An Incremental Algorithm for Concept Lattice Based on SSIM

Yu Hu (✉ huyu@bistu.edu.cn)

Beijing Information Science and Technology University <https://orcid.org/0000-0001-9978-0833>

Yan Zhu Hu

Beijing University of Posts and Telecommunications

Zhong Su

Beijing Information Science and Technology University

Xiao Li Li

Beijing University of Technology

Zhen Meng

Beijing Information Science and Technology University

Wen Jia Tian

Beijing University of Posts and Telecommunications

Research Article

Keywords: concept lattice, incremental algorithms, structural similarity index, edge detection, Fourier Descriptor

Posted Date: November 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-958313/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An incremental algorithm for concept lattice based on SSIM

Yu Hu¹, Yan Zhu Hu², Zhong Su¹, Xiao Li Li³, Zhen Meng¹, Wen Jia Tian²

¹School of Automation, Beijing Information Science and Technology University, Automation college, No. 12 Qing e Xiao Ying East Road, Beijing/100196, China.

² School of Automation, Beijing University of Posts and Telecommunications, No. 10 Xi Tu Cheng Road, Beijing/100876, China.

³The information department, Beijing university of technology, 100 Ping Le Yuan, Beijing/100124, China.

corresponding author: Yu Hu; E-mail: huyu@bistu.edu.cn;

Yan Zhu Hu, E-mail: yzhu@263.net

Abstract As an effective tool for data analysis, Formal Concept Analysis (FCA) is widely used in software engineering and machine learning. The construction of concept lattice is a key step of the FCA. How to effectively update the concept lattice is still an open, interesting and important issue. The main aim of this paper is to provide a solution to this problem. So, we propose an incremental algorithm for concept lattice based on image structure similarity (SsimAddExtent). In addition, we perform time complexity analysis and experiments to show effectiveness of algorithm.

Keywords: concept lattice; incremental algorithms; structural similarity index; edge detection; Fourier Descriptor

1 Introduction

Formal concept analysis is an effective method of data analysis based on lattice theory framework, it derives from the philosophical description of the concept. Conceptual hierarchies can be extracted from the formal context of a particular domain represented by a binary relationship. Concept lattice is the core data structure of formal concept analysis, it describes the essential relationship between objects and attributes, shows generalization and instantiation between concepts. Hasse diagram is used to visualize concept lattice. The rule acquisition methods which based on concept lattice has been widely applied to many areas, such as data mining[1], software engineering[2], linguistics[3], ontology engineering[4], bioinformatics[5][6]and information retrieval[7][8]. Poelmans[9] summarizes the wide application of FCA-based methods in different fields. However, existing concept lattice construction methods have many disadvantages, such as redundant rule acquisition, noise sensitivity, time and space complexity.

At present, the research about concept lattice mainly includes the construction of concept lattice construction, knowledge discovery based on concept lattice, generalization of concept lattice model, etc. the construction of concept lattice is one of the core tasks of the FCA, but the number of concepts of a formal context increases exponentially with the size of formal context. So far, many researchers have proposed a variety of concept lattice construction algorithms, those algorithms can be roughly divided into two categories: batch and incremental algorithms. Batch algorithm first generates all the concepts, then generates edges based on their direct precursor and successor relationships, and finally completes concept lattice construction. The batch algorithm constructs concept lattice from top to bottom or from bottom to up on given data, the concept lattice needs to be rebuilt when adding new objects. So, there is no flexibility for dense formal context. The incremental algorithm was proposed by Godin[12] in 1995. The key to the efficiency of the incremental algorithm is to generate

position of the added object and update the edge of node. Some researchers have made many improvements to the Godin algorithm. XieZhiPeng[13] uses a tree construction to organize conceptual grid nodes, which can greatly reduce the search space for searching father and child nodes of added object. Derrick[14] proposed a concept lattice generation algorithm based on set intersection. Li Yun[15][16] proposed an attribute-based concept lattice incremental algorithm and multi-concept lattice horizontal merging algorithm, especially illustrating the effect of attributes splitting strategy on concept lattice construction efficiency. Yuan[17] proposed an object-based efficiency improvement mechanism for concept lattice by tracking the maximum concept. These incremental generation algorithms can dynamically update and maintain on the basis of the original concept lattice, which is suitable for recording the background of the incremental transaction database, and the efficiency of the algorithm is improved to varying degrees. So, the algorithm is the main focus of this paper. However, the existing algorithm have universal defects: high computational complexity, high degree of redundancy in extraction rules, and over-fitting of sample data.

