Understanding Collaborative Filtering with Galois Connections

Dmitry I. Ignatov[0000−0002−6584−8534] and Vladimir Kursitys

National Research University Higher School of Economics, Moscow, Russia
dignatov@hse.ru, kursitys-vladimir@rambler.ru

Abstract. In this paper, we explain how Galois connection and related operators between sets of users and items naturally arise in user-item data for forming neighbourhoods of a target user or item for Collaborative Filtering. We compare the properties of these operators and their applicability in simple collaborative user-to-user and item-to-item setting. Moreover, we propose a new neighbourhood-forming operator based on pair-wise similarity ranking of users, which takes intermediate place between the studied closure operators and its relaxations in terms of neighbourhood size and demonstrates comparatively good Precision-Recall trade-off. In addition, we compare the studied neighbourhood-forming operators in the collaborative filtering setting against simple but strong benchmark, the SlopeOne algorithm, over bimodal cross-validation on MovieLens dataset.

Keywords: Collaborative Filtering · Galois Connection · Recommender Systems · Neighbourhood-forming operators

1 Introduction

Galois connections of different types as well as closure and kernel operators play important role not only in mathematics [14] but also in analysis of relational data, for example, object-attribute tables also known as transactional databases [13,40,39], information systems [16,38], formal contexts [18,33], user-item rating matrices [9,28], etc.

Thus, it has been shown that Boolean matrix factorisation performed by means of Galois operators on user-item binary matrix (obtained from user-item rating matrix under proper scaling) is not worse than ordinary SVD to capture similarity between users and items in terms of MAE, Precision and Recall [28]. The so called concept lattice, an ordered hierarchy of maximal submatrices (users, items) generated by Galois operators, has been proposed as a global search space for nearest neighbours of users and items [9]; however, such a lattice might be huge even for sparse input rating matrices and its generating is costly in terms of time and storage memory. An interesting attempt to scale this approach and form only necessary relevant neighbourhoods of users and items via Galois operators has been done in [5].

However, a systematic study of those useful connections and operators as well as their variants suitable for Collaborative filtering has not been performed yet. In this study, we introduce and discuss Galois operators for collaborative filtering setting to form neighbourhoods of users and items (as well as their sets in group recommendation scenario) and sets of prospective relevant items to rank by means of ordinary user-based (or item-based) approaches.

The remainder of the paper consists of five sections. In Section 2, we recall several definitions of Galois or derivation operators from Formal Concept Analysis and the associated closure operators. In Section 3, we explain how the existing operators can be used to form neighbourhoods of users (items) for a target user (item) as well as sets of prospective relevant items to recommend. Section 6 presents several simple experiments with user-based and item-based approaches where the formed neighbourhoods and sets of items used as parameters. Section 7 discusses related work and Section 8 concludes the paper.

2 Galois connections and related operators

First, we give the definition of Galois connection.

Let $\varphi : P \to Q$ and $\psi : Q \to P$

be maps between two ordered sets $(P, \leq)$ and $(Q, \leq)$. The pair of such maps is called a Galois connection between the ordered sets if:

a. $p_1 \leq p_2 \Rightarrow \varphi p_1 \geq \varphi p_2$;
b. $q_1 \leq q_2 \Rightarrow \varphi q_1 \geq \varphi q_2$;
c. p $\leq \psi\varphi p$ and $q \leq \varphi\psi q$.

The operators φ and ψ are called Galois operators.

Let us define concrete version of Galois operators as it is done in Formal Concept Analysis (FCA) [18] over relational object-attribute tables but in collaborative filtering setting. Here, the role of objects is played by users and the role of attributes by items.

Let us consider a triple (U, I, R) called formal context in FCA, where U is a set of users, I is a set of items, and $R \subseteq U \times I$. A pair $(u, i) \in R$ iff user $u \in U$ rated (liked or browsed) item $i \in I$.

In this case, for a subset of users $X \subseteq U$ and a subset of items $Y \subseteq I$ Galois operators (prime or derivation operators), $(\cdot)' : 2^U \to 2^I$ and $(\cdot)' : 2^I \to 2^U$, are defined as follows:

$X' = \{i \mid uRi \text{ for all } u \in X\}$,

$Y' = \{u \mid uRi \text{ for all } i \in Y\}$.

In fact, X' is the set of items that every user from X rated and Y' are those users, who rated every item from Y.

One may check that two operators $(\cdot)'$ form a Galois connection between $(2^U, \subseteq)$ and $(2^I, \subseteq)$.

Moreover, one may prove that $(\cdot)''$ is a closure operator, i.e. for $X, Z \subseteq U$ (or for $X, Z \subseteq I$).

1. $X \subseteq Z \Rightarrow X'' \subseteq Z''$ (monotony);
2. $X \subseteq X''$ (extensity);
3. $X' = X'''$ (idempotency).

A monotone and idempotent operator $op(\cdot)$ on 2^U is called a kernel operator iff for $X \subseteq U$: $op(X) \subseteq X$ (intensity). Operators with intensity property play important role in Social Choice Theory since they help to select relevant alternatives from their input set [4]. We provide an example of kernel operator in section 3.

Let us discuss the meaning of several important properties of the introduced Galois operators in terms of Collaborative Filtering domain.

For $X, X_1, X_2 \subseteq U$ (similarly for $Y, Y_1, Y_2 \subseteq I$):

4. $X_1 \subseteq X_2 \Rightarrow X_2' \subseteq X_1'$ (antitony);
5. $X' = X'''$.

