

The Research of Generalized Attributive Concept Lattice

Wang Hailong¹ and Meng Fanjun²

1 College of Network Technology, Inner Mongolia Normal University, Hohhot, China

*2 Computer and Information Engineering College, Inner Mongolia Normal University, Hohhot, China
lzjtuwhl@163.com*

Abstract

In order to improve the disadvantage in the process of constructing concept lattice which produces enormous meaningless association rule, based on the analysis of the existing concept lattice and its improved model, attribute-oriented generalization is introduced into the research of concept lattice. A new lattice is proposed, which is called Generalized Attributive Concept Lattice. At the same time, qualities of Generalized Attributive concept lattice are proved. The constructing theory of Generalized Attributive concept lattice is analyzed. A novel algorithm of constructing concept lattice is proposed.

Keywords: *attribute-oriented generalization; concept lattice; association rule*

1. Introduction

Attributive-oriented generalization (namely AOG) is a method of replacing the detailed information in records with relatively high-level concept hierarchy tree according to the concept hierarchy tree of various attributives.

The construction of concept lattice can be generalized into three types: the first can generate complete lattice, for instance, the ParGal algorithm [1] proposed by Njiwoua *et al.*, DisGodin algorithm [2] proposed by Zhang Qing, *etc.* Such algorithm may generate new concept node when merging the sub-lattice, and it is sensitive to the changes in formal context. As a result, the whole performance decreases drastically with the increasing lattice scale. The second can generate complete conceptual set, such as the NextClosure algorithm [3] proposed by Ganter. Such algorithm may generate substantial large-scale and intermediate concept for union during the construction process, and it is sensitive to the attributive changes in the formal context. However, it is limited in practical application. The third may generate concept tree, such as the LCBI algorithm [4] proposed by Tan Zhe, *et al.*, Such algorithm can simplify substantial pilot process, but it is sensitive to attributive changes. In this paper, a new lattice structure is proposed based on the three methods of constructing concept lattice, namely generalized attributive concept lattice, and meanwhile, a construction algorithm of the generalized attributive concept lattice (BGACL) is put forward.

2. Definition and Nature of Generalized Attributive Concept Lattice

Definition 1: $K = (O, A, R)$ is named as the generated attributive formal context, in which $O = \{o_1, o_2, \dots, o_n\}$ is the finite non-empty set of object, $A = \{a_1, a_2, \dots, a_n\}$ is the finite non-empty set of attributes, R is defined as the mapping between O and D , namely $R : O \times A \rightarrow X$, here, X is a set of generated attributive category $\{x_1, x_2, \dots, x_n\}$.

Definition 2: The binary group (X, Y) satisfies $X = g(Y), Y = f(X)$, in which, $X = 2^A, Y = 2^O$, and then (X, Y) is regarded as one of the generated attributive concepts of the generated attributive formal context $K = (O, A, R)$, X is the extension of the concept, and Y is the connotation of the concept [5].

Definition 3: The lattice $L(O, A, R, f, g)$ formed by the generalized attributive formal context $K = (O, A, R)$ is named as the generalized attributive concept lattice.

Definition 4: In the attributive discrete interval, the attribute a is discretized, the interval $[O_{\min}, O_{\max}]$ formed by the maximum and minimum limit O_{\max} and O_{\min} of the finite set O is the discrete interval of the finite set O .

Definition 5: For a point $\eta (\eta \in O)$ in the interval $[a, b]$, if there are δ elements in O of the sub-interval $[\eta - \frac{1}{2}\theta, \eta + \frac{1}{2}\theta]$, η is called the accumulation point on the finite set O , while θ and δ are the generalized attributive parameters.

Definition 6: The set of the accumulation point on the finite set O is the accumulation point set.

Definition 7: The generalized attributive operator is defined as:

$$c_j = \frac{\sum_{i=1}^n \lambda_{ij}^\alpha x_k}{\sum_{i=1}^n \lambda_{ij}^\alpha}, \text{ in which, } \lambda_{ij} = \frac{\theta_{ij}}{\delta_{ij}} = \frac{1}{\sum_c^{k=1} \left(\frac{\|x_k - c_j\|^2}{\|x_k - c_k\|^2} \right)^{\frac{2}{\delta-1}}},$$

α is the iteration parameter, and x_k is any of the element on the i level [6].

