Mining Spatial Co-location Patterns from Vague Set

Gao Shijian¹,a, Yu Jiankun²,b

¹Kunming University of Science and Technology Oxbridge College, 1369 Yunqiao Street, Guandu District, Kunming City, Yunnan Province
²Yunnan University of Finance and Economics, 237 Longquan Road, Kunming City, Yunnan Province

a2138gao@163.com, bjk1102@163.com

ABSTRACT. The target of mining spatial co-location patterns is discovering spatial association rule from spatial database, which plays an important role in many fields, investigate the distribution of plants for instance. Vague set extend the concept of fuzzy set, more flexible in terms of express and process fuzzy information. Some concepts such as vague spatial neighborhood relationship, vague participation ratio and vague participation index are defined in this paper first, after that, vague data filter rule and the algorithm of mining co-location patterns form vague set are proposed, the efficient prune strategy is presented in order to improve the efficiency of algorithm finally. The significance and efficiency of the techniques we propose are verified by experiments.

1. INTRODUCTION

Spatial co-location patterns representative a set of a group spatial feature, whose instances associated frequently in space, such as KFC tend to near to Wal-Mart in many cities. In fact, the task of mining spatial co-location patterns seeks to discover association relationship among objects in spatial database. However, traditional algorithms difficult to find the patterns from fuzzy data, since data in spatial database inaccurate in nature. Therefore, this paper aim to explore method of mining spatial co-location patterns in vague set, another representative of fuzzy data.¹

2. RELATED WORK

A great number of researchers have worked for improve the efficiency of mining co-location patterns, after this patterns proposed [2], a string of algorithms are proved efficient such as join-less [3], partial-join [4], order-clique-based [5] and CPI-tree [6]. Uncertain data processing is prevailing in data mining, authors proposed prune strategy and clustering pretreatment for uncertain value mining [7-8], deeply explored co-location patterns of possible world published in [9-10], patterns discovered from fuzzy objects that express by degree of membership and non-membership, sum of both degrees equal to 1.¹

In most cases, degree of membership adds to degree of non-membership is not necessarily 1. Suppose 5 out of 10 people vote favor for a bill, 3 against and 2 abstain, degree of membership is 0.5 and non-membership equal 0.3 in this situation, sum of the both degrees less than 1, only 0.8. This situation of sum of the two degrees within 1, Gau¹ call it vague set. So far, the research of vague set focus on similarity [12-13], distance [14-15], sorting and vague multi-attribute decision making [16-17]. Another study for algebra structure of vague set attract the attention of some scholars, such as vague soft ring [18], soft group [19]. A growing number of researcher interested in clustering [22-24] and mining
association rule \[^{[25]}\] from vague set.

3. DEFINITION

In this section, we define several related concepts about vague set and spatial co-location patterns.

**Definition 1 (Vague Set)** True membership degree \( t_v(x) \) and false membership degree \( f_v(x) \) are denoted as the bounds of membership degree, and satisfy \( 0 \leq t_v(x) + f_v(x) \leq 1 \), these two bounds constitute interval \([t_v(x), 1-f_v(x)]\), a sub interval of \([0,1]\). The voting model mentioned in section 2 represented as \([0.5, 0.7]\), where 0.5 is true membership degree and 1-0.7=0.3 is false membership degree, 1-0.5=0.3=0.2 denote as uncertainty degree. Obviously vague set is fuzzy set in nature when \( t_v(x)=1-f_v(x) \), both \( t_v(x) \) and \( 1-f_v(x) \) equal to 0 or 1 at the same time, vague set degenerates as traditional set, which enable people more flexibly to handle uncertain information.

**Definition 2 (Vague Filter Rule)** Vague set will be identical fuzzy set if all the data participate in computing regardless of how much the value is. Therefore vague data have to be filter before generate co-location patterns. The vague filter rules defined as:

1. For example, Amendments to the constitution must be passed by a two thirds majority. We call 2/3 as threshold, so \( t_v'(x)=t_v(x) \) if and only if \( t_v(x)\geq t_v(x)_\text{threshold} \) where \( t_v'(x) \) denotes true membership degree of vague object \( x \) after filter.

2. Another case when discussed where to travel in a family, may be obey the rule of the minority is subordinate to the majority, thus \( t_v'(x)=t_v(x) \) when \( t_v(x)>f_v(x) \).

These two rules are most common in daily life but take no account for the effect of uncertain degree. For instances, 10 people vote for a bill, 2 approvals and 1 oppose, so 7 people abstentions, this bill will be passed according to filter rule (2), because \( t_v(x)>f_v(x) \). However, such solution is unacceptable where support degree little more than 0.2. So new rule proposed to solve the problem.

