

Efficient Machine Learning Methods over Pairwise Space (keynote)

Hung Son Nguyen

University of Warsaw

Keywords

Rough sets, Support Vector Machine, Factorization Machine, Distance Metric Learning, Context-Aware Recommendation

Extended Abstract

In recent years many machine learning concepts and methods were developed on the set of pairs of objects. In this paper, the set of all pairs of objects is called *the pairwise space*. Let us notice that if the set of objects $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ consists of n instances, then the pairwise space contains $O(n^2)$ pairs. Thus why the straightforward implementations of those methods are not applicable for big data sets with millions of objects.

The main concepts in rough set theory (RS) such as reducts, lower and upper approximations, decision rules or discretizations have been defined in term of the discernibility matrix, which is a form of the pairwise space [1]. For example, in minimal decision reduct problem, we are looking for the minimal subset of features that preserves the discernibility between objects from different decision classes [2].

Support Vector Machine (SVM) is also a classification method described as an optimization problem over the pairwise space [3]. The initial idea of looking for the linear classifier with the maximal margin were transformed into the problem of looking for a set of coefficients $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ related to objects that maximizes an objective function

$$W(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j K(x_i, x_j).$$

defined on the set of dot products of all pairs of objects. In the above formula y_i denotes the decision class of the object \mathbf{x}_i and K is a kernel function chosen by the user.

Distance Metric Learning (DML) [4] is a machine learning discipline that looks for the best distance function (also divergence or similarity) from certain available information about similarity measures between different pairs or triplets of data. These similarities are determined

29th International Workshop on Concurrency, Specification and Programming (CS&P'21)

✉ son@mimuw.edu.pl (H. S. Nguyen)

🆔 0000-0002-3236-5456 (H. S. Nguyen)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

by the sets

$$\begin{aligned} S &= \{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X} \times \mathcal{X} : \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are similar.}\} \\ D &= \{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X} \times \mathcal{X} : \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not similar.}\} \\ R &= \{(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_l) \in \mathcal{X} \times \mathcal{X} \times \mathcal{X} : \mathbf{x}_i \text{ is more similar to } \mathbf{x}_j \text{ than to } \mathbf{x}_l.\} \end{aligned}$$

With these data and similarity constraints, the problem is to look for those distance functions (belonging to a predefined family of distances \mathcal{D}) that minimize a certain loss function ℓ determined on the base of the sets S, D and R . In other words, the objective of DML is to solve the optimization problem

$$\min_{d \in \mathcal{D}} \ell(d, S, D, R)$$

In recent years, many efficient implementations for the mentioned above disciplines have been proposed and developed. Most of them are based either on the approximation idea [5], [6] or deep learning [7].

Context aware recommendation systems is another machine learning concept that were defined on the pairwise space which was in fact defined as a regression problem over the set of transaction pairs [8].

In this talk we compare different techniques for the mentioned above machine learning concepts and we will pay an attention on application of factorization machine (FM). This method has been successfully applied for context aware recommendation systems [9]. The main idea is to transform the optimization problem established on the pairwise space into an equivalent problem where the time complexity for each iteration has been reduced from quadratic time into linear time [10]. We will show that factorization machine can be also applied for some problems in rough set theory, SVM or distance metric learning.

References

- [1] Z. Pawlak, Rough sets, *International Journal of Information and Computer Sciences* 11 (1982) 341–356.
- [2] Z. Pawlak, A. Skowron, Rudiments of rough sets, *Information Sciences* 177 (2007) 3–27.
- [3] C. Cortes, V. Vapnik, Support vector networks, *Machine Learning* 20 (1995) 273–297.
- [4] J.-L. Suárez, S. García, F. Herrera, A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges, *Neurocomputing* 425 (2021) 300–322.
- [5] H. S. Nguyen, *Approximate Boolean Reasoning: Foundations and Applications in Data Mining*, Springer-Verlag, Berlin, Heidelberg, 2006, p. 334–506.
- [6] T. Joachims, *Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms*, Kluwer/Springer, 2002.
- [7] P. H. Barros, F. Queiroz, F. Figueredo, J. A. dos Santos, H. S. Ramos, A new similarity space tailored for supervised deep metric learning, *CoRR abs/2011.08325* (2020). URL: <https://arxiv.org/abs/2011.08325>. arXiv:2011.08325.

- [8] S. Rendle, Z. Gantner, C. Freudenthaler, L. Schmidt-Thieme, Fast context-aware recommendations with factorization machines, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 635–644. URL: <https://doi.org/10.1145/2009916.2010002>. doi:10.1145/2009916.2010002.
- [9] X. Xin, B. Chen, X. He, D. Wang, Y. Ding, J. Jose, Cfm: Convolutional factorization machines for context-aware recommendation, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, 2019, pp. 3926–3932. URL: <https://doi.org/10.24963/ijcai.2019/545>. doi:10.24963/ijcai.2019/545.
- [10] S. Rendle, Factorization machines with libfm, ACM Trans. Intell. Syst. Technol. 3 (2012) 57:1–57:22. URL: <http://doi.acm.org/10.1145/2168752.2168771>. doi:10.1145/2168752.2168771.