Theorem 1: For given generalized attributive parameter θ and δ , in a formal context $K = (O, A, R)$, the extension of any element a in the attribute set is discretized as the finite set O , and there are finite accumulation points in its corresponding discrete interval $[O_{\min}, O_{\max}]$.

Prove: according to definition 4, $[O_{\min}, O_{\max}]$ is divided into two parts equally,

$$\left[O_{\min}, \frac{[O_{\max} - O_{\min}]}{2} \right], \left[\frac{[O_{\max} - O_{\min}]}{2}, O_{\max} \right]$$

If there are δ elements in the first part, it will continue to divide equally, or it will stop. Generally, it is assumed that the first part exists

$$\text{in an equal division, and then a nest of intervals will be formed } \left[O_{\min}, \frac{[O_{\max} - O_{\min}]}{2^n} \right].$$

When n is quite large, $(n > N), \frac{1}{2^n} < \frac{\theta}{2}$. It can be learnt from the construction method of

the nest of intervals, when $n > N$, there will be an accumulation point in the $\left[O_{\min}, \frac{[O_{\max} - O_{\min}]}{2^{N+1}} \right]$, and meanwhile, the interval $[O_{\min}, O_{\max}]$ will be divided into $N+1$ sub-interval, and as a result, there are at least $N+1$ accumulation points, the end [7].

Deduction 1: Different generalized attributive parameters θ and δ can form different accumulation points.

Prove: It can be learnt from the proof process of Theorem that different θ and δ may generate different N , and obtain different accumulation points.

Theorem 2: An accumulation point forms a concept

Prove: it can be learnt from definition 3 and definition 6.

For $\forall \alpha \in \left[\eta - \frac{1}{2}\theta, \eta + \frac{1}{2}\theta \right]$, and then $f(\alpha) = \eta$, as a result,

The binary group $\left(\left[\eta - \frac{1}{2}\theta, \eta + \frac{1}{2}\theta \right], \{\eta\} \right)$ is a concept.

Theorem 3: There is n (integer, $n \geq 2$), and there will be a group of θ and δ that can generalize the attributive discrete interval $[O_{\max}, O_{\min}]$ into n -level conceptual tree.

Prove: according to the construction of interval division in theorem 1, the attributive discrete interval $[O_{\max}, O_{\min}]$ is divided according to θ and δ . It can be learnt from theorem 1 that there are at least $N+1$ accumulation point sets. According to deduction 1, J forms the first-level concept set. Now, one of the accumulation points η is taken, and it can be learnt from definition 5 that $\left[\eta - \frac{1}{2}\theta, \eta + \frac{1}{2}\theta \right]$ is also the discrete interval. Meanwhile, according to theorem 1, θ_{η_1} and δ_{η_1} form finite accumulation point set J_{η} in the $\left[\eta - \frac{1}{2}\theta, \eta + \frac{1}{2}\theta \right]$, and then the first-level concept set is formed.

According to deduction 1, the second-level concept set is derived from the accumulation points η .

And so on, the concept set of the n^{th} level is obtained, the end.

3. One Construction Method of Generalized Attributive Concept Lattice

3.1. Basic Principles and Process of the Algorithm

After a certain attribute a in the attribute set A is selected as the generalized attributive target, determine if attribute a is a generalized attribute or objective generalized attribute. According to the type of attribute a , the generalized attributive process based on the cloud model is conducted. The concept lattice of information table K is constructed according to the progressive construction algorithm. The steps are shown as follows:

Input: Formal context K

Output: The lattice node after attribute generalization

Begin

Step1: Pre-process the original data set, discretize the data into new formal context.

Suppose there are n attributes in the formal context;

For $1 \leq i \leq n$

Suppose the i^{th} attribute is a , and now a is selected as the target of the attribute generalization (definition 3).

The concept central set $\{C_i(a, N, M_m)\}$ (a stands for the attribute, N stands for the total conceptual level divided by the attribute a , M_m stands for the concept set of the m^{th}

level) is calculated according to the generalized attributive operator $c_j = \frac{\sum_{i=1}^n \lambda_{ij}^\alpha x_k}{\sum_{i=1}^n \lambda_{ij}^\alpha}$ in

definition 7 [8].

