



HAL
open science

Interactive Mining of User-Preferred Co-location Patterns Based on SVM

Yuxiang Zhang, Xuguang Bao, Liang Chang, Tianlong Gu

► **To cite this version:**

Yuxiang Zhang, Xuguang Bao, Liang Chang, Tianlong Gu. Interactive Mining of User-Preferred Co-location Patterns Based on SVM. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.89-100, 10.1007/978-3-031-03948-5_8 . hal-04178709

HAL Id: hal-04178709

<https://inria.hal.science/hal-04178709v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Interactive Mining of User-Preferred Co-Location Patterns Based on SVM

Yuxiang Zhang², Xuguang Bao^{1,2}, Liang Chang^{1,2(✉)}, and Tianlong Gu^{1,3}

¹ Guangxi Key Laboratory of Trusted Software, Guilin 541004, China

² Guilin University of Electronic Technology, Guilin 541004, China

³ Jinan University, Guangzhou 510632, China

changl@guet.edu.cn

Abstract. Co-location pattern mining plays an important role in spatial data mining. With the rapid growth of spatial datasets, the usefulness of co-location patterns is strongly limited by the huge amount of the discovered patterns. To overcome this drawback, several statistics-based methods have been proposed to reduce the number of discovered co-location patterns. However, these methods cannot guarantee that their mined patterns are really user-preferred. Therefore, it is crucial to help users discover the co-location patterns they preferred through effective interactions. This paper proposes a newly interactive approach based on SVM, in order to discover user-preferred co-location patterns. First, we presented an originally interactive framework to help the user discover his/her preferred co-location patterns. Then, we designed a filtering algorithm with a small part of patterns as the training set of the SVM model for the provision of each interactive process, by which a high efficiency could be achieved by the optimization of the SVM model. Finally, the system was verified on both the real data sets and the synthetic data sets, accompanying with the 80% prediction accuracy.

Topics: Machine Learning, Knowledge-based Systems, Knowledge engineering

Contact author: Yuxiang Zhang, E-mail: zyxcs1206@163.com.

Interactive Mining of User-Preferred Co-Location Patterns Based on SVM

Yuxiang Zhang², Xuguang Bao^{1,2}, Liang Chang^{1,2(✉)}, and Tianlong Gu^{1,3}

¹ Guangxi Key Laboratory of Trusted Software, Guilin 541004, China

² Guilin University of Electronic Technology, Guilin 541004, China

³ Jinan University, Guangzhou 510632, China
changl@guet.edu.cn

Abstract. Co-location pattern mining plays an important role in spatial data mining. With the rapid growth of spatial datasets, the usefulness of co-location patterns is strongly limited by the huge amount of the discovered patterns. To overcome this drawback, several statistics-based methods have been proposed to reduce the number of discovered co-location patterns. However, these methods cannot guarantee that their mined patterns are really user-preferred. Therefore, it is crucial to help users discover the co-location patterns they preferred through effective interactions. This paper proposes a newly interactive approach based on SVM, in order to discover user-preferred co-location patterns. First, we presented an originally interactive framework to help the user discover his/her preferred co-location patterns. Then, we designed a filtering algorithm with a small part of patterns as the training set of the SVM model for the provision of each interactive process, by which a high efficiency could be achieved by the optimization of the SVM model. Finally, the system was verified on both the real data sets and the synthetic data sets, accompanying with the 80% prediction accuracy.

Keywords: Spatial data mining, Interesting co-location patterns, Machine learning, Support Vector Machines, Interactive system.

1 Introduction

Spatial co-location pattern mining is an important task in spatial data mining. A spatial co-location pattern represents a subset of spatial features whose instances are frequently located in spatial neighborhoods [1]. Co-location patterns can support decision making in various domains. For example, government agencies collect co-location patterns that residents are interested in to plan reasonable bus routes. Other application domains include biology, air pollution and earth science [2].

