MŰEGYETEM 1782

Convex polyhedron learning and its applications

Thesis booklet

Gábor Takács

Advisor: Béla Pataki, PhD

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Measurement and Information Systems

Contents

1 Research outline

From a possible engineer's point of view, *learning* can be considered as discovering the relationship between the features of a phenomenon. If the observations about the phenomenon are available as data, and the relationship between the features is discovered by a program, then we speak about *machine learning* (data mining). The machine learning approach fits well to such problems, in which it is difficult to formalize the connection between input and the desired output, but it is relatively easy to collect corresponding input–output pairs (e.g. spam filtering, credit risk analysis, face recognition, driving a car).

Two interesting special cases of learning are classification and collaborative filtering.

- In the problem of *classification* [1] the phenomenon is modeled by a random pair $(\mathbf{X}, Y)$, where the d-dimensional, continuous $\mathbf{X}$ is called input, and the discrete (often binary) Y is called label. The goal is to find a function g such that the error probability $\mathbf{P}\{g(\mathbf{X}) \neq Y\}$ is as small as possible.

- In the problem of *collaborative filtering* [2] the phenomenon is modeled by a random triplet (U, I, R), where the discrete U is called user identifier, the discrete I is called item identifier, and the continuous R is called rating value. The goal is to find a function g such that the mean squared error $\mathbf{E}\{(g(U, I) - R)^2\}$ is as small as possible. Collaborative filtering has moved into the focus of research in 2006, due to the 1 million dollar Netflix Prize [3].

The distribution of the random vector that models the phenomenon is usually not known. Instead of this we only have a finite *training set* that contains examples that were drawn according to the unknown distribution.

1.1 Convex polyhedron classification

Many classification problems arising in practice are *unbalanced* in the sense that one class label occurs much more frequently than the others. Such problems are quite common e.g. in the field of medical and technical diagnostics. For example, in the case of computer-aided breast cancer screening the vast majority of the patients are healthy, therefore the classification problem to be solved by the expert system is unbalanced.

Convex polyhedron classification is a special approach within classification that fits well to unbalanced problems. A convex polyhedron classifier is a function $g : \mathbb{R}^d \mapsto \{+1, -1\}$ with the property that the decision region $\{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) = 1\}$ is a convex polyhedron (the intersection of finitely many closed half-spaces).

A convex polyhedron classifier can be given for example by K hyperplanes. When classifying an input $\mathbf{x} \in \mathbb{R}^d$, one has to substitute $\mathbf{x}$ into the linear functions defining the hyperplanes. If $\mathbf{x}$ is on the negative side of any of the hyperplanes, then the function value $g(\mathbf{x})$ will be -1. This means that the calculation can be stopped immediately, if we get a negative number in any of the substitutions.

A main goal of my research was to give efficient and accurate algorithms for training convex polyhedron classifiers. Such methods were not known before, and therefore convex polyhedron classifiers were rarely used in practice.

1.2 Convex separability

An interesting problem related to convex polyhedron classification is determining the convex separability of point sets. The formal definition of the problem is the following: Let $\mathcal{P}$ and $\mathcal{Q}$ be

finite point sets in $\mathbb{R}^d$. The task is to decide if there exists a convex polyhedron that contains every elements from $\mathcal{P}$ and no elements from $\mathcal{Q}$.

In practice, determining the convex separability can be useful e.g. in the initial, data exploration phase of a data mining project. Analyzing the properties of the training set can help in choosing the right classification algorithm for the problem.

1.3 Collaborative filtering with convex polyhedrons

The applicability of convex polyhedron based methods is not restricted to classification of course. The approach has its analogue version e.g. for collaborative filtering too. However, in this case we lose the advantage of (extra) fast prediction, since we have to iterate over all hyperplanes for calculating the output. The utility of convex polyhedron based methods for collaborative filtering is that they solve the problem in a unique way, and they can be useful members of a blended solution consisting of many models.

1.4 Model complexity

From a point of view, machine learning can be seen as modeling. The input is a dataset that was collected by observing a phenomenon. The output is a model that explains certain aspects of the phenomenon, and that can be used for making prediction.