The fourth property means that the more users we add to the initial set X_1, the less is the number of their co-rated items (this property have been exploited in classic itemset mining algorithm, Apriori, in [2]). To understand the meaning of the remaining properties we need to discuss the interpretation of the result of $(\cdot)''$ to $X \subseteq U$. The first prime returns the set X' of all co-rated items for users from X, the second prime returns the set X'' of all users who rated all items from X'. In fact, this set X'' may become larger than X or remain the same (Property 2). If we have a group of users Z and its subgroup X, then after looking at the items that X and Z rated, i.e. X' and Z', we obtain by Property 4 that the set of items X' is larger or equal to Z'. By applying Property 4 one more time, we obtain that X'' is a subgroup of Z'' or equal to it. Property 2 says that by passing through items X' co-rated by X, we may obtain some more users who rated all items X' as well, i.e. our overlooked neighbours at the beginning. The third property says that it is not necessary to look at the co-rated items of the group X'' since everyone who rated all items from X' is in X''. That is X'' is a fixed point of operator $(\cdot)''$. These fixed points correspond to the called formal concepts in FCA, i.e. pairs (X'', X') for $X \subseteq U$ (for itemsets they are defined similarly).

In collaborative filtering setting, for a particular target user u from U we are mainly interested in $\{u\}'$, the items rated by u, and $\{u\}''$, all users from U, who rated all items $\{u\}'$. However, if we would require to show new items that those users also rated, applying the prime operator one more time, we would obtain $\{u\}''' = \{u\}'$, i.e. nothing to potentially recommend. One of the remedies would be to delete u from $\{u\}''$ and obtain $(\{u\}'' \setminus \{u\})' \setminus \{u\}'$, however we prefer to examine a richer set of possible alternatives and study their properties.

3 Connections for Collaborative Filtering

Let (U, I, R) be a formal context, then for a subset of users $X \subseteq U$ and a subset of items $Y \subseteq I$ then neighbourhood-forming operators, $(\cdot)^\diamond : 2^U \to 2^I$ and $(\cdot)^\diamond : 2^I \to 2^U$, are defined as follows:

$X^\diamond = \{i \mid uRi \text{ for some } u \in X\}$,
$Y^\diamond = \{u \mid uRi \text{ for some } i \in I\}$.

In fact, $X^\diamond$ can be considered as a query "show me all that have been bought by at least some user from X for $X \subseteq U$; $Y^\diamond$ is interpreted as all users that bought at least one item from Y for $Y \subseteq I$.

Property 1. $X^\diamond = \bigcup\limits_{x \in X} x'$ and $Y^\diamond = \bigcup\limits_{y \in Y} y'$.

Property 2. $X, Z \subseteq U \Rightarrow X^\diamond \subseteq Z^\diamond$ (monotony of $(\cdot)^\diamond$) (similarly for $X, Z \subseteq I$).

Now we have 2^2 combinations of operators $(\cdot)'$ and $(\cdot)^\diamond$ to form neighbours, and 2^3 operator combinations to list potentially relevant items. Let us figure it out theoretically which of the proposed combinations are relevant for collaborative filtering.

Theorem 1. *For $X, Z \subseteq U$ (similarly for $X, Z \subseteq I$) the following properties fulfil:*

- *If $X \subseteq Z$ then*
  1. *$X'^\diamond \supseteq Z'^\diamond$ (antitony);*
  2. *$X^{\diamond\prime} \supseteq Z^{\diamond\prime}$ (antitony);*
  3. *$X^{\diamond\diamond} \subseteq Z^{\diamond\diamond}$ (monotony);*
- 4. *$X \subseteq X'^\diamond$ (extensity);*
  5. *$X \supseteq X^{\diamond\prime}$ (intensity);*
  6. *$X \subseteq X^{\diamond\diamond}$ (extensity);*
- 7. *$X'^\diamond = X'^{\diamond\prime\diamond}$ (idempotency);*
  8. *$X^{\diamond\prime} = X^{\diamond\prime\diamond\prime}$ (idempotency);*
  9. *$(\cdot)^{\diamond\diamond}$ is not idempotent.*

Corollary 1. *Operator $(\cdot)^{\diamond\prime}$ is a kernel operator (antitone, extensive, and idempotent).*

In what follows, we mainly concentrate on one target user u, its neighbours founded by double combination of the derivation and neighbourhood forming operators, and items potentially relevant for that user obtained by triple combinations of those operators.

Lemma 1. *For $u \in U$: $u' = u^\diamond$ (similarly for $i \in I$).*

Theorem 2. *For $u \in U$ the following inclusions hold:*
$u'^{\diamond\prime} = u^{\diamond\diamond\prime} \subseteq u''' = u^{\diamond\prime\prime} = u' = u^{\diamond\prime\diamond} = u''^\diamond \subseteq u'^{\diamond\diamond} = u^{\diamond\diamond\diamond}$.

Thus, every triple operator on the left hand side from u' does not bring new items in comparison to those that u is rated. So, potentially we are interested in those from the right hand side, namely, $u''^\diamond$ and $u'^{\diamond\diamond}$.

Since we should eliminate the target user u from the set of his neighbours and his rated items from the set of potentially relevant items. Let us introduce two final neighbourhood forming operators:

$\mathcal{N}_{ij}(u) = u^{ij} \setminus u$ and $\mathcal{N}_{ijk}(u) = u^{ijk} \setminus u^i$, where $i, j, k \in \{\diamond, \prime\}$.

There is also an option to subtract u from its neighbourhood as soon as possible in the chain of operators applied to u. So, there is one more variant for forming potentially relevant items:

$\mathcal{N}_{ijk}^{-u}(u) = (u^{ij} \setminus u)^k \setminus u^i$, where $i, j, k \in \{\diamond, \prime\}$.