There exists a common problem in spatial co-location patterns mining, that is, only a small part of user-preferred patterns can be found by traditional mining methods. In general, the participation index (PI) value proposed by Shekhar and Huang [3] is used to measure the interest level of a co-location pattern [4], i.e., Given a user-specified prevalence threshold min_prev , for a co-location pattern c , if $\text{PI}(c) \geq \text{min_prev}$ holds, c is considered prevalent (interesting). However, the PI value is an objective measurement and satisfies the anti-monotonicity (i.e., $\text{PI}(c) \geq \text{PI}(c')$ if $c \subseteq c'$), if a co-location is

prevalent, all its subsets are also prevalent. Thus, growing the size of spatial datasets, common frameworks generate numerous prevalent co-location patterns, amongst them only a small fraction might be user-preferred. Thus, several condensed representations of prevalent co-locations are proposed in the related literature including maximal co-location [5], closed co-location patterns [6], and super participation index closed co-location patterns [7].

However, it is found that the co-location patterns with a high PI value are often common senses, e.g. {hospital, pharmacy}. Hence, they may be redundant to one user yet the above-mentioned condensed representations are unable to filter them out. Users assess whether they prefer a co-location pattern according to their subjective judgments. Therefore, the mining methods should be based on solid interactivity with the user to discover user-preferred co-location patterns. Motivated by the interactive mining methods in transactional data mining, several methods of user-preferred co-location patterns mining based on probability [8] and ontologies [9] are proposed. Although these interactive methods can discover co-location patterns that users are really interested in, they still have some shortcomings. The probabilistic method only concerns with the composition of co-location patterns, similar combinations of co-location patterns will be judged as user-preferred patterns. The ontology-based method only concerns the semantics of co-location patterns. It is assumed that if a co-location pattern is user-preferred, other co-location patterns with similar semantics are also user-preferred.

In order to overcome the above shortcomings and improve the accuracy of co-location patterns mining. In this paper, an interactive mining method based on support vector machine (IMMBS) is proposed to discover user-preferred co-location patterns. Our goal is to help a particular user interactively discover interesting co-location patterns according to his/her real interest. Instead of requiring the user to explicitly express his/her real interesting co-location patterns, we alleviate the user's burden by only asking him/her to choose a small set of sample co-location patterns according to his/her interest for several rounds.

SVM is a supervised machine learning algorithm in the field of machine learning, it is mainly used for classification problems and has been successfully applied in text classification [10], biological information [11], image recognition [12], etc. As a standard classification tool, SVM does not need to rely on all the data, that is, only a part of support vector is used to make hyperplane decisions. In IMMBS, we train the SVM model with a small subject of co-location patterns and the SVM model learns combinations of different features. Therefore, the shortcomings of the composition or semantics of the co-location patterns are solved.

Fig. 1 shows the description of IMMBS. It is similar to the process of traditional interactive co-location patterns mining, which requires a set of prevalent co-locations as the input. In each interaction, the filtering algorithm selects several (e.g., 10-30) co-location patterns from candidate set to present to the user. The user then indicates his/her preference on each provided co-location pattern, and these labelled co-location patterns are inputted into the SVM model as the training set. Next, the SVM model learns from the training set and predicts the candidate set. IMMBS outputs the prediction results of each co-location pattern in the candidate set and the distance from each pattern to the hyperplane, and updates the training set. The interaction process continues

for several rounds until no further changes are made in forecast results, or the user stops the process.