In a typical data mining project many experiments are performed and many models are created. It is non-trivial to decide which of them should be used for prediction in the final system. Obviously, if two models achieve the same accuracy on the training set, then it is reasonable to choose the simpler one. The question is how to characterize the complexity of machine learning models in a well-defined way.

The Vapnik-Chervonenkis dimension [4] is a widely accepted model complexity measure for binary classifiers. With the help of this concept it is possible to obtain a probabilistic bound for the error probability of a classifier, without using an independent test set.

2 Organization of the dissertation

The first chapter of the dissertation (Introduction) briefly introduces the field of machine learning and locates convex polyhedron learning in it. Then, without completeness it overviews a set known learning algorithms. The main selection criterion of the known algorithms was the degree of connection to the rest of the thesis. The part dealing with collaborative filtering contains novel results too.

The second chapter of the dissertation (Algorithms) is about algorithms that use convex polyhedrons for solving various machine learning problems. The first part of the chapter deals with the problem of linear and convex separation. The second part of the chapter gives algorithms for training convex polyhedron classifiers. The third part of the chapter introduces a convex polyhedron based algorithm for collaborative filtering. The first two parts contain known and novel methods too. The third part contains only new results, since in the case of collaborative filtering known convex polyhedron based approaches are not available.

The third chapter of the dissertation (Model complexity) collects the known facts about the Vapnik–Chervonenkis dimension of convex polyhedron classifiers and proves new results. The fourth chapter of the dissertation (Applications) presents the experiments performed with the given algorithms on real and artificial datasets.

3 Summary of new scientific results

- **Thesis group 1**: I gave new algorithms for determining the linear and the convex separability of point sets. I demonstrated it by experiments that the proposed algorithms are faster than conventional ones on numerous test problems. The results of the thesis group appeared in the following publications: [P10], [P11], [P24].

[T1.1] I investigated the possibility of deciding linear separability with an incremental method. I introduced a linear programming based direct method (LSEP$_2$) that has favorable properties for being applied as a component of an incremental solution. I proposed heuristics (LSEPX, LSEPY, LSEPZX, LSEPZY) for choosing the active constraints and variables of the next step.

[T1.2] I gave a novel approximate algorithm with low time requirement for the approximate convex separation of two point sets (CSEPC).

[T1.3] I gave a novel exact algorithm with low expected time requirement for determining the convex separability of two point sets (CSEPX). The algorithm applies the previous method as a preprocessor.

- **Thesis group 2**: I worked out new, derivative-based algorithms for training convex polyhedron based classifiers and predictors. The results of the thesis group appeared in the following publications: [P22], [P23].

[T2.1] I investigated the possibility of approximating the maximum operator by smooth functions, and I introduced six, parameterizable smooth maximum function families.

[T2.2] I worked out a novel, smooth maximum function based algorithm family for training convex polyhedron classifiers (SMAX). I demonstrated the efficiency of the algorithms in terms of classification accuracy and running time by experiments on artificial and real datasets.

[T2.3] I gave a novel, smooth maximum function based algorithm for training convex polyhedron models in the case of collaborative filtering (SMAX$_{CF}$). I demonstrated the efficiency of the algorithm in terms of prediction accuracy and running time by experiments on the currently largest available benchmark database (Netflix).

- **Thesis group 3**: I achieved new results in determining the Vapnik–Chervonenkis dimension of convex polyhedron classifiers. The results of the thesis group appeared in the following publications: [P8], [P9].

[T3.1] I determined the Vapnik–Chervonenkis dimension of convex K-gon classifiers on the plane. It is important to note that I do not restrict the label of the inner (convex) region in advance. (That is an easier, solved problem.)

[T3.2] I proved a lower bound for the Vapnik–Chervonenkis dimension of d-dimensional convex K-polyhedron classifiers that is better than the trivial lower bound. In the special cases $d = 3$ and $d = 4$ I improved the bound further.

- **Thesis group 4**: I introduced novel, scalable, and accurate algorithms for collaborative filtering. The results of the thesis group appeared in the following publications: [P1], [P2], [P3], [P4], [P5], [P6].