4 Similarity measure inspired by Galois operators

Given two users $u, \tilde{u} \in U$, define for them a measure of similarity as $\rho : U^2 \to \mathbb{N}$, $\rho(u, \tilde{u}) = |\{u, \tilde{u}\}'|$. For each user, we will reorder all users in ascending order: $\rho(u, u_1) \leq \rho(u, u_2) \leq \cdots \leq \rho(u, u_n)$. So, each user u generates its renumbering of the set U. Let us define the neighbourhood-forming operator $(.)^{\triangle m} : U \to 2^U$ as follows: $\{u\}^{\triangle m} = \bigcup_{i=1}^{m} \{u_i\}$.

Property 3. $\forall u \in U \ |\{u\}''| \leq m \leq |\{u\}^{\diamond\diamond}| \Leftrightarrow \{u\}'' \subseteq \{u\}^{\triangle m} \subseteq \{u\}^{\diamond\diamond}$

The neighbours of the target user will be considered as $\{u\}^{\triangle m} \setminus \{u\}$ for a given operator. This operator is useful because we exactly specify the number of neighbours, thereby solving the problem of the lack of neighbours or presence of too many of them. Then, we take the k nearest of these neighbours by the other measure (e.g. *cossim*). Since we will look for new items for the prediction in the set $(\{u\}^{\triangle m} \setminus \{u\})' \setminus \{u\}'$, we may need to solve the optimisation problem in this case for m and k choice w.r.t. MAE or Precision and Recall.

5 Algorithm

1. Find neighbours to the target user u using one of the following methods:
   - $\{u\}'' \setminus \{u\}$
   - $\{u\}'^{\diamond} \setminus \{u\}$
   - $\{u\}^{\triangle m} \setminus \{u\}$.
2. Find the set of top k nearest neighbours to the target user u among the neighbours found at the previous step using a measure of similarity *sim*. Denote this set by $\mathcal{N}_k$
3. Find new items for the user u using the method:

$$(\mathcal{N}_k \setminus \{u\})' \setminus \{u\}'$$

4. Make a prediction of the rating for each items found in the previous step. Choose top n of them, if necessary.

Since steps 1 and 3 have been considered, let us discuss steps 2 and 4.

5.1 Similarity measure

Note that this step is not necessary, but useful. The number of neighbours found at the first step can be large, which leads to fewer items for the recommendation. For our experiments, we will use the cosine measure of similarities. This measure is recognised as one of the best estimators of users' similarity [19]. Let u, $\tilde{u} \in U$, $r_{u,i}$ and $r_{\tilde{u},i}$ be the ratings of item $i \in I$ by users u and $\tilde{u}$, respectively, and the vector $\boldsymbol{r}_u = (r_{u,1}, r_{u,2}, \ldots, r_{u,n})$ be the vector of user ratings u. Then we define the cosine measure of users' similarity $cossim : U^2 \to [0, 1]$ as follows[1]:

$$cossim(u, \tilde{u}) = \frac{\boldsymbol{r}_u \cdot \boldsymbol{r}_{\tilde{u}}}{\|\boldsymbol{r}_u\| \, \|\boldsymbol{r}_{\tilde{u}}\|} = \frac{\sum_{i \in I} r_{u,i} r_{\tilde{u},i}}{\sqrt{\left(\sum_{i \in I} (r_{u,i})^2\right) \left(\sum_{i \in I} (r_{\tilde{u},i})^2\right)}}.$$

5.2 Rating prediction

The predicted rating $\hat{r}_{u,i}$ for an item $i \in I$ by a user $u \in U$ is a weighted combination of selected neighbours ratings, which is calculated as a weighted deviation from the average ratings of the neighbours. The general prediction formula is below:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{\tilde{u} \in U} (r_{\tilde{u},i} - \bar{r}_{\tilde{u}}) \, sim(u, \tilde{u})}{\sum_{\tilde{u} \in U} |\, sim(u, \tilde{u}) \,|}.$$

6 Experiments

6.1 Data

For test the model, we used data from the GroupLens[2] web site [20]. The data was collected through the MovieLens[3] recommender service during the seven-month period from September 19th, 1997 through April 22nd, 1998. This data has been cleaned up – users who had less than 20 ratings were removed from this data set.

This data set consists of:

- 100 000 ratings (1-5) from 943 users on 1682 movies.
- Each user has rated at least 20 movies.

The data represents 100 000 lines of the form:

$$|\ user\ id\ |\ item\ id\ |\ rating\ |\ timestep\ |.$$

[1] the formula should be adjusted by considering only commonly rated items in the numerator in case of missing ratings by u or $\tilde{u}$

[2] https://grouplens.org/datasets/movielens/

[3] https://movielens.umn.edu

6.2 Training/test set split procedure

We will partially imitate online testing when only a part of information on ratings for test users is known, as our operators are more focused on building recommendations than on forecasting ratings. We will follow the bimodal cross-validation procedure from [29]. To do this, we first find the sets U_{hidden} and I_{hidden}, where:

- U_{hidden} is a randomly selected 20% of all users U,
- I_{hidden} is a randomly selected 20% of all items I.

Then we hide all the information about the ratings at the intersection (U_{hidden}, I_{hidden}) as shown below and call this matrix *trainset*

$$
\begin{array}{c|cccc|ccc}
 & & & & & & I_{hidden} & \\
\hline
 & r_{1,1} & r_{1,2} & \cdots & r_{1,n} & r_{1,n+1} & \cdots & r_{1,l} \\
 & r_{2,1} & r_{2,2} & \cdots & r_{2,n} & r_{1,n+1} & \cdots & r_{2,l} \\
 & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
 & r_{m,1} & r_{m,2} & \cdots & r_{m,n} & r_{1,n+1} & \cdots & r_{m,l} \\
\hline
 & r_{m+1,1} & r_{m+1,2} & \cdots & r_{m+1,n} & * & \cdots & * \\
U_{hidden} & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
 & r_{k,1} & r_{k,2} & \cdots & r_{k,n} & * & \cdots & *
\end{array}
$$

where $r_{u,i}$ is the rating item i by users u or $*$ if this user did not rate this item yet.
Similarly, *testset* is a matrix containing all the hidden information.
Each experiment will be carried out 100 times to eliminate the dependence on random partitioning.