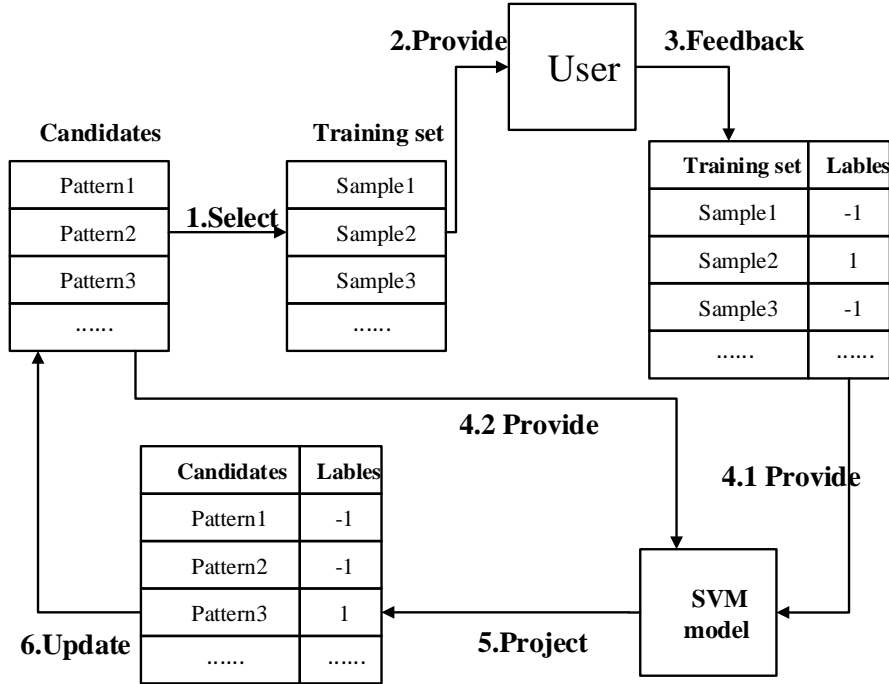


Fig. 1. Framework description.

In summary, this paper makes the following three contributions for interactive co-location pattern mining.

- 1) A novel interactive mining framework is designed to enable users to discover their preferred co-location patterns. This framework avoids using the objective PI value as a threshold, and assists different users in effectively discovering the patterns they are really interested in.
- 2) We propose a method to calculate the similarity between different co-location patterns based on the features of the pattern, and design a filtering algorithm based on this similarity and the support vectors of different co-location patterns.
- 3) The SVM model is optimized to assist IMMBS to achieve higher accuracy through fewer interactions, thus reducing the pressure of user selection.

2 Related Work and Problem Statement

2.1 Prevalent Co-location Patterns

In a spatial database, different kinds of things in space are represented by spatial features, and the occurrence of spatial features once in a spatial position is called an

instance of this spatial feature. Let F be a set of n features $F = \{f_1, f_2, \dots, f_n\}$, S be a set of instances of F , where each instance is a tuple $\langle \text{feature type, instance ID, location} \rangle$, and R be a neighbor relationship over locations of instances, where R is symmetric and reflexive. If the Euclidean distance between two different instances is not greater than the distance threshold d , the two spatial instances will satisfy the R .

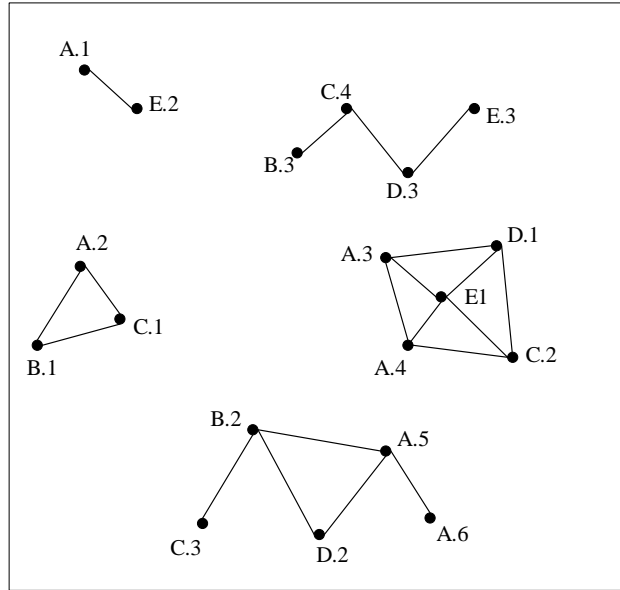


Fig. 2. An example neighbor graph.

A co-location pattern c is composed of a set of spatial features (i.e., c is a subset of F). The number of features in c is called the size of c . If any instance set satisfies the relationship R with other instances in the instance set, then a clique is formed by the instances in the instance set. A row instance I of a co-location c is a set of instances in S . The set of all row instances of c is called table instance of c . If the Participation index (PI) of a co-location pattern c is no less than a given prevalence threshold min_prev (i.e. $PI(c) \geq min_prev$), then c is defined as a prevalent co-location pattern.