[T4.1] I introduced a novel matrix factorization technique called BRISMF (biased regularized incremental simultaneous matrix factorization). The method has proved its efficiency in the Netflix Prize competition.

[T4.2] I gave a new training algorithm for Paterek's NSVD1 model [5], which is an important neighbor-based approach in collaborative filtering. The algorithm is significantly faster than the naive implementation of gradient descent, but the resulted model is exactly the same in the two cases.

4 Applicability of the results

The algorithms for determining the linear and convex separability can be useful in the initial, data exploration phase of data mining. Let us assume that we have a binary classification problem and a "sufficiently large" training set. Since the amount of resources available for solving the problem is limited, it is important to select good experiments to run.

If the classes are linearly separable in the training set, then it is reasonable to start experimenting with linear classifiers. If the classes are only convexly separable, then it is useful to start with convex polyhedron classifiers, and otherwise with general nonlinear classifiers. If we observe that the best approach is not that one that corresponds to the degree of separability of the training set, then this probably indicates noise or insufficient training set size.

The efficient algorithms for training convex polyhedron classifiers provide a new tool for data miners. The approach can be tried out for arbitrary tasks, and it can be particularly useful in the case of unbalanced problems.

The algorithms for collaborative filtering can be used for building recommender systems. If we have a predictor that is able to estimate the rating of a user for an item, then recommendation for that user can be implemented easily. We simply iterate over recommendable items and calculate the answer of the predictor for each of them. For the selected user we recommend items with the largest prediction values.

If the predictor used in the system is more accurate in predicting ratings, then relevant items will move up in the ranking list. Note that a small change in the prediction values can result a large change in the order of items, therefore it is often worthwhile to put effort into increasing the accuracy of the predictor.

The results about the Vapnik–Chervonenkis dimension of convex polyhedron classifiers can be useful in two ways. First, they make the estimation a bit more accurate for the error probability of the classifier. Second, they may give ideas to others for improving the lower and upper bounds further.

5   List of publications

[P1] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research* (Special Topic on Mining and Learning with Graphs and Relations), 10: 623–656, 2009.

[P2] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Matrix factorization and neighbor based algorithms for the Netflix Prize problem. *Proc. of the 2008 ACM Conference on Recommender Systems (RECSYS'08)*, pages 267–274, Lausanne, Switzerland, 2008.

[P3] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Investigation of various matrix factorization methods for large recommender systems. *Proc. of the 2nd KDD Workshop on Large Scale Recommender Systems and the Netflix Prize Competition*, Las Vegas, Nevada, USA, 2008.

[P4] G. Takács, I. Pilászy, B. Németh, and D. Tikk. A unified approach of factor models and neighbor based methods for large recommender systems. *Proc. of the 1th IEEE ICADIWT Workshop on Recommender Systems and Personalized Retrieval*, pages 186–191, Ostrava, Czech Republic, 2008.

[P5] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Major components of the Gravity Recommendation System. *ACM SIGKDD Explorations Newsletter*, 9(2): 80–83, 2007.

[P6] G. Takács, I. Pilászy, B. Németh, and D. Tikk. On the Gravity Recommendation System. *Proc. of the KDD Cup and Workshop 2007*, pages. 22–30, San Jose, California, USA, 2007.

[P7] G. Takács and B. Pataki. Case-level detection of mammographic masses. *International Journal of Applied Electromagnetics and Mechanics*, 25(1–4): 395–400, 2007.

[P8] G. Takács. The Vapnik–Chervonenkis dimension of convex n-gon classifiers. *Hungarian Electronic Journal of Sciences*, 2007.

[P9] G. Takács and B. Pataki. Lower bounds on the Vapnik–Chervonenkis dimension of convex polytope classifiers. *Proc. of the 11th International Conference on Intelligent Engineering Systems (INES 2007)*, Budapest, Hungary, 2007.

[P10] G. Takács and B. Pataki. Deciding the convex separability of pattern sets. *Proc. of the 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS'2007)*, Dortmund, Germany, 2007.