Many image processing and computer vision applications seek to maximize vision quality and/or minimize perceptual distortion in images, so it is important to develop more accurate computed image quality assessment (IQA) methods to estimate the vision quality of distorted images[18-20]. Some objective distortion measurement (e.g. MSE: Mean Square Error) are easy to calculate and have a clear physical meaning, but they are not greatly correlated with the subjective vision quality of the image. To overcome this problem, a structural similarity index (SSIM) was developed, which is an objective full reference image quality assessment index. SSIM[21] is one of the most popular IQA because of its high predictability and wide applicability to image quality optimization problems. SSIM is the product of three integrated measurement: luminance (mean), contrast (standard deviation), and structure (correlation). Therefore it has good performance in optical images[22][23]. Although SSIM is better, it can not effectively evaluate the image quality of severe distortion, which may be inconsistent with subjective assessment result. To overcome the problem, many researchers have proposed an improved SSIM method. Since humane eye is highly sensitive to image edges, image gradient and edge information can reflect change in image texture features, so Yang established gradient-based structural similarity index (GSSIM)[24] and edge-based structural similarity index (ESSIM)[25][26] that is an SSIM-based optical image improvement method to introduce image edges into SSIM[27] proposed regression SSIM to measure the similarity of images of two similar modes, it is more effective than SSIM in suppressing noise and outliers in data. Marziliaono calculates the continuous width of the image to measure the degree of image blur[28]. Grete uses different gradient changes in different resolution images to evaluate the ambiguity of color images after Gaussian blur[29]. Yuan proposed gradient information based on luminance components to evaluate the sharpness of image fusion[17].

In this paper, the formal context analysis theory and concept lattice incremental construction algorithm are analyzed. Combined with SSIM, an incremental concept lattice construction algorithm based on SSIM (SsimAddExtent) is proposed. The main ideas are as follows: First, we consider the essence of the concept lattice construction is the knowledge division process. That is, each node on the Hasse diagram is a class with common attributes. Secondly, we consider when a class is mapped to a graphics, the center point of the class is strongly attractive to other points in the class, and other points will be as close as possible to the center

point, away from the class boundary. Therefore, the mapping graphics of class will have a stable boundary, even if some elements are added or remove from the class, the boundary of the mapping graphic will be unchanged or only slightly changed. That is to say, the mapping graphics of the classes before and after changing is highly similar. Based on the above considerations, we first propose an image structural similarity index based on Elliptic Fourier Descriptor (EFD-SSIM). Secondly, an improved concept lattice incremental construction algorithm based on EFE-SSIM is proposed (SsimAddExtent). The degree of influence of each newly added object on the boundary contour change of the mapping graphics of the original node on the Hasse diagram is used as the basis for judging the position of the newly added object. Theoretical analysis and experimental comparison demonstrate that SsimAddExtent is superior to other algorithms and construction efficiency is higher.

The structure of this paper is as follows. In the second section, we introduce some basic definitions and propositions of FCA. The third section describe our algorithm. The fourth section evaluate the performance of the algorithm. Our work is summarized in the fifth section.

2 Preliminaries

In this section, we will introduce the basic notions and conventions about FCA and SSIM. All definitions and propositions are assumed they can be found in[30][31] which the reader is kindly referred to for more detailed description.

2.1 Formal concept analysis (FCA)

Definition1. A formal context is triple of sets $K=(U, M, I)$, where I is a binary relationship between u and m . The element in u and m are called objects and attributes, respectively. $(u, m) \in I$ or uIm indicates the object u has attribute m .

A formal context can be represented by a cross-table or matrix where every row is an object and every column are an attribute. Cross in table represent incidence relation I . Table 1 show an example of the formal context.

Table 1 an example of the formal context

object	a	b	c	d	e	f	g	h	i
1	x				x		x		x
2	x	x	x	x					
3	x			x		x			
4	x			x	x				
5	x				x	x	x	x	
6	x	x	x				x		
7	x		x		x				
8	x				x	x	x	x	x

Definition2. $K=(U, M, I)$ is an formal context. $A \subseteq U, B \subseteq M$, Let $f(A) = \{m \in M \mid \forall u \in A, (u, m) \in I\}$, $g(B) = \{u \in U \mid \forall m \in B, (u, m) \in I\}$, if $f(A) = B, g(B) = A$, (A, B) is called a formal concept, A and B are called the extent and intent of the concept (A, B) , respectively. $B = (U, M, I)$ or $B = (K)$ represents the sets of all concepts on the formal context.

Definition3. Let (A_1, B_1) and (A_2, B_2) are two formal concepts of a given formal context $K=(U, M, I)$. If $A_1 \subseteq A_2$ (or equivalently $B_2 \subseteq B_1$), (A_1, B_1) is called an sub-concept of (A_2, B_2) which can be denoted by $(A_1, B_1) < (A_2, B_2)$. If $A_1 \subseteq A_2$ and there is no other concept (A_3, B_3) which satisfies $A_1 \subset A_3 \subset A_2$. The relationship is denoted by $(A_1, B_1) p (A_2, B_2)$ and is called

partial order relation, we denote (A_1, B_1) as the child of (A_2, B_2) and (A_2, B_2) as the parent node of (A_1, B_1) . Using $\underline{B}=(U, M, I)$ ($B=(U, M, I), \leq$) represent a semi-order sets formed by $B=(U, M, I)$ with partial order relation.

Definition4. If any subset of $B=(U, M, I)$ has supremum and infimum, $\underline{B}=(U, M, I)$ is a complete set, called concept lattice.

Like any partial order set, the concept lattice can be represented by a line graph or an Hasse diagram. In the Hasse diagram, only adjacent nodes can be connected by edges. If $(A_1, B_1) \leq (A_2, B_2)$, then the position of (A_1, B_1) should be lower than (A_2, B_2) in Hasse diagram. The Hasse diagram of Table 1 (Fig 1).