Fig. 2 shows an example of spatial instance distribution. It contains 5 features (A, B, C, D, E), where $A.1$ is represented as the first instance of A . This space contains 5 instances of A , 3 instances of B , 4 instances of C , 3 instances of D , and 3 instances of E . The connection lines indicate that two spatial instances meet the neighbor relationship. For example, $A.1$ and $E.2$ are mutual neighbors. $\{A.2, B.1, C.1\}$ is formed into a clique, and it is a row instance of a 3-size co-location pattern $\{A, B, C\}$. Since, the table instance of the co-location pattern $\{A, B, C\}$ is namely $\{\{A.2, B.1, C.1\}\}$ for the reason that there are no other instances that contain the features A, B , and C . If $PI(\{A, B, C\}) \geq min_prev$, $\{A, B, C\}$ will be defined as a prevalent co-location pattern.

Huang et al. previously reported that a join-based algorithm discovery prevalent co-location patterns [4]. However, this algorithm is inefficient when processing dense

spatial databases. Yoo and Shekhar proposed the partial-join approach [13] and join-less approach [14], which solved the efficiency disadvantage of the join-based approach by constructing a neighboring list through the materialization of the whole spatial data.

2.2 Interaction with Users in Pattern Mining

In transactional data mining, many interactive systems have been proposed. The common ideas are as follows: given a set of candidate patterns, the system selects several sample patterns for the user, then requires the user to express his/her preferences on the sample patterns. Finally, the personalization model is utilized to estimate the user's prior knowledge. This system outputs all user's preferred patterns through multiple interactions. However, spatial data is much more complex than transactional data, which leads to the fact that interactive mining methods in transactional data cannot be applied to spatial data mining.

In co-location patterns mining, several interactive systems have been proposed to discover user-preferred co-location patterns. Bao et al. [15] proposed an interactive ontology-based method, where ontologies were used to select the sample patterns and update the candidates. Wang et al [8] presented an interactive probabilistic postmining method to discover user-preferred co-location patterns by iteratively obtaining user feedback and probabilistically refining preferred patterns. However, two similar co-location patterns are outputting the same result using these two methods. The user's preferences are subjective, so that small differences make the user give different choices in practical cases. Therefore, in this paper, we present a novel filtering algorithm and an interactive system, which are more adaptable than the existing methods, to assist the user in discovering his/her preferred co-location patterns through an interactive process.

2.3 Problem Statement

The problem of post-mining preferred spatial co-location patterns through interactive feedback can be stated as follows. Given a set of prevalent co-location patterns, the system outputs the ideal user-preferred co-location patterns according to the user's feedback, and at the same time minimize the user's efforts in providing feedback.

Considering the uncertainties of the user's ideal preference, in this paper we design a novel filtering to learn the prior knowledge of the user and use the SVM model to predict the user-preferred co-location patterns. The basic idea of this method is: given a set F of prevalent co-location patterns, there exists a set F_l of ideal preference co-location patterns in F for a user. Actually, the system does not know which pattern belong to F_l in the interactive process. The pattern in F_l is selected by the filtering algorithm as much as possible. Then the user labels the selected patterns and put them into the SVM model for training. According to the output data of the SVM model, the patterns in F is selected again by the system. The resultant output set would finally be close to the user's ideal preference result set after repeated interactions.

3 Algorithm

3.1 Similarity Calculation of Co-location Patterns

In traditional transactional data mining, Jaccard index is widely adopted to measure the distance between two item-sets. However, this index cannot be used to calculate the similarity between two co-location patterns in co-location data mining because there exist no conceptions of transactions in spatial data. Thus, we calculate the similarity between co-location patterns based on the characteristics of different co-location patterns, and the similarity measure is defined as follows:

$$similarity(c_1, c_2) = \frac{|T(c_1) \cap T(c_2)|}{|T(c_1) \cup T(c_2)|} \quad (1)$$

Where c_1 and c_2 are two different co-location patterns, $T(c_1)$ is the features contained in c_1 . A larger value of $similarity(c_1, c_2)$ demonstrates the similarity of the two co-locations, i.e., c_1 and c_2

For example, let $c_1 = \{A, B, C, D, E\}$, $c_2 = \{B, D, F, G\}$, $T(c_1) \cap T(c_2) = \{B, D\}$, $T(c_1) \cup T(c_2) = \{A, B, C, D, E, F, G\}$, so $similarity(c_1, c_2) = 2/7$.