End For

Step 2: Scan the data set, establish the node $l_{NEW} (O_{NEW}, A_{NEW})$, and for any existing nodes $l_{OLD} (O_{OLD}, A_{OLD})$, there are three conditions:

- i. If $A_{OLD} \subseteq A_{NEW}$, and then A_{OLD} remains unchanged, and $O_{OLD} = O_{OLD} \cup O_{NEW}$;
- ii. If $A_{OLD} \cap A_{NEW} = \Phi$, and then l_{OLD} remains unchanged.
- iii. If it does not satisfy i, ii shall increase the nodes $(O_{OLD} \cup \Theta_{NEW}, A_{OLD} \cup A_{NEW})$ [9].

Step 3: For any newly added nodes, the parent node and sub-node shall be found, and one side shall be added respectively to the parent node.

Step 4: The minimum confidence coefficient threshold \min_conf is defined, the association rules are deduced, and then $A \Rightarrow B$, in which $A, B \subseteq 2^Y$, $B = Y - A$, and

$$\frac{g(Y)}{g(A)} > \min_conf .$$

Step 5: All relevant rules shall be deduced.

End

3.2. Basic Model of the Algorithm

Basic model of the algorithm shows in Figure 1.

It can be seen from the basic model of the algorithm in Figure 1 that this algorithm can be divided into four steps: In the first step, the original formal context shall be generalized; in the second step, the original attribute set shall be established according to the generalized attribute operator; in the third step, it shall match the attribute set of the newly-added formal context. In the fourth step, corresponding attribute set operations shall be conducted according to the conditions of three different newly-added attribute set, thus to form new attribute set [10].

- i. If the intersection of newly-added attribute set (A_{NEW}) and old attribute set (A_{OLD}) is empty, the new attribute set should be added in the old set.
- ii. If the newly-added attribute set (A_{NEW}) belongs to the old attribute set (A_{OLD}), the original attribute set will remain the same [11].

iii. If the intersection of newly-added attribute set (A_{NEW}) and old attribute set (A_{OLD}) is not empty, and the new attribute set does not belong to the old attribute set, and then the new and old attribute set shall be combined. For newly-added nodes, minimum confidence coefficient min_conf is adopted for the induction of association rules [12].

