, pp. 297–303, 2008.
31. Z. Switalski, "General transitivity conditions for fuzzy reciprocal preference matrices," *Fuzzy Sets and Systems*, vol. 137, pp. 85–100, 2003.
32. B. De Baets and H. De Meyer, "Transitivity frameworks for reciprocal relations: cycle-transitivity versus FG -transitivity," *Fuzzy Sets and Systems*, vol. 152, pp. 249–270, 2005.
33. B. De Baets and R. Mesiar, "Metrics and T -equalities," *Journal of Mathematical Analysis and Applications*, vol. 267, pp. 531–547, 2002.
34. B. Moser, "On representing and generating kernels by fuzzy equivalence relations," *Journal of Machine Learning Research*, vol. 7, pp. 2603–2620, 2006.
35. A. Billot, "An existence theorem for fuzzy utility functions: A new elementary proof," *Fuzzy Sets and Systems*, vol. 74, pp. 271–276, 1995.
36. M. Koppen, "Random utility representation of binary choice probabilities: Critical graphs yielding critical necessary conditions," *Journal of Mathematical Psychology*, vol. 39, pp. 21–39, 1995.
37. L. Fono and N. Andjiga, "Utility function of fuzzy preferences on a countable set under max*-transitivity," *Social Choice and Welfare*, vol. 28, pp. 667–683, 2007.
38. U. Bodenhofer, B. De Baets, and J. Fodor, "A compendium of fuzzy weak orders," *Fuzzy Sets and Systems*, vol. 158, pp. 811–829, 2007.
39. R. Herbrich, T. Graepel, and K. Obermayer, "Large margin rank boundaries for ordinal regression," in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. MIT Press, 2000, pp. 115–132.
40. T. Pahikkala, E. Tsivtsivadze, A. Airola, J. Järvinen, and J. Boberg, "An efficient algorithm for learning to rank from preference graphs," *Machine Learning*, vol. 75, no. 1, pp. 129–165, 2009.
41. T. Pahikkala, W. Waegeman, E. Tsivtsivadze, T. Salakoski, and B. De Baets, "Learning intransitive reciprocal relations with kernel methods," *European Journal of Operational Research*, vol. 206, pp. 676–685, 2010.
42. T. Pahikkala, W. Waegeman, A. Airola, T. Salakoski, and B. De Baets, "Conditional ranking on relational data," in *Proceedings of the European Conference on Machine Learning*, ser. Lecture Notes in Computer Science, J. Balczar, F. Bonchi, A. Gionis, and M. Sebag, Eds. Springer Berlin / Heidelberg, 2010, vol. 6322, pp. 499–514.