Fig 1 the concept lattice of Table 1

2.2 Structural similarity index (SSIM)

SSIM is proposed as an image quality assessment algorithm, and its correction with human perceived quality is higher than Mean Square Error (MSE)[21]. SSIM includes the luminance information, contrast information and structural information in original image and distortion image in, they are as follows:

$$\begin{aligned}
 SSIM(x, y) &= L(x, y) \cdot C(x, y) \cdot S(x, y) \\
 L(x, y) &= \frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\
 C(x, y) &= \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \\
 S(x, y) &= \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 c_1 &= (k_1 \cdot l)^2 \\
 c_2 &= (k_2 \cdot l)^2 \\
 c_3 &= c_2 / 2
 \end{aligned} \tag{2}$$

Wherein, μ_x and μ_y are the average luminance value of original image and distortion image; σ_x and σ_y are standard deviation of original image and distortion image; σ_{xy} are correlation coefficient of two images. Small fixed positive number k_1, k_2, l are constants in order to avoid the denominator is zero, $k_1 = 0.01, k_2 = 0.03, l = 255$ (the dynamic range of pixel values, it is generally taken as 255).

3 An incremental algorithm for concept lattice based on SSIM

In this section, we will expatiate Elliptic Fourier Descriptor, and establish an improved SSIM algorithm based on Elliptic Fourier Descriptor (EFD-SSIM). After carefully studying FCA theory and incremental algorithm, we realize the construction process of concept lattice is actually the knowledge classification, each node on the concept lattice is a set where element with common attributes. When a new object is added to a concept lattice, the EFD-SSIM will be used to judge which node has the closest distance from the new object on the upper layer of the concept lattice, thus finding the first position of the new object on the concept lattice. After the first location is done, the final position of the new object and connection relationship with other nodes will be determined along the top-to-bottom direction. Therefore, we establish an incremental algorithm based on EFE-SSIM. The algorithm is based on the graphical representation of knowledge classification, and use EFD-SSIM to calculate the nearest node of the new object on the concept lattice, thus realizing location of the new object.

3.1 EFD-SSIM

Fourier Descriptor (FD) is a classic contour-based shape representation that originally proposed by Cosgriff (1960). The main idea is to use a set of data representing the overall frequency of the shape to describe the contour features, and to have invariance to operations such as rotation and translation. It is hot of the shape representation research. In terms of algorithm research, the many researchers have done a lot of work to improve shape representation algorithm based on Fourier operator in order to enhance the ability of shape representation. D Zhang, G Lu proposed an enhanced universal Fourier Descriptor to extract the key content of graph, which resolves the shortcoming that the most of descriptors are not suitable for generic shape representation[32]. SS Li ,YD Huang ,JW Yang proposed a region-based affine invariant ring Fourier Descriptor for affine invariant feature extraction , which can be used to extract contour feature of object with multiple component[33]. R Kasaudhan, SH Son proposed an enhanced version of the grid distance Fourier Descriptor to calculate image similarity and improve the image matching ratio. B Belkhaoui, AToumi combined Fourier Descriptor with watershed algorithm (WS), a process and method for auto target recognition based on inverse synthetic aperture radar image to solve the target recognition of the radar image[34]. The principle and related work of Fourier Descriptor will be described in detail below.

The first, we define a continuous curve $s(t)$ (see Fig 2). Let (x_0, y_0) as starting point of the target boundary. After moving at a certain speed in the counterclockwise direction, the target boundary can be described by the coordinates of the boundary points.

Fig 2 a continuous curve

The boundary curve is defined as:

$$s(t) = x(t) + jy(t) \quad t = 0, 1, L, N-1 \quad (3)$$

Where, t is the unit arc-length that moves along the boundary curve. To describe the contour of the image, the selected starting point must move one circle along the boundary. So $s(t)$ is a periodic function, $T = 2\pi$. For obtaining Elliptic Fourier Descriptors (EFD) of boundary, we need to obtain Fourier expansion of boundary.

$$\begin{aligned}
 s(t) &= \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kwt) + b_k \sin(kwt)) \\
 &= \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[\frac{a_k}{2} (e^{jkw t} + e^{-jkw t}) - j \frac{b_k}{2} (e^{jkw t} - e^{-jkw t}) \right] \\
 &= \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[\frac{a_k - jb_k}{2} e^{jkw t} + \frac{a_k + jb_k}{2} e^{-jkw t} \right] \\
 &= s_0 + \sum_{k=1}^{\infty} [s_k e^{jkw t} + s_{-k} e^{-jkw t}] \\
 &= \sum_{k=-\infty}^{\infty} s'_k e^{jkw t}
 \end{aligned} \tag{4}$$

Let $s_k = s_{k1} - js_{k2}$ $s_{-k} = s_{k1} + js_{k2}$

Equation(4) can be represented as:

$$\begin{aligned}
 s(t) &= s_0 + 2 \sum_{k=1}^{\infty} \left[s_{k1} \frac{e^{jkw t} + e^{-jkw t}}{2} + js_{k2} \frac{-e^{jkw t} + e^{-jkw t}}{2} \right] \\
 &= s_0 + 2 \sum_{k=1}^{\infty} [s_{k1} \cos(kwt) + s_{k2} \sin(kwt)] \\
 &= s_0 + \sum_{k=1}^{\infty} [a_k \cos(kwt) + b_k \sin(kwt)]
 \end{aligned} \tag{5}$$