3. Bill pass when true membership greater than false membership and true membership greater than threshold, formalize as \( t_v'(x)=t_v(x) \) if and only if \( t_v(x)\geq t_v(x)_\text{threshold} \) and \( t_v(x)>f_v(x) \).

**Definition 3** (Vague spatial neighborhood relationship, \( VR \)) Suppose that two vague object \( a \) and \( b \), if both of their true membership degree satisfy the same filter rule, and Euclidean distance between two objects less than distance threshold meanwhile, these two objects satisfy Vague spatial neighborhood relationship.

![Figure 1. Vague spatial neighborhood relationship](image)

**Example 1.** Spatial vague object plot as point in Figure 1, object instances ID compose of letter and number, vague value place beside ID, line between two points mean satisfy vague spatial neighborhood relationship. If we adapt filter rule (1) and set the true membership threshold is 0.5, there is no line connect \( a1 \) and \( b1 \) although they are very close, since that the true membership degree of object \( a \) less than 0.5. Similarity, \( a1 \) and \( c1 \) as well as do not satisfy the vague spatial neighborhood relationship.

**Definition 4** (co-location patterns) Spatial co-location patterns is a set \( c \) of a group of vague
objects, \{A, B, C\} is a co-location pattern in Fig 1. The number of vague objects is order of pattern, the order of pattern \{A, B, C\} is 3. Suppose a spatial instance \(I=\{i_1, i_2, \ldots, i_n\}\), if satisfy

\[
\{VR(i_j, i_k) \mid 1 \leq j \leq m, 1 \leq k \leq m\}
\]

then call \(I\) as clique. In the event of all attributes of co-location pattern \(c\) included in clique \(I\), \(I\) is a row instance of co-location pattern \(c\), a set that constitutes all the row instances of pattern \(c\) define as table instance, denote as table_instance(c).

**Example 2.** As showed the Fig1, table instance of co-location pattern \{A, C\} is \{\{a2, c1\}, \{a3, c2\}\}, for 3-order co-location pattern \{A, B, C\}, \{\{a2, b2, c1\}\} is table instance.

**Definition 5 (Vague participate ratio, VPR)** Assume that \(v_i\) is a vague object, vague participate ratio of \(v_i\) in k-order co-location pattern denote as VPR(c, \(v_i\)), defined as following

\[
VPR(c,v_i) = \frac{\sum t'_v(x)}{\text{table_instance(\{\(v_i\}\})}}
\]

Where \(\sum t'_v(x)\) is the sum of the true membership degree of non-repeatable instance of \(v_i\) in table instance of pattern \(c\), denominator is total instance number in \(v_i\). Furthermore, \(x \in \Pi_{v}(\text{table_instance(c)})\), \(\Pi\) is projection of relationship.

**Definition 6 (Vague participate index, VPI)** The VPI of co-location pattern \(c = \{v_1, v_2, \ldots, v_k\}\) defined as minimum value of VFR for all the vague objects,

\[
VPI(c) = \min(VFR(c,v_i)), 1 \leq i \leq k
\]

Given that min VPI is the participate threshold set by user, co-location pattern \(c\) is frequent pattern when \(VPI(c) \geq \text{min VPI}\).

**Example 3.** Continue to use Fig 1 to illustrate, for co-location pattern \{A,C\}, \(VFR(c, A) = (0.5+0.7)/3 = 0.4\), \(VFR(c, C) = (0.7+0.5)/2 = 0.6\), \(VPI(c) = \min(0.4,0.6) = 0.4\). Also easy to compute that for co-location pattern \{A,B,C\}, \(VFR(c, A) = 0.5/3 = 0.17\), \(VFR(c, B) = 0.6/2 = 0.3\), \(VFR(c, C) = 0.7/2 = 0.35\), \(\min(0.17,0.3,0.35) = 0.17\). If \(\text{min VPI} = 0.3\), co-location pattern \{A, C\} is frequent but \{A, B, C\} is not.

4. CO-LOCATION PATTER MINING FROM VAGUE SET

Firstly, we introduce a join base algorithm of mining spatial co-location pattern in vague set, and then propose prune strategy to enable the method more efficient.