3.2 SVM Model

The co-location patterns marked by the user are used as the training set of SVM model. Assuming a training set $S = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$, where $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ represents a vector of input; y_i denotes the corresponding output vector; n means the sample number; m denotes the number of input vector dimension.

The distance from each pattern to the hyperplane is an important measure in the filtering algorithm. This distance is calculated by the SVM model in each prediction process. The equation dividing the hyperplane is expressed as $\omega^T x + b = 0$, where the direction of the hyperplane is determined by the normal vector ω and b is the displacement term which determines the distance between the hyperplane and the origin. The distance from each pattern to the hyperplane is denoted as d .

$$d = \frac{|\omega \cdot x + b|}{\|\omega\|} \quad (2)$$

The SVM performance and overfitting phenomena are affected by kernel function selection and hyperparameter optimization [16]. Thus, accuracy is evaluated in each training step, so if all parameter combinations have been examined, then the best kernel parameter will be used to build the SVM. Finally, the classification accuracy of the obtained SVM will be evaluated.

The specific process of the SVM model is shown in Algorithm 1, where *kernels* contain kernel functions such as *liner*, *rbf*. The prediction score of each kernel function is calculated, and the maximal one is selected as the kernel function of the training

model. Then the candidate set is predicted. Finally, the distance from each pattern to the hyperplane and the prediction result is the output.

Algorithm 1 SVM model

Input:

CS: A set of candidates.

TS: A set of training samples.

Output:

Result: The label of each pattern in the candidate set, *distance_set*.

Variables:

kernels: the set of kernel functions.

Method:

- 1) **For Each** *kernel* in *kernels* **Do**
 - 2) *score_list.add(svc(kernel,TS));*
 - 3) *train_model(max_sore_list,TS);*
 - 4) *predict_model(CS);*
 - 5) *cal_distance();*
 - 6) **Return** *Result*;
-

Assume that there are n labeled patterns in the training set, SVM model training typically needs $O(n^2)$ calculation.

3.3 Filtering Algorithm

A good selecting method contributes to the improvement of interaction efficiency. To enable users to obtain co-location patterns with different features more efficiently, the patterns outputted by the SVM model near the hyperplane (i.e., the predicted ambiguous patterns) are also given to users for re-label. Therefore, we design a new filtering algorithm based on the similarity calculation method of transactional data and the distance from each co-location pattern output from the SVM model to the hyperplane. The idea is as follows:

- 1) The proportion of different quantitative feature patterns in the candidate sets is calculated and divided into several small candidate sets according to the number of features.
- 2) The filtering algorithm selects a pattern with the maximum PI value into the training sets.
- 3) The union operation is performed for each feature of the training set.
- 4) Calculate the similarity between the candidate set and the union generated in the previous step.
- 5) The similarity of each pattern in the candidate set is added to the distance from each pattern to the hyperplane output by the SVM model, and the value is denoted as *sum*

In each small candidate set generated in step 1, the algorithm could choose the pattern corresponding to the minimum sum value and put it into the training set. Algorithm

2 shows the pseudocode of the filtering algorithm. It contains three parts: 1) calculating the proportions of the patterns of different sizes (lines 1 and 2); 2) adding the filter parameters (similarity and distance) for each pattern (lines 5-7); 3) filtering by proportion (lines 8 and 9). The computational complexity of the three parts is $O(|CS|)$, $O(|CS|)$, and $O(|sizes|)$. Note that TS is empty in the initial and the pattern with the maximum PI is put into TS . DIS is the distance output by the SVM model. If the system interacts with the user for the first time, the DIS will be empty and only the similarity will be calculated.

Algorithm 2 Filtering Algorithm

Input:

CS : A set of candidates.

TS : A set of training samples.

DIS : A set of each pattern to hyperplane.

Output:

TS

Variables:

$pattern$: The co-location pattern in the candidate set.

$sizes$: The proportion of different quantitative feature patterns.

u : The union of all features in the training set.

sum : The sum of similarity and distance of each pattern.