Where, a_k, b_k is called Fourier Descriptor. According to the orthogonal properties of the trigonometric functions in Equation(5), we can see that:

$$a_k = \frac{2}{T} \int_0^T s(t) \cos(kwt) dt, b_k = \frac{2}{T} \int_0^T s(t) \sin(kwt) dt \tag{6}$$

According Equation(4):

$$\begin{aligned}
 s_k &= \frac{1}{T} \int_0^T s(t) e^{-jkw t} dt \\
 &= \frac{1}{T} \int_0^T x(t) e^{-jkw t} dt + \frac{1}{T} \int_0^T y(t) e^{-jkw t} dt \\
 &= s_{xk} + js_{yk}
 \end{aligned} \tag{7}$$

According to Equation(4) and Equation(5):

$$\begin{aligned}
 s_{xk} &= \frac{a_{xk} - jb_{xk}}{2} & s_{yk} &= \frac{a_{yk} - jb_{yk}}{2} \\
 s_{s-k} &= \frac{a_{xk} + jb_{xk}}{2} & x_{y-k} &= \frac{a_{yk} + jb_{yk}}{2}
 \end{aligned} \tag{8}$$

Then, according Equation(6):

$$\begin{aligned}
 a_{xk} &= \frac{2}{T} \int_0^T x(t) \cos(kwt) dt & b_{xk} &= \frac{2}{T} \int_0^T x(t) \sin(kwt) dt \\
 a_{yk} &= \frac{2}{T} \int_0^T y(t) \cos(kwt) dt & b_{yk} &= \frac{2}{T} \int_0^T y(t) \sin(kwt) dt
 \end{aligned} \tag{9}$$

Because curve $s(t)$ is non-continuous, we use the most direct Riemann summation method to approximate the integral value to obtain the discrete approximation of Equation(9):

$$\begin{aligned} a_{xk} &= \frac{2}{m} \sum_{i=1}^m x_i \cos(kwi\tau) & b_{xk} &= \frac{2}{m} \sum_{i=1}^m x_i \sin(kwi\tau) \\ a_{yk} &= \frac{2}{m} \sum_{i=1}^m y_i \cos(kwi\tau) & b_{yk} &= \frac{2}{m} \sum_{i=1}^m y_i \sin(kwi\tau) \end{aligned} \quad (10)$$

Where $a_{xk}, a_{yk}, b_{xk}, b_{yk}$ denote as an ellipse; m is the number of sampling points along the boundary, it is a half of the number of sampling points in general; $\tau = T/m$ is sampling period; x_i and y_i are value at the i -th sample point.

According Equation(7) and Equation(8), s_k can be seen as the sum of the complex pairs

$$s_k = A_k - jB_k \quad s_{-k} = A_k + jB_k \quad (11)$$

We assume that

$$A_k = \frac{a_{xk} + ja_{yk}}{2} \quad B_k = \frac{b_{xk} + jb_{yk}}{2} \quad (12)$$

Equation(4) can be expressed by:

$$s(t) = s_0 + \sum_{k=1}^{\infty} (A_k - jB_k) e^{jkwt} + \sum_{k=-\infty}^{-1} (A_k + jB_k) e^{jkwt} \quad (13)$$

We assume that $s(t)$ shift Δx and Δy along XY -coordinate, respectively. Rotating angle is φ in anti-clockwise direction. Then, we can obtain a new curve $\hat{s}(t) = \hat{x}(t) + j\hat{y}(t)$. The relationship between $s(t)$ and $\hat{s}(t)$ is obtained by Equation(4).

$$\begin{pmatrix} \hat{x}(t) \\ \hat{y}(t) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} a_{x0} + 2\Delta x \\ a_{y0} + 2\Delta y \end{pmatrix} + s \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix} \sum_{k=1}^{\infty} \begin{pmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{pmatrix} \begin{pmatrix} \cos(kwt) \\ \sin(kwt) \end{pmatrix} \quad (14)$$

The coefficients in Equation(14) are computed by:

$$\begin{aligned} \hat{a}_{xk} &= s(a_{xk} \cos(\varphi) + a_{yk} \sin(\varphi)) \\ \hat{b}_{xk} &= s(b_{xk} \cos(\varphi) + b_{yk} \sin(\varphi)) \\ \hat{a}_{yk} &= s(-a_{xk} \sin(\varphi) + a_{yk} \cos(\varphi)) \\ \hat{b}_{yk} &= s(-b_{xk} \sin(\varphi) + b_{yk} \cos(\varphi)) \\ \hat{a}_{x0} &= a_{x0} + 2\Delta x \\ \hat{a}_{y0} &= a_{y0} + 2\Delta y \end{aligned} \quad (15)$$

The following conclusion can be proved;

$$\frac{|\hat{A}_k|}{|\hat{A}_1|} + \frac{|\hat{B}_k|}{|\hat{B}_1|} = \frac{|A_k|}{|A_1|} + \frac{|B_k|}{|B_1|} = \frac{\sqrt{a_{xk}^2 + a_{yk}^2}}{\sqrt{a_{x1}^2 + a_{y1}^2}} + \frac{\sqrt{b_{xk}^2 + b_{yk}^2}}{\sqrt{b_{x1}^2 + b_{y1}^2}} \quad (16)$$

