Disjunctive Boolean Kernel based Collaborative Filtering for top-N item recommendation

Disjunctive Boolean Kernel based Collaborative Filtering for top-N item recommendation

Mirko Polato and Fabio Aiolli

University of Padova - Department of Mathematics
Via Trieste, 63, 35121 Padova - Italy
{mpolato,aiolli}@math.unipd.it

Abstract. In many real-world recommendation tasks the available data consists only of simple interactions between users and items, such as clicks and views, called implicit feedback. In this kind of scenarios model based pairwise methods have shown of being one of the most promising approaches. In this paper, we propose a principled and efficient kernel-based collaborative filtering method for top-N item recommendation inspired by pairwise preference learning. We also propose a new boolean kernel, called Monotone Disjunctive Kernel, which is able to alleviate the sparsity issue that is one of the main problem in collaborative filtering contexts. The embedding of this kernel is composed by all the combinations of a certain degree d of the input variables, and these combined features are semantically interpreted as disjunctions of the input variables. Experiments on several CF datasets have shown the effectiveness and the efficiency of the proposed kernel-based method.

Keywords: collaborative filtering, implicit feedback, top-n recommendation, kernel methods, boolean kernels

1 Introduction

Collaborative Filtering (CF) is the *de facto* approach for making personalized recommendation. Even though, in the past, the rating prediction task has got most of the attention by the research community, recently the focus has shifted towards the implicit feedback scenario, where the task is the, so called, top-N recommendation task. This drift in favour of the implicit setting is due to the fact that in real-world recommendation scenarios (i) implicit data are much more easy to gather as they do not require any active action by the user, and (ii) they are simply more common.

In literature, many model-based methods for implicit feedback data have been proposed, such as *SLIM* [1], *WRMF* [2], *BPR*[3], and *CF-OMD* [4]. Despite the complexity of the above mentioned approaches, most of them are linear models, and in general, in CF contexts easier models tend to achieve very good results. This behaviour is typical of CF datasets because they usually own the following two characteristics: (i) they are very sparse ($\sim 99\%$ sparsity), and (ii) the distribution of the interactions over the items and/or over the users are long tailed. For these reasons, for this kind of data, it is not ideal to use more complex and, consequently, even more sparse representations.

In this paper, we present a kernel-based CF framework, based on the seminal work [4], for top-N item recommendation. This method, even though it is based on kernels, it is very efficient, highly scalable and it is very suitable for datasets with few positive and many negative/unlabeled examples (e.g., CF datasets).

We also propose a new representation for boolean valued data which is less expressive than the linear one. Specifically, we propose a (boolean) kernel, called Monotone Disjunctive Kernel (mD-Kernel), in which the feature space is composed by all the combinations of the input variables of a given degree d, and these combined features are semantically interpreted as logical disjunctions of the input variables. The underpinning idea behind the proposed mD-Kernel is to define higher-level features that are a kind of generalization of the linear ones so to obtain more general representations that hopefully can alleviate the sparsity issue. Moreover, since it is based on boolean logic we also aim to build an efficient and effective algorithm able to extract from this kernel the most relevant features, namely the most relevant boolean rules. In this way we will be able to provide explanations for the recommendations.

2 The proposed Framework

In this section we present a collaborative filtering method called CF-KOMD [5] for top-N recommendation which is inspired by preference learning [4], and designed to explicitly maximize the AUC (Area Under the ROC Curve).

In the following, we give a brief explanation of the algorithm, for further details please refer to [5]. Let $\mathbf{W} \in \mathbb{R}^{n \times k}$ be the embeddings of users ($\mathcal{U}$) in a latent factor space and $\mathbf{X} \in \mathbb{R}^{k \times m}$ be the embeddings of items ($\mathcal{I}$) in that space. Given a user u, a ranking over items can be induced by the factorization $\hat{\mathbf{R}} = \mathbf{W}\mathbf{X}$, where $\hat{r}_{ui} = \mathbf{w}_u^\top \mathbf{x}_i$ with the constraint $\|\mathbf{w}_u\| = \|\mathbf{x}_i\| = 1$.

CF-KOMD tries to learn (implicitly) the user representation $\mathbf{w}_u^*$, by solving the optimization problem

$$\boldsymbol{\alpha}_{u^+}^* = \arg\min_{\boldsymbol{\alpha}_{u^+}} \ \boldsymbol{\alpha}_{u^+}^\top \mathbf{K}_{u^+} \boldsymbol{\alpha}_{u^+} + \lambda_p \|\boldsymbol{\alpha}_{u^+}\|^2 - 2\boldsymbol{\alpha}_{u^+}^\top \mathbf{q}_u, \tag{1}$$

where $\boldsymbol{\alpha}_{u^+}$ is a probability distribution over the positive examples (i.e., rated items) for a given user u, $\mathbf{K} \in \mathbb{R}^{m \times m}$ is a kernel matrix between items induced by a given kernel function κ, and the elements of the vector $\mathbf{q}_u \in \mathbb{R}^{|\mathcal{I}_u|}$ are defined by $q_{ui} = \frac{1}{|\mathcal{I}|} \sum_{j \in \mathcal{I}} \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

The induced ranking is obtained using the scoring function $\hat{\mathbf{r}}_u = \mathbf{X}^\top \mathbf{w}_u^* = \mathbf{K}_{u^+:}^\top \boldsymbol{\alpha}_{u^+} - \mathbf{q}$ where $\mathbf{K}_{u^+:} \in \mathbb{R}^{|\mathcal{I}_u| \times |\mathcal{I}|}$ is the matrix obtained by taking the subset of rows corresponding to the positive set of items for the user u, and $\mathbf{q} \in \mathbb{R}^{|\mathcal{I}|}$ is like $\mathbf{q}_u$ but defined over the whole set of items.