[P11] G. Takács and B. Pataki. An efficient algorithm for deciding the convex separability of point sets. *Proc. of the 14th PhD Mini-Symposium, Budapest University of Technology and Economics, Department of Measurement and Information Systems*, pages 54–57, Budapest, Hungary, 2007.

[P12] G. Takács and B. Pataki. Nearest local hyperplane rules for pattern classification. *AI*IA 2007: Artificial Intelligence and Human-Oriented Computing*, pages 302–313, Rome, Italy, 2007.

[P13] G. Takács and B. Pataki. A lépcsőzetes döntéshozás elvének műszaki alkalmazásai, (in Hungarian). *Elektronet*, 16(8): 76–78, 2007.

[P14] M. Altrichter, G. Horváth, B. Pataki, Gy. Strausz, G. Takács and J. Valyon. *Neurális hálózatok*, (in Hungarian). Panem, 2006.

[P15] G. Takács and B. Pataki. Local hyperplane classifiers. *Proc. of the 13th PhD Mini-Symposium, Budapest University of Technology and Economics, Department of Measurement and Information Systems*, pages 44–45, Budapest, Hungary, 2006.

[P16] G. Takács and B. Pataki. Fast detection of masses in mammograms with difficult case exclusion. *International Scientific Journal of Computing*, 4(3): 70–75, 2005.

[P17] G. Takács and B. Pataki. Case-level detection of mammographic masses. *Proc. of the 12th International Symposium on Interdisciplinary Electromagnetic, Mechanic and Biomedical Problems (ISEM 2005)*, pages 214–215, Bad Gastein, Austria, 2005.

[P18] G. Takács and B. Pataki. Fast detection of mammographic masses with difficult case exclusion. *Proc. of the 3rd IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS'2005)*, pages 424–428, Sofia, Bulgaria, 2005.

[P19] G. Takács and B. Pataki. Computer-aided detection of mammographic masses. *Proc. of the 12th PhD Mini-Symposium, Budapest University of Technology and Economics, Department of Measurement and Information Systems*, pages 24–25, Budapest, Hungary, 2005.

[P20] N. Tóth, G. Takács, and B. Pataki. Mass detection in mammograms combining two methods. *Proc. of the 3rd European Medical & Biological Engineering Conference (EMBEC'05)*, Prague, Czech Republic, 2005.

[P21] G. Horváth, B. Pataki, Á. Horváth, G. Takács, and G. Balogh. Detection of microcalcification clusters in screening mammography. *Proc. of the 3rd European Medical & Biological Engineering Conference (EMBEC'05)*, Prague, Czech Republic, 2005.

[P22] G. Takács. The smooth maximum classifier. Accepted at: Second Győr Symposium on Computational Intelligence, 2009.

[P23] G. Takács. Smooth maximum based algorithms for classification, regression, and collaborative filtering. Accepted at: *Acta Technica Jaurinensis, Series Intelligentia Computatorica*, 2009.

[P24] G. Takács. Efficient algorithms for determining the linear and convex separability of point sets. Accepted at: *Acta Technica Jaurinensis, Series Intelligentia Computatorica*, 2009.

[P25] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Unifying collaborative filtering approaches. Veszprém Optimization Conference: Advanced Algorithms (VOCAL 2008), Veszprém, Hungary, 2008.

[P26] R. Horváth-Bokor, Z. Horváth, and G. Takács. Kockázatelemzés logisztikus regresszióval nagy adathalmazokon, (in Hungarian). 28. Magyar Operációkutatási Konferencia, Balatonőszöd, Hungary, 2009.

References

[1] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition.* Springer, New York, 1996.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749, 2005.

[3] J. Bennett and S. Lanning. The Netflix Prize. In *Proc. of the KDD Cup and Workshop 2007*, pages 3–6, 2007. (URL: `http://www.netflixprize.com/`)

[4] O. Bousquet, S. Boucheron, and G. Lugosi. Introduction to statistical learning theory. *Lecture Notes in Artificial Intelligence*, 3176:169–207, 2004.

[5] A. Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proc. of the KDD Cup and Workshop 2007*, pages 39–42, 2007.