Thus, EFD with translation rotation and scale transformation invariant can be expressed as by:

$$EFD = \frac{|A_k|}{|A_1|} + \frac{|B_k|}{|B_1|} \quad (17)$$

In order to measure the degree of changing of the boundary contour of image, the overall structural information of visible region can no longer be extracted, and only the structural information of boundary is utilized, which is represented by EFD. So, the Equation(1) is modified, which include luminance information, distortion information and edge structure information. EFD-SSIM is defined by Equation(18) and(19). EFD-SSIM is closer 1 means that the boundary contour of two images have highly similarity.

$$EFD-SSIM(x, y) = L(x, y) \cdot C(x, y) \cdot ES(x, y) \in [0, 1] \quad (18)$$

$$ES = \frac{|\text{cov}(EFD1, EFD2)|}{\sigma_{EFD1} \cdot \sigma_{EFD2}} \in [0, 1] \quad (19)$$

Wherein, ES is the edge structural similarity of original image and distortion image; $\text{cov}()$ is covariance of edge descriptor of two images; σ_{EFD_i} is standard deviation of edge descriptor of two images.

3.2 SsimAddExtent

Definition5. Let L, L' be the concept lattice before and after inserting the new object $m, g(m')$ is attributes set of m . (A, B) is a concept in L' . Then.,

- (1). (A, B) is a new concept if B is not an intent of any concept in L
- (2). (A, B) is a modified concept if $B \subseteq g(m')$ and B is an intent of one concept in L , (A, B) is updated to $(A \cup m, B)$
- (3). (A, B) is an old concept if it is unchanged from L to L'
- (4). Assuming (X, Y) is a new concept and (A, B) is an old concept, if they satisfy $B \cap g(m') = X \neq B$, $B \cap g(m') \neq \emptyset$, the concept (A, B) is the generator of the concept (X, Y) . Otherwise, it is a general old concept.

According to the above definition, we can find that searching modified concept and new concept are very important to increment algorithm. Modified concept is easy to identify because its attributes set is subset of $g(m')$. We only need to insert new object m into their scope. On the other hand, new concepts need more time in the process because they do not exist until new attributes are inserted. To decide which a concept need to be created, incremental algorithm usually make use of the relationship between generator and new concept, which is defined by definition5. Each new concept (X, Y) has at least one generator, but there is only one canonical generator in all generators. Because the properties, we can create a new concept by recognizing the corresponding canonical generator.

The SsimAddExtent algorithm proposed in this paper construct concept lattice recursively from the maximum upper bound of the concept lattice, just like the original AddExtent algorithm[35]. Their difference is SsimAddExtent uses EFD-SSIM function to find the maximum general concept called MaximalConcept in the recursive process. The algorithm will be summarized as follows:

Assuming L is a concept lattice of formal context $K=(U, M, I)$, Extent is object sets of any node in L , Intent is attributes set of any node in L , B_1 is the attributes corresponding new object S . To insert S into L , the MaximalConcept of S in L should be decided at first, S will be put in the mapping graphic of the direct child node of the root node, respectively, and uses EFD-SSIM algorithm to decide where S should be inserted into. If $0 < \text{EFD-SSIM} < 0.5$, S will be not added as new object to the corresponding node. If $0.5 < \text{EFD-SSIM} < 1$, the corresponding node in L will be updated to $(\text{Extent} \cup S, \text{Intent})$, and as the maximum upper bound of new concept and modified concept, namely MaximalConcept. Thereafter, the attributes set of S is sequentially intersected with the direct child node of MaximalConcept. If the intersection of two sets is empty, the direct child is called an invariant node. If the attributes set of the direct child node is a subset of B_1 , the node will be updated to $(\text{Extent} \cup S, \text{Intent})$. If the intersection of two sets is not empty and it different from attributes set of any child node of MaximalConcept, then $(\text{Extent} \cup S, \text{Intent} \cap B_1)$ is called new node. Follow the above steps to recursively search until all child nodes of MaximalConcept are traversed.

SsimAddExtent is as follows:

$$C_0 = (\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \emptyset)$$

$$C_1 = (\{13, 14, 5\}, l)$$

$$C_2 = (\{10, 3, 4, 5\}, j)$$

$$C_3 = (\{2, 3, 4, 5, 8, 9\}, d)$$

$$C_4 = (\{1, 2\}, b)$$

$$C_5 = (\{3, 6, 8\}, f)$$

$$C_6 = (\{1, 10, 11, 12, 4, 5, 6, 7, 9\}, a)$$

$$C_7 = (\{11, 6\}, \{a, o\})$$

$$C_8 = (\{7\}, \{a, p, q\})$$

$$C_9 = (\{11, 12\}, \{a, r\})$$

$$C_{10} = (\{6, 8\}, \{f, n\})$$

$$C_{11} = (\{1\}, \{a, b, c\})$$

$$C_{12} = (\{10, 4, 5\}, \{a, j\})$$

$$C_{13} = (\{4, 5, 9\}, \{a, d\})$$

$$C_{14} = (\{3, 8\}, \{d, f\})$$

$$C_{15} = (\{2\}, \{b, d\})$$

$$C_{16} = (\{3, 4, 5\}, \{d, j\})$$

$$C_{17} = (\{11\}, \{a, o, r, s\})$$

$$C_{18} = (\{6\}, \{a, f, m, n, o\})$$

$$C_{19} = (\{8\}, \{d, f, n\})$$

$$C_{20} = (\{4, 5\}, \{a, d, j\})$$

$$C_{21} = (\{3, 5\}, \{d, h, j\})$$

$$C_{22} = (\{4\}, \{a, d, l, j\})$$

$$C_{23} = (\{5\}, \{a, d, h, j, k, l\})$$

$$C_{24} = (\{3\}, \{d, f, g, h, j\})$$

$$C_{25} = (\emptyset, \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r\})$$