4.1 VJoin-base Algorithm

The core of VJoin-base algorithm is candidate pattern generation. Candidate patterns of \(k\) order generated by connect two \(k-1\) order patterns with the same \(k-2\) order, and required last instance between two patterns have the relationship of vague neighborhood.

\[
\begin{array}{ccc}
A & B & A \quad C \\
2 & 2 & 1 \\
3 & 2 & 1 \\
\end{array}
\]

\[
\begin{array}{ccc}
B & C & A \quad B \quad C \\
2 & 1 & 2 \\
2 & 2 & 1 \\
\end{array}
\]

Figure 2. Generation for 3-order candidate patterns

**Example 4.** 3-order candidate pattern generation show in figure 4. Pattern \{A, B\} and \{A, C\} have common instance \(a2\), instance of \(b2\) and \(c1\) is one of row instances of pattern \{B,C\}, so \(a2, b2, c1\) constitute a row instance of candidate pattern \{A,B,C\}.

This algorithm generate vague set firstly, then filter vague data according to the filter rule, after that, generate 2-order candidate pattern and frequent pattern by using vague participate index, finally
generate high order candidate patterns based on low order patterns until no new patterns generated.

Algorithm 1 VJoin-based algorithm

Input: Vague set Vdata, Spatial instance set I, Vague instance set after filter VI, Vague participate index threshold VPI threshold, distance threshold d_threshold, True membership degree threshold t_threshold.
Output: Frequent co-location PC.
Variables: order of pattern k, k-order frequent co-location pattern Pk, k-order candidate co-location pattern Ck, table instance set of co-location pattern in CkTk.
1. Vdata, I=gen_data();
2. VI=filter_Vdata(t_threshold, I);
3. P1=VFI, FC=∅;
4. for (k=2; Pk;i≠∅; k++)
   4.1 Ck=gen_candidate_co-location (Pk, i);
   4.2 Tk=gen_table_instance (Tk, Ci);
   4.3 Pk=gen_prevailant_co-location (VPI_threshold, Tk);
5. return ∪P

Suppose that c1 generated by connecting c1 and c2, k1, k2, k3 denote the number of row instances for c1, c2 and c3 respectively, the join operation for the algorithms mean that connect two table instances, and search the frequent patterns in table instances of c3, each row takes k-2 times of comparison while connecting table instances. So the time complexity of this algorithm is

\[ O((k-2) \times k_1 \times k_2 \times k_3) \]

This complexity have poor performance while increasing the number of instances, we introduce prune strategy to improve the efficiency in next section.

4.2 Vjoin-base-prune Algorithm

1 lemma and 2 theorem are proposed in this section, and introduced pruned method based them.

**Lemma 1.** Vague participate ratio and Vague participate index decrease monotonously with the increase the order of co-location patterns.

**Prove.** Suppose that a vague object included in row instance of co-location pattern c, if \( c' \subseteq c \), then this instance must also included in row instance \( c' \), but inverse is not. Therefore, vague participate ratio decrease when increasing the order of co-location patterns, and vague participate index defined as the minimum of all the vague participate ratios, so lemma proved.

**Theorem 1.** If k-order co-location pattern is frequent, then arbitrary sub patterns are frequent.

**Prove.** If k-order co-location pattern is frequent, according to the definition, its vague participate index great than or equal to vague participate threshold. Vague participate index of k-order pattern less than Vague participate index of arbitrary sub patterns, based on lemma 1, so its arbitrary sub co-location patterns are frequent.

**Theorem 2.** If co-location pattern is non-frequent, and \( c' \supseteq c \), then pattern \( c' \) must be non-frequent.

**Prove.** Because \( c' \supseteq c \), so \( VPI(c) \leq VPI(c') \) holds, if c is non-frequent, so \( c' \) is non-frequent too.

**Example 4.** According to example 3, vague participate index of pattern \( \{A, C\} \) equal to 0.5, if vague participate index threshold set as 0.5, then assert that pattern \( \{A, C\} \) is non-frequent, so pattern \( \{A, B, C\} \) is as well as non-frequent on the basis of theorem 2.

Summary up the lemma and theorems, prune strategy of algorithm 1 described as following:

Algorithm 2 VJoin-based-prune algorithm

Input: k-order candidate co-location pattern Ck, k+1-order candidate co-location pattern Ck+1, vague participate index threshold VPI threshold.
Output: k+1-order frequent co-location pattern Pk+1.
1. if \( VPI(C_k) < VPI\_threshold \) and \( c_k \subset c_{k+1} \) \( P_{k+1}=\emptyset \);
2. return \( P_{k+1} \);
5. EXPERIMENTS
Experiments executed in Windows 7 with memory of 4GB, and using C#. Data generated randomly 2dimensional space over [0,100] × [0,100], parameters and its default value show as following.