2.1 Monotone Disjunctive Kernel

Since we are working inside a binary input space (i.e., binary rating matrix, $\mathbf{R} \in \{0,1\}^{n \times m}$), the kernel κ can be a boolean kernel function, i.e., $\kappa : \{0,1\}^n \times \{0,1\}^n \to \mathbb{N}$. From our previous experiments [6] and other studies, e.g., [7], it is possible to notice that linear models, in general, achieve very good results in CF contexts. For this reason, instead of using more expressive kernels, such as, the polynomial or the Gaussian, we propose a new kernel which is less expressive than the linear kernel.

The key idea of the Monotone Disjunctive Kernel (mD-kernel) is the creation of boolean disjunctions of the input variables. Specifically, the kernel creates all the combinations of variable of a certain degree d and it interprets them as disjunctions of boolean variables, e.g., $x_1x_3x_7 \equiv x_1 \vee x_3 \vee x_7$, assuming 1 as *true* and 0 as *false*. A disjunction is satisfied if and only if at least one of its literals is *true*, so in the feature space a feature is active if and only if one of its (input) variables is active. Formally, the embedding of the mD-kernel of degree d is given by $\boldsymbol{\phi}^d_\vee : \mathbf{x} \mapsto (\boldsymbol{\phi}^{(\mathbf{b})}_\vee(\mathbf{x}))_{\mathbf{b} \in \mathbb{B}_d}$, with $\boldsymbol{\phi}^{(\mathbf{b})}_\vee(\mathbf{x}) = [\![\langle \mathbf{x}, \mathbf{b} \rangle]\!] = [\![\sum_{i=1}^n x_i b_i]\!]$, where $[\![\cdot]\!]$ is the indicator function. The dimension of the mD-kernel embedding is $\binom{n}{d}$. It can be demonstrated [8] that the mD-kernel of degree d between $\mathbf{x}$ and $\mathbf{z}$, i.e., $\kappa^d_\vee(\mathbf{x}, \mathbf{z})$, can be computed by

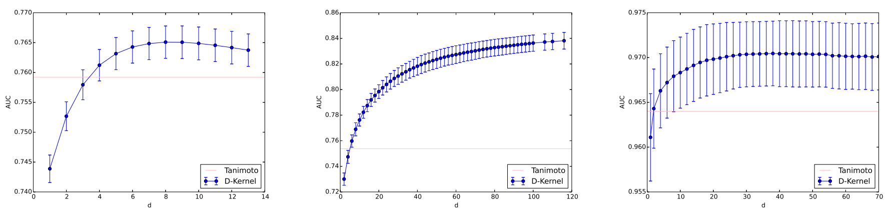
$$\kappa^d_\vee(\mathbf{x}, \mathbf{z}) = \binom{n}{d} - \binom{n - \|\mathbf{x}\|_2^2}{d} - \binom{n - \|\mathbf{z}\|_2^2}{d} + \binom{n - \|\mathbf{x}\|_2^2 - \|\mathbf{z}\|_2^2 + \langle \mathbf{x}, \mathbf{z} \rangle}{d}.$$

This kernel owns many interesting properties that are described in [8].

3 Results

In this section we report some of the empirical results obtained with our method. For more details and more extensive empirical results please refer to [5, 6, 8].

Figure 1 shows the effect of the degree of the mD-kernel on the AUC performances of *CF-KOMD* on three CF datasets.



(a) BookCrossing (b) Ciao (c) Film Trust

Fig. 1. Performance of different mD-kernel degrees.

In Table 1 there are the AUC performances achieved by our method with different kernels (only the best results are reported) against *WRMF* and *BPR*.

	MovieLens	**Netflix**	**FilmTrust**	**Ciao**
CF-KOMD	**89.6**	**94.1**	**96.4**	**73.4**
WRMF	87.0	80.4	94.7	56.5
BPR	85.4	84.3	95.4	54.9

Table 1. AUC results. Best results are reported in **bold**.

4 Conclusions and future work

We proposed a kernel-based CF method for top-N recommendation. The method has demonstrated of being efficient and very effective in many CF datasets. Moreover, we proposed a new boolean kernel, called mD-kernel, able to deal with the sparsity issue in CF contexts. We leveraged on the observation made in our previous work [6] to come up with the idea of creating a data representation less expressive than the linear one in order to mitigate the sparsity and the long tail issues. In the future we aim to build an algorithm able to efficiently extract from this kernel the most relevant features, i.e., the most relevant rules. In this way we will be able to provide explanation for the given recommendations improving the user experience.

References

1. Xia Ning and George Karypis. SLIM: sparse linear methods for top-n recommender systems. In *11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011*, pages 497–506, 2011.
2. Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *ICDM*, pages 263–272, 2008.
3. Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.
4. Fabio Aiolli. Convex AUC optimization for top-N recommendation with implicit feedback. In *ACM Recommender Systems Conference*, pages 293–296, New York, USA, 2014.
5. Mirko Polato and Fabio Aiolli. Kernel based collaborative filtering for very large scale top-n item recommendation. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 11–16, 2016.
6. Mirko Polato and Fabio Aiolli. Exploiting sparsity to build efficient kernel based collaborative filtering for top-n item recommendation. 2017 forthcoming.
7. J. Zhou and T. Luo. A novel approach to solve the sparsity problem in collaborative filtering. In *2010 International Conference on Networking, Sensing and Control (ICNSC)*, pages 165–170, 2010.
8. Mirko Polato and Fabio Aiolli. Disjunctive boolean kernels for collaborative filtering in top-n recommendation. `https://arxiv.org/abs/1612.07025`, 2016.