Method:

- 1) **For Each** $pattern$ in CS **Do**
 - 2) $sizes = cal_size()$;
 - 3) **If** $TS == NULL$ **Then**
 - 4) $TS.Add(max_PI(CS))$;
 - 5) $u = union(TS)$;
 - 6) **For Each** $pattern$ in CS **Do**
 - 7) $sim_list.Add(cal_similarity(pattern,u))$;
 - 8) $sum = sim_list + DIS$;
 - 9) **For Each** $size$ in $sizes$ **Do**
 - 10) $TS.Add(min_sim_list(size))$;
 - 11) **return** TS ;
-

4 Evaluation of the experiment

In this chapter, we used both real data sets and synthetic data sets to verify the accuracy of the IMMBS and the effectiveness of the algorithm.

4.1 Experimental analysis of real data

Experimental analysis of the real dataset was mainly used to verify the accuracy of IMMBS. The real dataset was a spatial POI dataset of urban elements in Beijing, which contained 16 spatial features. The number of spatial instances was 90 000 and the spatial

range was $18\text{ km} \times 18\text{ km}$. We mined 960 prevalent co-location patterns by setting the minimum prevalence threshold as 0.4 and setting the distance threshold as 80 m.

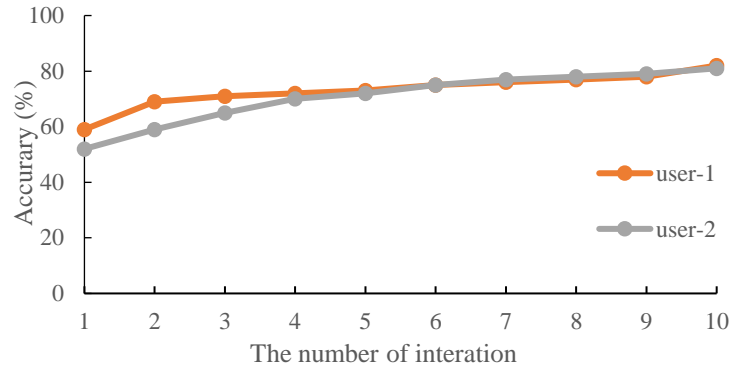
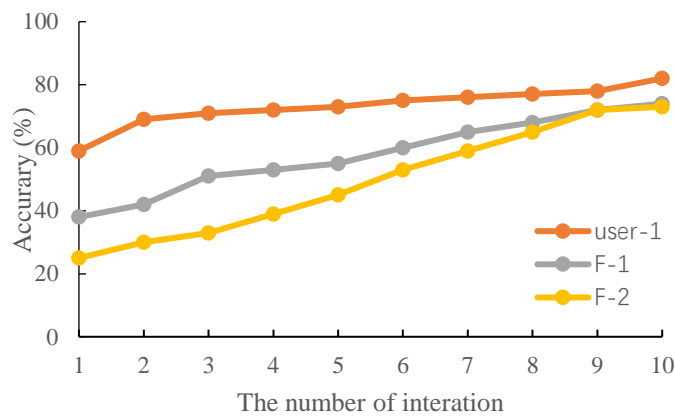


Fig. 3. Forecast results of different users.

The effectiveness of IMMBS for users with different prior knowledge was proved more intuitively. We simulated two different user preferences as the standard set user-1 and user-2 to verify the accuracy of IMMBS. The measurement formula of accuracy in the experiment is $\text{Accuracy} = (M \cap N) / (M \cup N)$, M is expressed as a standard set (user-1 or user-2), N is expressed as the result set output by IMMBS.

Fig. 3 shows the accuracy of IMMBS under different users. In the previous few interactions, the prediction accuracy of the two users was different. This was because the screened co-location patterns in the previous few times were rarely interesting to user-2. IMMBS's prediction accuracy for the two users tended to be the same after seven interactions. Thus, the prediction result showed that IMMBS was effective in predicting according to the preferences of different users.



(a) Comparison result of user-1.