6.3 Adjusted Precision and Recall

We used standard measures to compare studied models: *Precision* and *Recall*. They can be defined as follows:

$$
Precision = \frac{|\{relevant\} \cap \{retrieved\} \cap I_{hidden}|}{|\{retrieved\} \cap I_{hidden}|},
$$
$$
Recall = \frac{|\{relevant\} \cap \{retrieved\} \cap I_{hidden}|}{|\{relevant\} \cap I_{hidden}|},
$$

where for user $u \in U_{hidden}$:

- $\{relevant\}$ is the set of all items that the user u rated,
- $\{retrieved\}$ is the set of all items that we recommended to the user u.

Note special cases:

- $Precision = 1$, if $\{retrieved\} \cap I_{hidden} = \varnothing$,
- $Recall = 1$, if $\{relevant\} \cap I_{hidden} = \varnothing$.

6.4 Testing models

We will test the models based on the algorithm described in Section 5. Since these algorithms different only in the initial obtaining of the neighbours of the target user, we denote them as $//$, $/\diamond$, and $\triangle m$.

Model $//$ For this experiment, after applying all the operators, we took $top_k = 5$ of the nearest neighbours by the cosine similarity measure. The result can be observed in table 1:

Table 1. Precision and Recall for the model $//$ over 100K MovieLens dataset

top n recommendation	Precision	Recall	time, sec
1	1.8%	0.01%	6.28
2	0.8%	0.06%	6.28
3	0.7%	0.08%	6.28
⋮	⋮	⋮	⋮
all	7.2%	99%	6.28

Further studies of this model were not carried out since unacceptable trade-off between such low values of Precision and Recall. The problem with this model is that we have neighbours only for $< 3\%$ users, and according to our Galois operator we have $\{\varnothing\}' = I$. Thus, we get the same predicted rating for all available movies, equals the average rating of the target user. Therefore we obtain low values of Precision and Recall.

Model $/\diamond$ For this experiment, after applying all the operators, we took $top_k = 5$ of the nearest neighbours by the cosine measure. The result can be observed in table 2:

Table 2. Precision and Recall for the model $/\diamond$ over 100K MovieLens dataset

top n recommendation	Precision	Recall	time, sec
1	97.6%	0.5%	1.51
2	97.4%	0.6%	1.51
3	97.6%	0.7%	1.51
⋮	⋮	⋮	⋮
all	97.1%	0.9%	1.51

Further studies of this model were not carried out. In this model, we have a very good Precision, but its Recall does not suit us. This is due to the fact that in this case we have too many neighbours. The average number of neighbours is more than half of the total number of users U. Therefore, the cosine measure does not correctly give us top_k nearest neighbours, e.g., 100% of similarity for the only one commonly rated item. After applying our Galois operator, we get that for $\approx 93\%$ users we have nothing to recommend.

Model $\triangle m$ For this experiment, $top_m = 50$ of the nearest neighbours by measure based on the Galois connection was taken and then $top_k = 5$ of the nearest neighbours by the cosine measure were selected.The result can be observed in table 3:

Table 3. Precision and Recall for the model $\triangle m$ over 100K MovieLens dataset

top n recommendation	Precision	Recall	time, sec
1	72.8%	4.9%	1.44
3	69.5%	10.4%	1.44
5	67.5%	13.3%	1.44
10	67.0%	16.3%	1.44
⋮	⋮	⋮	⋮
all	65.8%	17.7%	1.44

This model produced acceptable results, and thus we can work with it. We have at least one recommendation for $\approx 87.3\%$ users. It can be considered normal for our model. The maximum possible number of recommendations for the user on average is 3-4 movies. So we do not have high recall for all of recommendations. Possibly, this problem should be solved in the transition to larger data (like 10M ratings) or its portion with moderately large profiles of users. Also note that Precision is not greatly reduced by recommending a large number of movies. This is unusual for most recommendation systems. So we can immediately recommend to the user u all the movies that fall into the set $(\mathcal{N}_k \setminus \{u\})' \setminus \{u\}'$ and do not predict the ratings for these movies.

Now we need to understand whether it is necessary to solve the optimisation problem. To do this, we first look at how the values of Precision, Recall, and F_1 score change, fixing k, where F_1 score is considered according as follows: $F_1 = 2\,\frac{Precision \cdot Recall}{Precision + Recall}$.

Then fixing m and see how the Precision and Recall change in depending on k.

We can see from figures 1, 2, 3, and 4 that the Precision is directly proportional to m and k, and the Recall is inversely proportional to m and k. Therefore, it is not necessary to solve an optimisation problem for any of the parameters.