In the process of adding object 15 whose attributes set is $\{a, d, f, r\}$. Firstly, the spatial distribution of $C_1 \sim C_6$ before and after adding object 15 is mapped into graphic, in order to realize graphical representation of high dimensional data, the dimension reduction process is necessary. After the dimensionality reduction using the t -SNE method in this paper, the graph mapping is performed using Delaunay Triangulation algorithm (see Fig 4)

Fig 4 Mapping graph of lay1 before and after adding object 15

Making use of canny edge detection operator, the edge contour of Fig 4 can be achieved (see Fig 5)

Fig 5 The edge contour in Fig 4

Using Equation(17)~(19) to calculate the EFD-SSIM value between A and B in Fig 5. Table 4 shows EFD-SSIM value of $C_1 \sim C_6$ before and after adding object 15.

Table 4 EFD-SSIM value between A and B in Fig 5

node	C1	C2	C3	C4	C5	C6
	0.0964	0.2171	0.8353	0.4479	0.5929	0.6015

Since the value of C_3, C_5, C_6 are greater than 0.5. Therefore, they are modified nodes, their expressions are as follow:

$$\begin{aligned}
 C_3 &= (\{2, 3, 4, 5, 8, 9, 15\}, d) \\
 C_5 &= (\{3, 6, 8, 15\}, f) \\
 C_6 &= (\{1, 10, 11, 12, 4, 5, 6, 7, 9, 15\}, a)
 \end{aligned}$$

We can see from Fig 3 that the direct child node of C_3 are $C_{13} \sim C_{16}$, the direct child node of C_5 are C_{10}, C_{14} and the direct child node of C_6 are $C_7 \sim C_{13}$. Making attributes set of object 15 intersect with $C_{13} \sim C_{16}$, respectively.

$$\begin{aligned}
 Intent(C_{13}) \cap Intent(15) &= \{a, d\} \subset \{a, d, f, r\} \\
 Intent(C_{14}) \cap Intent(15) &= \{d, f\} \subset \{a, d, f, r\}
 \end{aligned}$$

According to definition 5 C_{13}, C_{14} are modified concept and C_{15}, C_{16} are general old concept. Similarly, C_{14} which is the direct child node of C_5 is modified concept, C_{10} is general old concept; C_9, C_{13} which are the direct child nodes of C_6 are modified concept, the others are general old concept. Modified concepts are as follows:

$$\begin{aligned}
 C_9 &= (\{11, 12, 15\}, \{a, r\}) \\
 C_{13} &= (\{4, 5, 9, 15\}, \{a, d\}) \\
 C_{14} &= (\{3, 8, 15\}, \{d, f\})
 \end{aligned}$$

The direct child nodes of C_7 are C_{17}, C_{18} ; the direct child node of C_9 is C_{17} ; the direct child nodes of C_{10} are C_{18}, C_{19} ; the direct child node of C_{11} is C_{19} ; the direct child node of C_{12}, C_{13} is C_{20} ; the direct child nodes of C_{14} are C_{19}, C_{24} . Let attributes sets of these nodes intersect with $\{a, d, f, r\}$. According to definition 5, we can

$$Intent(C_{18}) \cap Intent(15) = \{a, f\} \neq Intent(C_7) \cap Intent(15) \neq Intent(C_{10}) \cap Intent(15)$$

Confirm that $\{(6,16), (a, f)\}$ is new concept which is generalized by C_{18} , the others are general old concept. Determining the father-son relationship between new concepts and modified concepts to update to the edges in Hasse diagram (see Fig 6).

Fig 6 The Hasse diagram after adding object 15

4 Experimental Evaluation and Analysis

In order to prove the efficiency of the algorithm proposed in this paper, we use R language to implement it and original AddExtent algorithm. SsimAddExtent algorithm has no superiority in time complexity, it is $O(|L| \cdot |G| \cdot |M|^3)$ [36]. However, since EFD-SSIM is introduced and only the child nodes of modified concepts or new concepts are traversed, some unnecessary comparison is avoided. Therefore, the position of new node can be decided quickly on the large-scale concept lattice, the runtime is greatly reduced.

The data set used throughout the experiments was randomly generated with different integrity ratio and they are 10%, 30%, and 50%. The data set have 80 objects, the number of attributes is variant, and each attribute may have different number of objects.

Fig 7 Results for random datasets with 10% integrity ratio

We compared the runtime between the SsimAddExtent algorithm and the AddExtent algorithm on a dataset with 10% integrity ratio (see Fig 7). The number of attributes is

increased gradually from 200 to 35000, the number of objects is 80. When the number of attributes is smaller, the SsimAddExtent algorithm differs from the AddExtent algorithm slightly. When the number of attributes is 2700, the runtime of SsimAddExtent algorithm to be less than the AddExtent algorithm, and following the, and as the number of attributes increases, the gap of two algorithm becomes larger.