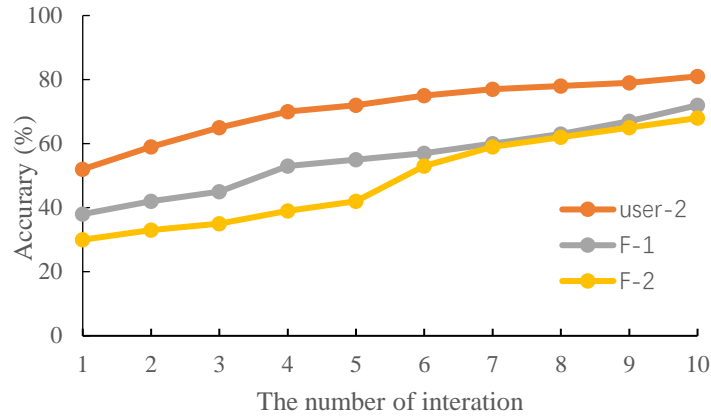


Fig. 4. (b)Comparison result of user-2.

The accuracy comparison between IMMBS and other interactive methods was exhibited in Fig. 4. F-1 was the accuracy rate of the interactive probabilistic post-mining method [8] under the same conditions and F-2 was the accuracy rate of the interactive mining method based on ontology [17] under the same conditions. The accuracy rate of IMMBS was higher than that of F-1 and F-2 under the same number of interactions. Thus, the prediction result showed that IMMBS was effective in predicting according to the preferences of different users.

4.2 Experimental analysis of synthetic data

Evaluations based on the synthetic datasets illustrated the effectiveness of IMMBS on the dataset of different sizes and the effectiveness of the filtering algorithm.

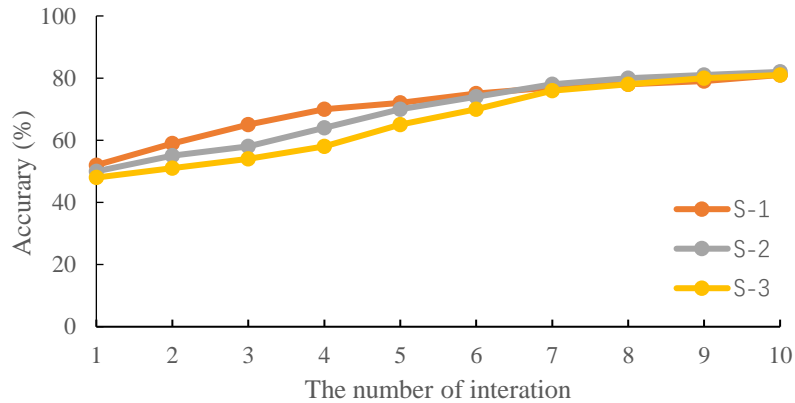


Fig. 5. Forecast results for different data set sizes.

Fig. 5 showed the accuracy of 10 interactions using IMMBS in datasets of different sizes. S-1, S-2, S-3 was the accuracy rate with different capacity of input dataset, corresponding to 15000, 25000 and 35000 respectively. The results verify that IMMBS is effective on the dataset of different sizes.

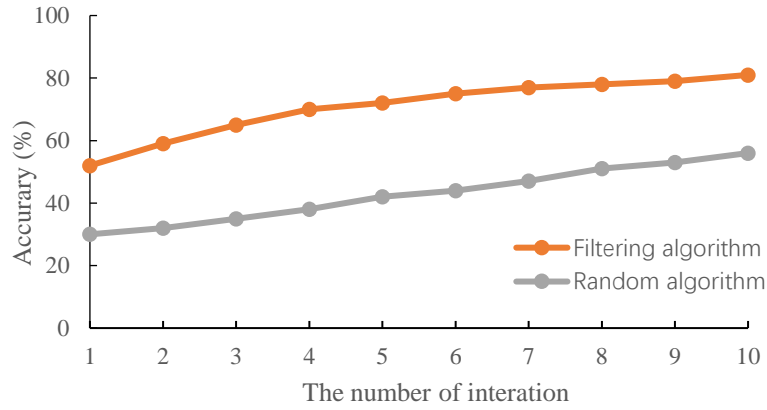


Fig. 6. Forecast results of different filtering algorithms.

The accuracy of IMMBS interacting was shown with the same user under the two algorithms (Fig. 6). The novel filtering algorithm was confirmed to be effective for IMMBS, by comparing a random algorithm. Obviously, the novel filtering algorithm was superior in the improvement of interaction efficiency and prediction accuracy of IMMBS.