Next, we experimented with the MovieLens dataset composed by 1M ratings and look at the results. For this experiment, $top_m = 100$ of the nearest neighbours by measure based on the Galois operators and then $top_k = 10$ of the nearest neighbours by the cosine measure were taken. We have found out

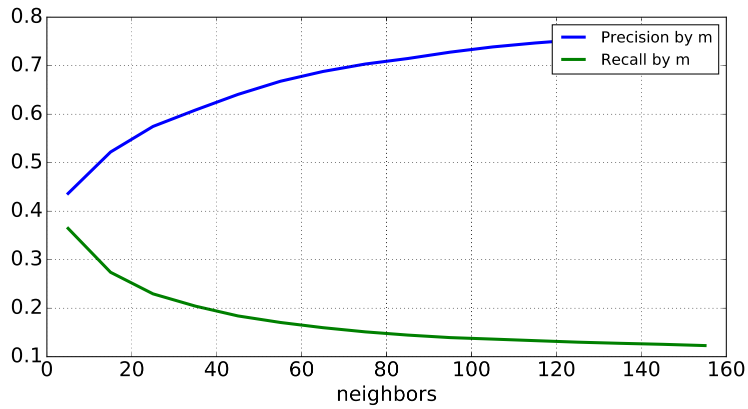

Fig. 1. Precision and Recall by m for the model $\triangle m$

that the Precision is not greatly reduced by recommending a larger number of movies. Therefore, we made an estimate of Prediction and Recall only for all possible recommendations. The result can be observed in table 4:

Table 4. Precision and Recall for the model $\triangle m$ over 1M MovieLens dataset

top n recommendation	Precision	Recall	time, sec
all	61.4%	13.1%	25.7

From this experiment we can conclude that Precision varies slightly and Recall has a small increase in the larger sample. However, we have at least one recommendation for $\approx 97.3\%$ users. That can be considered as an acceptable result since in most cases even a few recommendations is enough.

6.5 Slope One

For comparison, we took a well-known model Slope One [34]. In a reputed survey [10] it was shown that this model is one of the best for offline predicting the rating of items, while in our test we see how this model works for making recommendations.The result can be observed in table 5.

This model works worse than ours. This is due to the fact that we do not limit the set of items for prediction.

We can see in figure 5, that sorting by the top n predicted ratings does not give strong effect on the precision of the recommendation by SlopeOne.

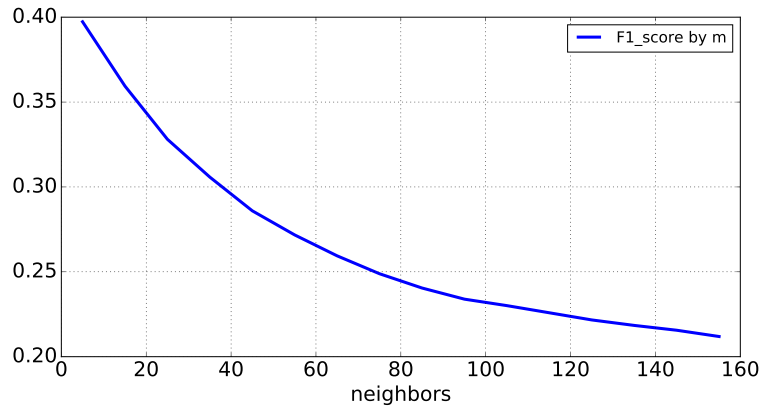

Fig. 2. F_1 score by m for the model $\triangle m$

Table 5. Precision and Recall for SlopeOne over 100K MovieLens dataset

top n recommendation	Precision	Recall	time, sec
1	3.5%	0.2%	7.66
3	8.1%	1.2%	7.66
5	10.2%	2.5%	7.66
10	13.1%	6.8%	7.66
⋮	⋮	⋮	⋮
all	7.7%	100%	7.66

7 Related Work

To the best of our knowledge, the first paper that uses Formal Concept Analysis (FCA) and Galois Connections for Collaborative Filtering was [9]. Later on, a paper on concept-based biclustering for making recommendations over firms-terms contextual advertising problem appeared [25] based on a prior study on the same datset from Yahoo! (former Overture) with spectral clustering techniques [41]; its latest version with revisited experiments and study of biclustering properties is presented in [26]. In parallel, maximal-inclusion biclusters (in fact, formal concepts) were used in similar collaborative filtering scenario [37] based on BiMax algorithm from [36]. A reincarnated study in explicit FCA-terms was done in [6] with large real commercial datasets like PayPal. FCA-based biclusters were also used [21] for recommender system to facilitate educational orientation of Russian school graduates. In [27,23], the authors used concept-based biclustering for making recommendations for crowdsourcing platform Witology to find similar user's ideas and the so-called users-antagonists for stronger team building.

As for interval-like ratings ranges, recommendations based on pattern structures [17] (an extension of FCA-approach for complex data) were firstly intro-

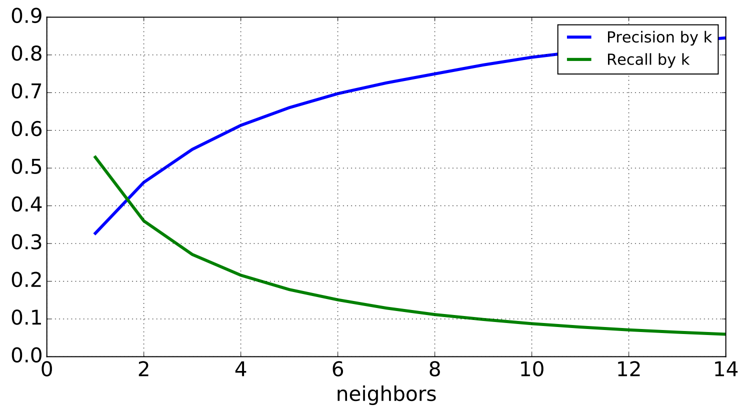

Fig. 3. Precision and Recall by k for the model $\triangle m$

duced and compared with SlopeOne approach in [24]. Another attempt to build interval-based biclusters on MovieLens data was done in [11].