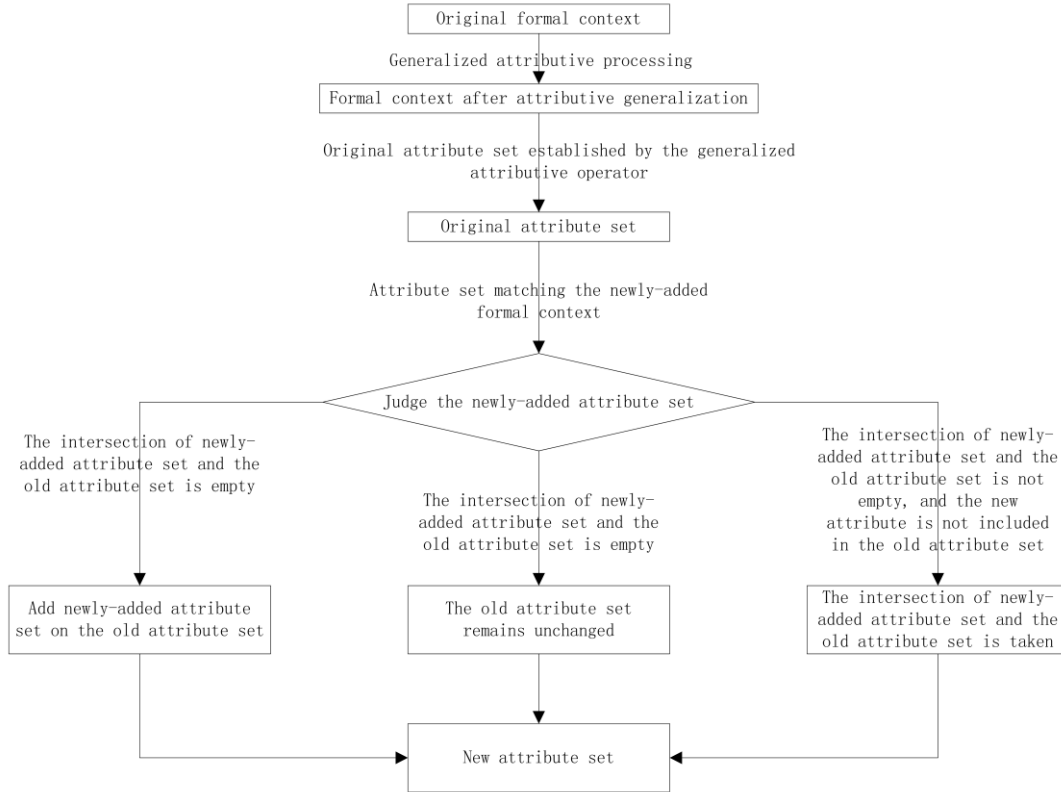


Figure 1. Basic Model of the Algorithm

3.3. Experiment and Analysis

In the window 2003 operating system, the database ORACLE10g, and programming language JAVA, the comparison of GODIN algorithm is realized, DISGODIN algorithm and BGACL is conducted.

The sample data extracted randomly is selected as the experimental data for testing the time variation of the three algorithms with the increasing data in time of unchanged attributes, as well as the time variation of the three algorithms with the number of attributes in time of unchanged records. In addition, the time needed by the three algorithms is the average value of three tests [13].

When the number of records remains unchanged (3000), the time comparison table of the three algorithms shows in Table 1.

It can be seen from the data in Table 1 that the BGACL algorithm proposed in this paper has certain advantages when compared to the classic GODIN algorithm and DISGODIN algorithm. Meanwhile, with increasing data size, the advantage of the algorithm grows more and more evident. It can be learnt from the experimental result that when compared to the classic algorithm, the advantage of this algorithm is: on one hand, the data in the formal

context is divided rationally with the generalized attributive operator, and as a result, much reasonable generalized formal context is obtained, which simplifies the calculated amount. On the other hand, according to the size of the formal context, the minimum confidence coefficient threshold value *min_conf* is modified dynamically through the parameter operation, which accelerates the speed of inducing the association rules [14].

Table 1. When the Number of Records Remains Unchanged (3000), the Time Comparison Table of the Three Algorithms

Attribute	GODIN (ms)	DISGODIN(ms)	BGACL(ms)
10	48753	8202	6009
20	289876	29500	12541
30	345302	75211	55412
40	6845713	268472	256840

When the number of attributes (20) remains unchanged, the time comparison table of the three algorithms shows in Table 2.

Table 2. When the Number of Attributes (20) Remains Unchanged, the Time Comparison Table of the Three Algorithms

Records	GODIN (ms)	DISGODIN(ms)	BGACL(ms)
2000	12800	5002	5421
4000	30294	8501	6877
6000	92010	13541	9420
8000	560041	196874	11204

It can be seen from Table 2 that when compared to the classic algorithm GODIN and DISGODIN, this algorithm have certain advantages. When the number of attributes remains unchanged, the time variation speed is relatively slow with the increasing records, for the changes of records has little influence on the generalized attributive results, and generally, there won't be substantial newly-added leaf nodes [15].

4. Summarization and Conclusion

The construction of traditional concept lattice may generate substantial valueless association rules. In this paper, based on the attributive-oriented generalization, such a rule can be discovered during the data research process: a conceptual distance algorithm can be discovered (theorem 3), and it is calculated that the distance of the two concepts is smaller than a certain threshold, and then it can be judged that the two concepts fall into the same type. According to such repeated process, the complex attribute combination can generalize regular attribute combination (or it is called concept). In the end, valuable association rules can be discovered from the database. Such method is especially applicable for discovering the association rules from the high-volume database. In this paper, the construction of distributed generalized attributive concept lattice will be the emphasis of the follow-up work.