5 Conclusions

In traditional co-location patterns mining approaches, the measurement was objective and users passively accept the results. In this paper, we designed a system IMMBS to find interesting patterns by interacting with users. Different from traditional mining methods, IMMBS with the SVM model was applied to predict the candidate sets by acquiring the user's prior knowledge. A filtering algorithm was designed to select the training sets for reducing the number of interactions and improving the efficiency of IMMBS.

We will apply new machine learning methods to interactive spatial co-location patterns mining in the following study. A new filter algorithm not only assists users to filter most of the co-location patterns, but also keeps the representative co-location patterns.

Acknowledgements. This work was supported in part by grants (No. U1811264, No. U1711263, No. 61966009, No. 62006057, 61762027) from the National Natural Science Foundation of China, in part by grants (No. 2018GXNSFDA281045, No. 2019GXNSFBA245059) from the Natural Science Foundation of Guangxi Province,

and in parts by grants (No.AD19245011) from the Key Research and Development Program of Guangxi Province.

References

1. Verhein, Florian, and Ghazi Al-Naymat. "Fast mining of complex spatial co-location patterns using GLIMIT." Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007). IEEE, pp. 679-684 (2007).
2. Anderson, James M., et al. "Simulations of VLBI observations of a geodetic satellite providing co-location in space." *Journal of Geodesy* 92.9 pp: 1023-1046 (2018).
3. Shekhar, Shashi, and Yan Huang. "Discovering spatial co-location patterns: A summary of results." *International symposium on spatial and temporal databases*. Springer, Berlin, Heidelberg, pp. 1472-1485 (2001).
4. Huang, Yan, Shashi Shekhar, and Hui Xiong. "Discovering colocation patterns from spatial data sets: a general approach." *IEEE Transactions on Knowledge and data engineering* 16.12 pp: 1472-1485 (2004).
5. Wang, Lizhen, et al. "An order-clique-based approach for mining maximal co-locations." *Information Sciences* 179.19 pp: 3370-3382 (2009).
6. J. S. Yoo and M. Bow, "Mining top-k closed co-location patterns," in Proc. IEEE Int. Conf. Spatial Data Min. Geogr. Knowl. Services, pp. 100–105 (2011).
7. Wang, Lizhen, et al. "Effective lossless condensed representation and discovery of spatial co-location patterns." *Information Sciences* 436 pp: 197-213 (2018).
8. Wang, L., Bao, X., & Cao, L. Interactive probabilistic post-mining of user-preferred spatial co-location patterns. In 2018 IEEE 34th International Conference on Data Engineering (ICDE) (pp. 1256-1259). IEEE (2018).
9. Bao, Xuguang, et al. "Knowledge-Based Interactive Postmining of User-Preferred Co-Location Patterns Using Ontologies." *IEEE Transactions on Cybernetics*, pp.1-14 (2021).
10. Haddoud, Mounia, et al. "Combining supervised term-weighting metrics for SVM text classification with extended term representation." *Knowledge and Information Systems* 49.3 pp: 909-931 (2016).
11. Bertoni, Alberto, Raffaella Folgieri, and Giorgio Valentini. "Bio-molecular cancer prediction with random subspace ensembles of support vector machines." *Neurocomputing* 63 pp: 535-539 (2005).
12. Inglada, Jordi. "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features." *ISPRS journal of photogrammetry and remote sensing* pp: 236-248 (2007).
13. J. S. Yoo, S. Shekhar, J. Smith, and J. P. Kumquat, "A partial join approach for mining co-location patterns," in Proc. ACM IWGIS, pp. 241–249 (2004).
14. J. S. Yoo and S. Shekhar, "A joinless approach for mining spatial colocation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1323–1337 (2006).
15. X. Bao and L. Wang, "Discovering interesting co-location patterns interactively using ontologies," in Proc. DASFAA Workshops, pp. 75–89 (2017).
16. Yang, Keyu, et al. "Towards factorized svm with gaussian kernels over normalized data." 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE (2020).
17. Bao, X., Wang, L., & Xiao, Q. OICPM: An Interactive System to Find Interesting Co-location Patterns Using Ontologies. In Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data (pp. 329-332). Springer, Cham (2017).