Since after its success in the NetflixPrize competition, a widely accepted method for Recommender Systems is matrix factorization [32], the question on whether Boolean Matrix Factorisation (BMF) provides a competitive approach here emerged. The first answer was received in two works [35,28], where BMF-based solution was compared with Singular Value Decomposition for Collaborative Filtering in terms of MAE and demonstrated equal quality. In the subsequent paper [3], BMF was studied against SVD (compared in terms of MAE, Precision and Recall) over matrices extended by user's and item's features representing the so-called context-aware approach. The main advantages of BMF lie in its high interpretability and promising efficiency of bit-wise Boolean operations whereas its main drawback resides in higher complexity due to combinatorial nature of the optimal number of factors determination (the cover or dimension problem) [8,7].

A separate venue is recommendation for Folksonomies based on higher order extenstions of FCA; let us cite only one recent paper with a detailed introduction of the problem [30].

As an unexpected example, FCA-based collaborative filtering can be also used as an ensemble technique to suggest a proper classifier within classification framework [31].

An interested reader may also refer to a tutorial on FCA for Information Retrieval (IR) [22] and related fields with examples of FCA-based recommender systems as well as a survey on FCA for IR [12].

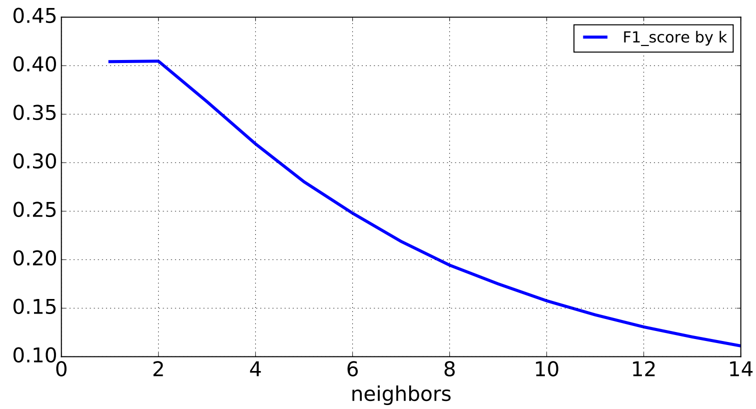

Fig. 4. F_1 score by k for the model $\triangle m$

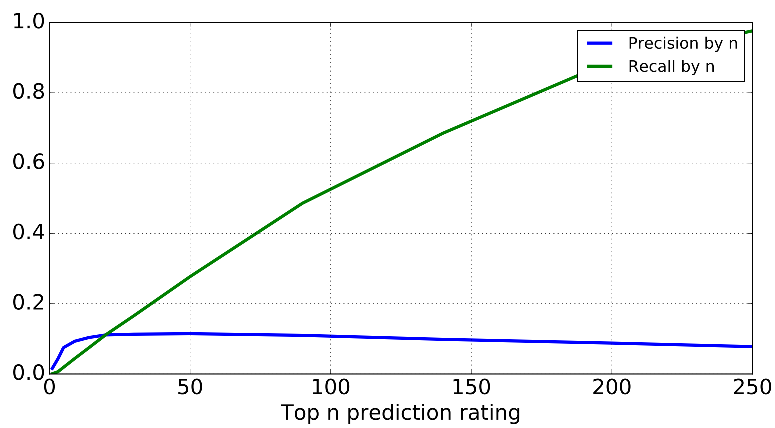

Fig. 5. Precision and Recall for SlopeOne over 100K MovieLens dataset

8 Conclusion

The obtained results seems to be a promising attempt to rethink neighbourhood-based methods in terms of Galois connections and see their theoretical comprehensiveness and limits.

The problem of finding items that the user has not yet looked at, but should see in the near future is relevant for many models. We have managed to treat this problem with the help of Galois operators. Thus this paper being not only an interesting theoretical exercise, again indirectly confirmed the hypothesis: users who rate the same items tend to rate other items similarly.

As for possible venues of the forthcoming work one may take: 1) group recommendations by means of Galois operators; 2) explicit decomposition of $\triangle m$ operator into a combination of two operators from the set of users to the set of items and vice versa; 3) extensive set of experiments with other large real datasets and more recent nearest-neighbours based techniques [1]; 4) scalability issues. A richer set of possible neighbourhood forming operators can be potentially found in [15].

Acknowledgements The first author would like to thank our colleagues either for their piece of advice or earlier contribution, Rakesh Agrawal, Marat Akhamtnurov, Shlomo Berkovskiy, Vladimir Bobrikov, Ivo Düntsch, Chandan Reddy, Mehdy Kaytoue, Sergei Kuznetsov, Denis Kornilov, Maria Mikhailova, Amedeo Napoli, Elena Nenova, Alexander Tuzhilin, and Konstantin Vorontsov.

The work was supported by the Russian Science Foundation under grant 17-11-01294 and performed at National Research University Higher School of Economics, Russia.