In this paper, the algorithm of this paper is main different from the traditional GODIN and DISGODIN algorithm in the generalized attributive process. The traditional generalized attributive process is mainly realized with manual intervention, and it is quite objective and difficult for realizing the machine learning for the complicated formal background with enormous concept. Therefore, a new generalized attributive operator is defined in this paper,

and it is applied for dividing the data in the context. As a result, rational division method is obtained for certain discussion about the realization of machine learning.

Acknowledgments

This research was supported by Scientific Research Project of Higher Education of Inner Mongolia Autonomous Region, China (NJZY13052) (NJZY14034).

References

- [1] P. Njiwoua and E. M. Nguifo, "A parallel algorithm to build concept lattice", Proceedings of 4th Groningen Intl Information Tech., (1997), pp. 103-107.
- [2] Z. Zhenyu, "Study on the Concept Lattice Model based on the Attribute Generalization", Hefei University of Technology, (2011).
- [3] B. Ganter, "Formal: Algorithmic aspects", <http://www.math.tude/ganter/el02/>, (2009).
- [4] T. Zhe and H. Xuegang, "Distributed Algorithm for Constructing Concept Lattice Based on Index", Computer Application, vol. 5, (2009), pp. 1409-1411.
- [5] P. Njiwoua and E. Nguifo, "A parallel algorithm to build concept lattice", Proceedings of 4th Groninged International Information Technical Conference for Students, (2010), pp. 103-107.
- [6] H. Xuegang, W. Xiang and W. Dexing, "A parallel algorithm to construct concept lattice", IEEE Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, (2011), pp. 119-123.
- [7] B. Peter, P. Jozef and P. Jana, "On equivalence of conceptual scaling and generalized one-sided concept lattices", Information Sciences, vol. 259, no. 5, (2014), pp. 57-70.
- [8] F. Li and Z. Liu, "Concept Lattice Based on the Rough Sets", International Journal of Advanced Intelligence, vol. 1, no. 1, (2009), pp. 141-151.
- [9] W. Zhang, Y. Zhang and T. Kim, "Detecting bad information in mobile wireless networks based on the wireless application protocol", Computing, <http://dx.doi.org/10.1007/s00607-013-0325-1>, (2013).
- [10] M. D. S. Obiedkov and D. Kourie, "AddIntent: A New Incremental Algorithm for Constructing Concept Lattices", Notes in Computer Science, Springer Berlin Heidelberg, (2004), pp. 372-385.
- [11] P. Valtchev, R. Missaoui and P. Lebrun, "A partition-based approach towards constructing Galois (concept) lattices", Discrete Math, vol. 256, no. 3, (2002), pp. 801-829.
- [12] B. Ganter, "Two basic algorithms in concept analysis", Proceedings of the 8th international conference on Formal Concept Analysis. ICFA-10, Berlin, Heidelberg, Springer-Verlag, (2010), pp. 312-340.
- [13] R. Godin, R. Missaoui and H. Alaoui, "Incremental concept formation algorithms based on Galois (Concept) lattices", Computational Intelligence, vol. 11, no. 2, (1995), pp. 246-267.
- [14] W. Zhang, X. Li, H. He and X. Wang, "Identifying Network Public Opinion Leaders Based on Markov Logic Networks", The Scientific World Journal, Article ID 268592, 8 pages. [http://dx.doi.org/10.1155/\(2014\)/268592](http://dx.doi.org/10.1155/(2014)/268592), (2014).
- [15] G. Ammons, D. Mandelin, R. Bodik and J. Larus, "Debugging temporal specifications with concept analysis", In ACM SIGPLAN Conference on Programming Language Design and Implementation, (2003), pp. 182-193.

Authors



Hailong Wang, he received the BS in computer science from North Jiaotong University, China, in 1998, and received the MS in computer science from Lanzhou Jiaotong University, China, in 2007. Currently, he is an assistant professor in Network Technology College at Inner Mongolia Normal University, China. His research interests include embedded system and multi-core processors, and also fault tolerance and real-time database.



Fanjun Meng, he received the BS and MS degrees in computer science from Inner Mongolia Normal University, China, in 1999 and 2007. Currently, his research interests include scheduling techniques and parallel algorithms for clusters, and also multi-core processors and software techniques for I/O-intensive applications.