Fig 8 shows that the difference the SsimAddExtent algorithm and AddExtent algorithm on dataset with 30 integrity ratio. The number of attributes is increased from 30 to 12000, the number of objects is invariant. Under the same conditions, the runtime gap is bigger compared to Fig 7. The SsimAddExtent algorithm has superiority whether the number of attributes is large or small.

Fig 8 Results for random datasets with 30% integrity ratio

We compared the runtime between the SsimAddExtent algorithm and the AddExtent algorithm on a dataset with 50% integrity ratio (see Fig 9). The number of attributes is increased from 5 to 1200. Since the size of concepts is greatly large and the consumption of memory resource is very fast, we can only test on 1200 attributes. Figure shows that the runtime is rising quickly. The crossing point of two algorithms appears much earlier compared to Fig 7 and Fig 8. Meanwhile, the performance of the SsimAddExtent algorithm superiority to the AddExtent algorithm.

Fig 9 Results for random datasets with 50% integrity ratio

5 Conclusions

The incremental algorithm by adding objects to concept lattice , which not only can construct concept lattice, but also update concept lattice. In this paper, we introduce an fast and efficient

incremental algorithm construct concept lattice. SsimAddExtent algorithm improves AddExtent algorithm, it uses EFD-SSIM to decide the MaximalConcept, and the attributes of new object only compares with the child nodes of MaximalConcept to decide modified concepts and new concepts. The algorithm saves runtime since it reduces the number of iterations.

SsimAddExtent algorithm has obvious advantage over AddExtent algorithm at almost every test dataset even when the number of attributes is small and the dataset has low integrity ratio. Meanwhile, the performance gap between two algorithms will increase as the number of attributes increases. Both theoretical analysis and performance test show SsimAddExtent algorithm is better choice than AddExtent algorithm when we apply the FCA method with large scale data or high data integrity ratio.

Acknowledgements

We thank all participants for their invaluable contribution. We thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

Funding

This study was funded by Beijing postdoctoral science foundation (2021-ZZ-063).

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] Poelmans, Jonas, Paul Elzinga, Stijn Viaene, and Guido Dedene. "Formal Concept Analysis in Knowledge Discovery: A Survey." Paper presented at the International Conference on Conceptual Structures: from Information to Intelligence, 2010.
- [2] Wermelinger, Michel, Yijun Yu, and Markus Strohmaier. "Using Formal Concept Analysis to Construct and Visualise Hierarchies of Socio-Technical Relations." Paper presented at the International Conference on Software Engineering-companion Volume, 2009.
- [3] Priss, Uta. "Linguistic Applications of Formal Concept Analysis." Paper presented at the Formal Concept Analysis, Foundations & Applications, 2005.
- [4] Kang, Xiangping, Deyu Li, and Suge Wang. "Research on Domain Ontology in Different Granulations Based on Concept Lattice." *Knowledge-Based Systems* 27, no. 3 (2012): 152-161.
- [5] Amin, Islam Ibrahim, Samar K. Kassim, Aboul Ella Hassanien, and Hesham A. Hefny. "Using Formal Concept Analysis for Mining Hyomethylated Genes among Breast Cancer Tumors Subtypes." Paper presented at the International Conference on Advances in Computing, 2013.
- [6] Besson, Jrmy, Cline Robardet, Jean Franois Boulicaut, and Sophie Rome. "Constraint-Based Concept Mining and Its Application to Microarray Data Analysis." *Intelligent Data Analysis* 9, no. 1 (2005): 59-82.
- [7] Dau, Frithjof, Jon Ducrou, and Peter Eklund. "Concept Similarity and Related Categories in Searchleuth." Paper presented at the International Conference on Conceptual Structures: Knowledge Visualization & Reasoning, 2008.
- [8] Formica, Anna. "Semantic Web Search Based on Rough Sets and Fuzzy Formal Concept Analysis." *Knowledge-Based Systems* 26, no. 9 (2012): 40-47.