Table 1. Parameters and its default value

| Parameters                           | Default value |
|--------------------------------------|---------------|
| Number of instances                  | 300           |
| True membership threshold            | 0.4           |
| Distance threshold                   | 10            |
| Vague participation index threshold  | 0.4           |

5.1 Effect of Different Filter Rules
Firstly, we investigate the effect of different filter rules to mining result, using vague data set as table 1. In figure 3 shows the result of different rules, r1, r2 and r3 representative three filter rules in definition 2, data have not filtered for r4, patterns number display in Y-axis. Algorithms that without filer data find the patterns more than filter it according to this figure, but this makes no sense, since that is not vague set absolutely. The third rules explores least numbers among three different filter rules, because Vague set have to satisfied true membership threshold and true membership greater than false membership.

![Figure 3. Effect of different rules](image)

5.2 Performance Comparison for Two Algorithms
Experiments evaluate the effect of parameters to algorithms in this session, and test the performance of prune methods. Other parameters are fixed when change the certain parameter, and the third filter rule adopt in definition 2.

Effect of instances number to algorithms showed in figure 4. The time increasing sharply with the growing instances number intuitively, the more instance, the more co-location patterns generated. Furthermore, the running time of two algorithms very close, since that the low-order candidate are frequented mostly, therefore the join computing ignored very few in the processing of generate high-order candidates.
Figure 4. Effect of instances number

Figure 5 demonstrates the outcomes of different true membership threshold. Both two algorithms convergence earlier by increasing the true membership threshold, easy to know that fewer instances left with greater threshold and the few instances leads to few high-order candidates, so the time decrease gradually. By the mean time we find that time of base algorithm descends smoothly but prune method drop down when threshold from 0.3 increases to 0.4, indicates the massive candidates pruned during the period of increase threshold 0.3 to 0.4.

Figure 5. Effect of true membership threshold

The execute time of VJoin-base algorithm growing quickly while increasing distance threshold according to the figure 6. For the larger distance threshold is, the more instances satisfy vague spatial neighborhood relationship, and cost more time to join the candidate patterns. However, distance threshold has little effect to VJoin-base-prune algorithm, a mount of candidates generated while vague participation index equals 0.4, but the vague participation index of these candidates co-location patterns less than 0.4, so they are pruned while generated higher order candidates.
Look at the figure 7, we obviously find that the running time of base algorithm have no change while increase the vague participation index threshold, all the join computing have to complete whatever the vague participation index threshold is, therefore there is any candidate patterns to be pruned. By contrary, vague participation index threshold plays an important role in prune algorithm, when the threshold of vague participation increases, the execution time of pruning algorithm decreases. Especially the threshold value change from 0.3 to 0.4, time fall down sharply because of large number of candidate patterns, whose vague participation index between 0.3 and 0.4, and pruned in the next step, this phenomenon as well as explain the figure 6, in which time is almost static.

May be it is not very meaningful to discuss the time within 1 seconds, for the default value of instance number is 300, two algorithms have distinct performance when we set the instance number
equals to 600, this figure illustrates the prune algorithm improve obviously while we have large number of instance and distance threshold.

6. CONCLUSIONS
Vague set could be handle fuzzy information flexibly. Method of mining spatial co-location patterns from vague set discussed in this paper, algorithm generate co-location patterns after data filtering, and introduce prune strategy to reduce the join cost. More efficient will be considered in the future work, and the nature of vague set continue to be explored.

ACKNOWLEDGMENTS
Research supported in part by Key Laboratory of Intelligent Manufacturing in Universities of Yunnan Province, Business Intelligence Technology Innovation Team Fund of Yunnan Province, Scientific Research Fund of Yunnan Education Department(#2019J0364), and Research Fund of Education and Teaching Reform of Oxbridge College, KMUST.