References

1. Adamopoulos, P., Tuzhilin, A.: On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. In: Proceedings of the 8th ACM Conference on Recommender Systems. pp. 153–160. RecSys '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2645710.2645752, http://doi.acm.org/10.1145/2645710.2645752
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile. pp. 487–499 (1994), http://www.vldb.org/conf/1994/P487.PDF
3. Akhmatnurov, M., Ignatov, D.I.: Context-aware recommender system based on boolean matrix factorisation. In: Yahia, S.B., Konecny, J. (eds.) Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, Clermont-Ferrand, France, October 13-16, 2015. CEUR Workshop Proceedings, vol. 1466, pp. 99–110. CEUR-WS.org (2015), http://ceur-ws.org/Vol-1466/paper08.pdf
4. Aleskerov, F., Bouyssou, D., Monjardet, B.: Utility Maximization, Choice and Preference. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
5. Alqadah, F., Reddy, C.K., Hu, J., Alqadah, H.F.: Biclustering neighborhood-based collaborative filtering method for top-n recommender systems. Knowl. Inf. Syst. **44**(2), 475–491 (2015). https://doi.org/10.1007/s10115-014-0771-x, http://dx.doi.org/10.1007/s10115-014-0771-x
6. Alqadah, F., Reddy, C.K., Hu, J., Alqadah, H.F.: Biclustering neighborhood-based collaborative filtering method for top-n recommender systems. Knowl. Inf. Syst. **44**(2), 475–491 (2015). https://doi.org/10.1007/s10115-014-0771-x, https://doi.org/10.1007/s10115-014-0771-x
7. Belohlávek, R., Trnecka, M.: From-below approximations in boolean matrix factorization: Geometry and new algorithm. J. Comput. Syst. Sci. **81**(8), 1678–1697

(2015). https://doi.org/10.1016/j.jcss.2015.06.002, https://doi.org/10.1016/j.jcss.2015.06.002
8. Belohlávek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. Syst. Sci. **76**(1), 3–20 (2010). https://doi.org/10.1016/j.jcss.2009.05.002, https://doi.org/10.1016/j.jcss.2009.05.002
9. du Boucher-Ryan, P., Bridge, D.G.: Collaborative recommending using formal concept analysis. Knowl.-Based Syst. **19**(5), 309–315 (2006). https://doi.org/10.1016/j.knosys.2005.11.017, http://dx.doi.org/10.1016/j.knosys.2005.11.017
10. Cacheda, F., Carneiro, V., Fernández, D., Formoso, V.: Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. ACM Trans. Web **5**(1), 2:1–2:33 (Feb 2011). https://doi.org/10.1145/1921591.1921593, http://doi.acm.org/10.1145/1921591.1921593
11. Codocedo, V., Napoli, A.: Lattice-based biclustering using partition pattern structures. In: Schaub, T., Friedrich, G., O'Sullivan, B. (eds.) ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014). Frontiers in Artificial Intelligence and Applications, vol. 263, pp. 213–218. IOS Press (2014). https://doi.org/10.3233/978-1-61499-419-0-213, https://doi.org/10.3233/978-1-61499-419-0-213
12. Codocedo, V., Napoli, A.: Formal concept analysis and information retrieval - A survey. In: Baixeries, J., Sacarea, C., Ojeda-Aciego, M. (eds.) Formal Concept Analysis - 13th International Conference, ICFCA 2015, Nerja, Spain, June 23-26, 2015, Proceedings. Lecture Notes in Computer Science, vol. 9113, pp. 61–77. Springer (2015). https://doi.org/10.1007/978-3-319-19545-2_4, https://doi.org/10.1007/978-3-319-19545-2_4
13. Cristofor, D., Cristofor, L., Simovici, D.A.: Galois connections and data mining. J. UCS **6**(1), 60–73 (2000). https://doi.org/10.3217/jucs-006-01-0060, http://dx.doi.org/10.3217/jucs-006-01-0060
14. Denecke, K., Erné, M., Wismath, S.L.: Galois Connections and Applications. Springer Netherlands (2004)
15. Dubois, D., Prade, H., Rico, A.: The cube of opposition: A structure underlying many knowledge representation formalisms. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 2933–2939 (2015), http://ijcai.org/Abstract/15/415
16. Düntsch, I., Gediga, G.: Approximation operators in qualitative data analysis. In: Theory and Applications of Relational Structures as Knowledge Instruments, COST Action 274, TARSKI, Revised Papers. pp. 214–230 (2003). https://doi.org/10.1007/978-3-540-24615-2_10, http://dx.doi.org/10.1007/978-3-540-24615-2_10
17. Ganter, B., Kuznetsov, S.: Pattern structures and their projections. In: Delugach, H., Stumme, G. (eds.) Conceptual Structures: Broadening the Base, Lecture Notes in Computer Science, vol. 2120, pp. 129–142. Springer Berlin Heidelberg (2001). https://doi.org/10.1007/3-540-44583-8_10, http://dx.doi.org/10.1007/3-540-44583-8_10
18. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edn. (1999)

19. Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. J. Mach. Learn. Res. **10**, 2935–2962 (Dec 2009), http://dl.acm.org/citation.cfm?id=1577069.1755883
20. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. TiiS **5**(4), 19:1–19:19 (2016). https://doi.org/10.1145/2827872, http://doi.acm.org/10.1145/2827872
21. Ignatov, D., Poelmans, J., Zaharchuk, V.: Recommender System Based on Algorithm of Bicluster Analysis RecBi. In: Ignatov, D., Poelmans, J., Kuznetsov, S. (eds.) CEUR Workshop proceedings Vol-757, CDUD'11 - Concept Discovery in Unstructured Data. pp. pp. 122–126 (2011)
22. Ignatov, D.I.: Introduction to formal concept analysis and its applications in information retrieval and related fields. In: Braslavski, P., Karpov, N., Worring, M., Volkovich, Y., Ignatov, D.I. (eds.) Information Retrieval - 8th Russian Summer School, RuSSIR 2014, Nizhniy, Novgorod, Russia, August 18-22, 2014, Revised Selected Papers. Communications in Computer and Information Science, vol. 505, pp. 42–141. Springer (2014). https://doi.org/10.1007/978-3-319-25485-2_3, https://doi.org/10.1007/978-3-319-25485-2_3
23. Ignatov, D.I., Kaminskaya, A.Y., Konstantinova, N., Malioukov, A., Poelmans, J.: Fca-based recommender models and data analysis for crowdsourcing platform witology. In: Graph-Based Representation and Reasoning - 21st International Conference on Conceptual Structures, ICCS 2014, Iaşi, Romania, July 27-30, 2014, Proceedings. pp. 287–292 (2014). https://doi.org/10.1007/978-3-319-08389-6_24, http://dx.doi.org/10.1007/978-3-319-08389-6_24
24. Ignatov, D.I., Kornilov, D.: RAPS: A recommender algorithm based on pattern structures. In: Proceedings of the 4th International Workshop "What can FCA do for Artificial Intelligence?", FCA4AI 2015, co-located with the International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina, July 25, 2015. pp. 87–98 (2015), http://ceur-ws.org/Vol-1430/paper9.pdf
25. Ignatov, D.I., Kuznetsov, S.O.: Concept-based Recommendations for Internet Advertisement. In: Belohlavek, R., Kuznetsov, S.O. (eds.) Proc. CLA 2008. CEUR WS, vol. Vol. 433, pp. 157–166. Palack University, Olomouc, 2008 (2008)
26. Ignatov, D.I., Kuznetsov, S.O., Poelmans, J.: Concept-based biclustering for internet advertisement. In: ICDM Workshops. pp. 123–130 (2012)
27. Ignatov, D.I., Mikhailova, M., Zakirova, A.Y., Malioukov, A.: Recommendation of ideas and antagonists for crowdsourcing platform witology. In: Information Retrieval - 8th Russian Summer School, RuSSIR 2014, Nizhniy, Novgorod, Russia, August 18-22, 2014, Revised Selected Papers. pp. 276–296 (2014). https://doi.org/10.1007/978-3-319-25485-2_9, http://dx.doi.org/10.1007/978-3-319-25485-2_9
28. Ignatov, D.I., Nenova, E., Konstantinova, N., Konstantinov, A.V.: Boolean matrix factorisation for collaborative filtering: An fca-based approach. In: Artificial Intelligence: Methodology, Systems, and Applications - 16th International Conference, AIMSA 2014, Varna, Bulgaria, September 11-13, 2014. Proceedings. pp. 47–58 (2014). https://doi.org/10.1007/978-3-319-10554-3_5, http://dx.doi.org/10.1007/978-3-319-10554-3_5
29. Ignatov, D.I., Poelmans, J., Dedene, G., Viaene, S.: A new cross-validation technique to evaluate quality of recommender systems. In: Kundu, M., Mitra, S., Mazumdar, D., Pal, S. (eds.) Perception and Machine Intelligence, Lecture Notes in Computer Science, vol. 7143, pp. 195–202. Springer Berlin Heidel-

berg (2012). https://doi.org/10.1007/978-3-642-27387-2_25, http://dx.doi.org/10.1007/978-3-642-27387-2_25
30. Jelassi, M.N., Yahia, S.B., Nguifo, E.M.: Towards more targeted recommendations in folksonomies. Social Netw. Analys. Mining **5**(1), 68:1–68:18 (2015). https://doi.org/10.1007/s13278-015-0307-8, https://doi.org/10.1007/s13278-015-0307-8
31. Kashnitsky, Y., Ignatov, D.I.: Can fca-based recommender system suggest a proper classifier? In: Proceedings of the 4th International Workshop "What can FCA do for Artificial Intelligence"? co-located with the European Conference on Artificial Intelligence, FCA4AI@ECAI 2014. pp. pp. 17–26 (2014), http://arxiv.org/abs/1504.05473
32. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009). https://doi.org/http://doi.ieeecomputersociety.org/10.1109/MC.2009.263
33. Kuznetsov, S.O.: Galois connections in data analysis: Contributions from the soviet era and modern russian research. In: Formal Concept Analysis, Foundations and Applications. pp. 196–225 (2005). https://doi.org/10.1007/11528784_11, http://dx.doi.org/10.1007/11528784_11
34. Lemire, D., Maclachlan, A.: Slope one predictors for online rating-based collaborative filtering. In: Proceedings of the 2005 SIAM International Conference on Data Mining. pp. 471–475 (2005). https://doi.org/10.1137/1.9781611972757.43, http://epubs.siam.org/doi/abs/10.1137/1.9781611972757.43
35. Nenova, E., Ignatov, D.I., Konstantinov, A.V.: An fca-based boolean matrix factorisation for collaborative filtering. In: International Workshop FCA meets IR at ECIR 2013. vol. Vol. 977, CEUR Workshop Proceeding, pp. 57–73 (2013)
36. Prelic, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Hennig, L., Thiele, L., Zitzler, E.: A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics **22**(9), 1122–1129 (2006). https://doi.org/10.1093/bioinformatics/btl060, https://doi.org/10.1093/bioinformatics/btl060
37. Symeonidis, P., Nanopoulos, A., Papadopoulos, A.N., Manolopoulos, Y.: Nearest-biclusters collaborative filtering based on constant and coherent values. Inf. Retr. **11**(1), 51–75 (2008)
38. Wolski, M.: Galois connections and data analysis. Fundam. Inform. **60**(1-4), 401–415 (2004), http://content.iospress.com/articles/fundamenta-informaticae/fi60-1-4-26
39. Zaki, M.J., Meira Jr., W.: Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press (2014), http://www.dataminingbook.info/
40. Zaki, M.J., Hsiao, C.: Efficient algorithms for mining closed itemsets and their lattice structure. IEEE Trans. Knowl. Data Eng. **17**(4), 462–478 (2005). https://doi.org/10.1109/TKDE.2005.60
41. Zhukov, L.E.: Spectral clustering of large advertiser datasets. Technical report, Overture R&D (April 2004)