-
- [9] Poelmans, Jonas, Dmitry I. Ignatov, Sergei O. Kuznetsov, and Guido Dedene. "Formal Concept Analysis in Knowledge Processing: A Survey on Applications." *Expert Systems with Applications An International Journal* 40, no. 16 (2013): 6538-6560.
- [10] Zhang, Kai, H. U. Yun Fa, and Yu Wang. "An Ist-Based Algorithm for Construction of Concept Lattices." *Journal of Computer Research & Development* (2004).
- [11] Hong, Q. I., Da You Liu, H. U. Cheng-Quan, L. U. Ming, and Liang Zhao. "An Algorithm for Generating Concepts Based on Search Space Partition." *Journal of Software* 16, no. 12 (2005): 2029-2035.
- [12] Godin, Robert, Rokia Missaoui, and Hassan Alaoui. "Incremental Concept Formation Algorithms Based on Galois (Concept) Lattices." *Computational Intelligence* 11, no. 2 (2010): 246-267.
- [13] Xie, Zhi Peng, and Zong Tian Liu. "A Fast Incremental Algorithm for Building Concept Lattice." *Chinese Journal of Computers*, no. 5 (2002): 490-496.
- [14] Kourie, Derrick G., Sergei Obiedkov, Bruce W. Watson, and Dean Van Der Merwe. "An Incremental Algorithm to Construct a Lattice of Set Intersections." *Science of Computer Programming* 74, no. 3 (2009): 128-142.
- [15] Li Yun, Liu Zong Tian, Chen Jun, Shen Xia Jiong, and Xu Xiao Hua. 'Attribute-based concept lattice progressive generation algorithm.' *Small microcomputer system* 25, no. 10 (2004): 1768-1771.
- [16] 'Horizontal merge algorithm for multiple concept lattices.' *Electronic Journal* 32, no. 11 (2004): 1849-1854.
- [17] Yuan, Y. U., Xu Qian, Feng Zhong, and L. I. Xiao-Rui. "Increment Construction Algorithm for Concept Lattice Based on Maximal Concept." *Computer Engineering* 35, no. 21 (2009): 62-64.
- [18] Lin, Zhang, Zhang Lei, Xuanqin Mou, and David Zhang. "A Comprehensive Evaluation of Full Reference Image Quality Assessment Algorithms." Paper presented at the IEEE International Conference on Image Processing, 2013.
- [19] Bae, Sung Ho, Jaeil Kim, Munchurl Kim, Sukhee Cho, and Soo Choi Jin. "Assessments of Subjective Video Quality on Hvc-Encoded 4k-Uhd Video for Beyond-Hdtv Broadcasting Services." *IEEE Transactions on Broadcasting* 59, no. 2 (2013): 209-222.
- [20] Choi, Jae Seok, Sung Ho Bae, and Munchurl Kim. "Single Image Super-Resolution Based on Self-Examples Using Context-Dependent Subpatches." Paper presented at the IEEE International Conference on Image Processing, 2015.
- [21] Zhou, Wang, Bovik Alan Conrad, Sheikh Hamid Rahim, and Eero P Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Trans Image Process* 13, no. 4 (2004): 600-612.
- [22] Wang, Zhou, A. C. Bovik, and Ligang Lu. "Why Is Image Quality Assessment So Difficult?" Paper presented at the IEEE International Conference on Acoustics, 2011.
- [23] Narwaria, Manish, and Weisi Lin. "Scalable Image Quality Assessment Based on Structural Vectors." Paper presented at the IEEE International Workshop on Multimedia Signal Processing, 2009.
- [24] Yang, Chun Ling. "Gradient-Based Structural Similarity for Image Quality Assessment." *Journal of South China University of Technology* 2, no. 9 (2006): II-II.
- [25] Chen, Guan Hao, Chun Ling Yang, Lai Man Po, and Sheng Li Xie. "Edge-Based Structural Similarity for Image Quality Assessment." Paper presented at the IEEE International Conference on Image Processing, 2007.
- [26] Peng, L. V., and Jian Qiu Zhang. "Edge-Weighted Structural Similarity Index." *Computer Engineering* 37, no. 14 (2011): 226-219.
- [27] Yi, Ke Wang, Lianlin Li, Yang Zhou Xiao, and Tie Jun Cui. "Supervised Automatic Detection of Uwb Ground-Penetrating Radar Targets Using the Regression Ssim Measure." *IEEE Geoscience & Remote Sensing Letters* 13, no. 5 (2017): 621-625.
- [28] Mi, Zengzhen. "Image Quality Assessment in Multiband Dct Domain Based on Ssim." *Optik - International Journal for Light and Electron Optics* 125, no. 21 (2014): 6470-6473.
- [29] Crété-Roffet, Frédérique, Thierry Dolmiere, Patricia Ladret, and Marina Nicolas. "The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric." *Proceedings of SPIE - The International Society for Optical*

Engineering 12 (2007): 64920I-20I-211.

- [30] Wille, Rudolf. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. 2009.
- [31] Zou, Caifeng, Daqiang Zhang, Jiafu Wan, Mohammad Mehedi Hassan, and Jaime Lloret. "Using Concept Lattice for Personalized Recommendation System Design." *IEEE Systems Journal* 11, no. 1 (2017): 305-314.
- [32] Zhang, Dengsheng, and Guojun Lu. "Enhanced Generic Fourier Descriptors for Object-Based Image Retrieval." Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011.
- [33] Li, San San, Yong Dong Huang, and Jian Wei Yang. "Affine Invariant Ring Fourier Descriptors." Paper presented at the International Conference on Wavelet Analysis and Pattern Recognition, 2013.
- [34] Belkhaoui, B., Abdelmalek Toumi, and A. Khalfallah. "Fusion Fourier Descriptors from the Em." *International Journal of Computer & Information Technology* (2014).
- [35] Merwe, Dean Van Der, Sergei Obiedkov, and Derrick Kourie. "Addintent: A New Incremental Algorithm for Constructing Concept Lattices." *Concept Lattices* 2961 (2004): 205-206.
- [36] Zou, Ligeng, Zuping Zhang, Jun Long, and Zhang Hao. "A Fast Incremental Algorithm for Deleting Objects from a Concept Lattice." *Knowledge-Based Systems* 89, no. C (2015): 411-419.