REFERENCES
[1] Gau W L, Buehrer D J. Vague sets[J]. IEEE Trans on Systems Man and Cybernetics, 1993, 23(2):610-614.
[2] Huang Y, Shekhar S, Xiong H. Discovering co-location patterns from spatial data sets: A general approach. IEEE Trans on Knowledge and Data Engineering, 2004, 16 (12): 1472 - 1485.
[3] Yoo J S, Shekhar S, Celik M.A Join-Less Approach for Co-Location Pattern Mining: A Summary of Results//Proceedings of the 5th IEEE International Conference on Data Mining, ICDM 2005, Houston, USA, 2005: 813-816.
[4] Yoo J S, Shekhar S.A partial Join Approach for Mining Co-location Patterns// Proceedings of the international workshop on Geographic information systems, Washington, USA, 2004: 241-249.
[5] Wang Li-zhen, Zhou Li-Hua, Lu J, Yip J. An order-clique-based approach for mining maximal co-locations. Information Science, 2009, 179(19):3370-3382.
[6] Wang Li-zhen, Bao Yu-Zhen, Lu J, Yip J. A New Join-less Approach for Co-location Pattern Mining//Proceedings of the IEEE 8th International Conference on Computer and Information Technology (CIT2008), Syney, Australia, 2008:197-202.
[7] Lu Ye, Wang Li-zhen, Zhang Xiao-feng. Mining Frequent Co-location Patterns from Uncertain Data. Journal of Frontiers of Computer Science and Technology, 2009, 3(6):656-664.
[8] Gao Shi-jian, Wang Li-zhen, Xiao Qing. A U-AHC-based Algorithm for Mining Spatial Co-location Patterns over Uncertain Data. Journal of Computer Research and Development, 2011, 48(Supplement): 60-66.
[9] Lu Ye, Wang Li-zhen, Chen Hong-mei, Zhao Li-hong. Spatial Co-location Patterns Mining over Uncertain Data based on Possible Worlds. Journal of Computer Research and Development, 2010,47(Supplement):215-221.
[10] Wang Li-zhen, Wu Ping-ping, Chen Hong-mei. Finding Probabilistic Prevalent Co-locations in Spatially Uncertain Data Sets. IEEE Trans on Knowledge and Data Engineering, 2013, 25(4): 790-804.
[11] OUYang Zhi-ping, Wang Li-zhen, Chen Hong-mei. Mining Spatial Co-location Patterns for Fuzzy Objects. Chinese Journal of Computers, 2011,10(34):1947-1955.
[12] Wang W, Peng J. A new Similarity Measure between Vague Sets based on Information Similarity Coefficient and its Application//Proceedings of the 3rd International Conference on Frontiers of Manufacturing Science and Measuring Technology, LiJiang,China,2013: 959-963.
[13] Zhou Meng, Yu Jian-kun. A new Vague Set Similarity measurement. Computer Applications and Software, 2012, 29(2):138-141.
[14] Papakostas G, Hatzimichailidis A, Kaburlasos V. Distance and Similarity Measures between Intuitionistic Fuzzy Sets: a Comparative Analysis from a Pattern Recognition Point of View.
Pattern Recognition Letters, 2013, 34(14): 1609-1622.

[15] Hatzimichailidis A, Papakostas G, Kaburlasos V. A Novel Distance Measure of Intuitionistic Fuzzy Sets and its Application to Pattern Recognition Problems. International Journal of Intelligent Systems, 2012, 27(4): 396-409.

[16] Selvachandran G, Salleh A. A Vague Soft Set Theoretic Approach to Multiattribute Decision Making Problems. Applied Mathematical Sciences, 2014, 8:6937-6949.

[17] Alhazaymeh K, Hassan N. Possibility Vague Soft Set and Its Application in Decision Making. International Journal of Pure and Applied Mathematics, 2012, 77(4): 549-563.

[18] Selvachandran G. Vague Soft Rings and Vague Soft Ideals. International Journal of Algebra, 2012, 6(12):557-572.

[19] Biswas R. Vague Groups. International Journal of Computational Cognition, 2006, 4(2):20-23.

[20] Hao Zhong-xiao, Li Song. Representation and Compound Reasoning of the Vague Temporal Interval Relations and the Vague Region Relations. Chinese Journal of Computers, 2014, 37(8):1743-1753.

[21] Li Song, Zhang Li-ping, Hao Xiao-hong, Hao Zhong-xiao. Representation and Compound Reasoning of Vague Region Relations and Direction Relations. Journal of Computer Research and Development, 2015, 52(4):918-928.

[22] Wang Chang, Liu Ya-ya. Similarity measure of Vague sets and its application in clustering analysis. Computer Engineering and Applications, 2011, 47(29):8-11.

[23] Yu Jian-kun, Shen Xiao-hu. Direct clustering method based on vague sets. Computer Engineering and Application, 2010, 46(31):143-147.

[24] Liu Xin-lin, Liang Jia-rong. Optimal tree clustering based on Vague diagram. Computer Engineering and Application, 2016, 52(3):123-126.

[25] Shen Xiao-hu, Yu Jian-kun. Improvement of Apriori algorithm based on Vague sets. Computer Engineering and Applications, 2010, 46(25):131-133.