

Bounding Stability in Formal Concept Analysis



Bikram P. Bhuyan, Arindam Karmakar and Shyamanta M. Hazarika

Abstract In Formal Concept Analysis, stability is an important utility measure to rank concepts. However, computation of stability is considered to be a hard problem. Efficient algorithms having good bounds to estimate stability holds promise. In this paper, an effective graph-based technique is proposed to estimate stability. Our estimation algorithm has a polynomial time complexity of $O(|A|^2)$ where $|A|$ is the number of vertices.

Keywords Formal concept analysis · Algorithms · Graph algorithm · Estimation algorithm · Combinatorial problem

1 Introduction

Formal Concept Analysis (FCA) acquires grouping called formal concepts from Boolean data comprising objects and attributes [1]. A formal concept is a pair comprising a. extent—containing the objects which belong to the concept, and b. intent—include all the properties and meanings which apply to all objects in the extent [2]. The set of all formal concepts of a given context forms a complete lattice, called the concept lattice. However, such approaches using concept lattices often encounter the problem of the exponential growth rate of concepts in the lattice with the size of

B. P. Bhuyan (✉)

Asian Institute of Management and Technology, Guwahati, Assam, India
e-mail: bikram23bhuyan@gmail.com

A. Karmakar

Biomimetic and Cognitive Robotics Lab, Computer Science and Engineering,
Tezpur University, Tezpur, India
e-mail: arindam@tezu.ernet.in

S. M. Hazarika

Department of Mechanical Engineering, Indian Institute of Technology
Guwahati, Guwahati, India
e-mail: s.m.hazarika@iitg.ernet.in

© Springer Nature Singapore Pte Ltd. 2018

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Advances in Intelligent Systems and Computing 706,
https://doi.org/10.1007/978-981-10-8237-5_53

545

the context [3, 4]. A measure to rank concepts is often used; thereby only relevant concepts are retained.

In this regard, ‘stability’ was introduced as one such ranking measure [1]. However, the computation of stability is proven to be an NP-complete problem [5, 6]. One of the best known algorithms was proposed by Obiedkov et. al [7] whose complexity is itself quadratic to the size of the lattice. Therefore, efficient algorithms to estimate stability hold promise. Although there exist many algorithms to estimate stability in literature [5, 8, 9], there lacks an efficient method having good bounds to estimate stability index. In this paper, a new technique is proposed to estimate the stability of a formal concept exploiting results on bounds for counting cliques in spectral graphs.

2 Terminology and Main Definitions

2.1 Formal Concept Analysis

Formal Concept Analysis was first developed by Wille [10] and is implemented in many fields like psychology, sociology, anthropology, medicine, biology, linguistics, computer sciences, mathematics and industrial engineering [11]. Here we briefly present some basic definitions of FCA [10, 12]. The root of FCA is a formal context from which concepts are generated and are organized in a hierarchical order in the form of a lattice. Formal context is a triple (G, M, I) where G denotes the set of objects, M is the set of attributes and I is a binary relation between G and M , i.e. $I \subseteq G \times M$. $(g, m) \in I$ indicates that object g has the property (attribute) m . If $A \subseteq G$ and $B \subseteq M$, then the following derivation operators produce the Galois connection:

$$A' = \{m \in M | (g, m) \in I, \forall g \in A\} \quad (1)$$

(the set of attributes common to the object set in A)

$$B' = \{g \in G | (g, m) \in I, \forall m \in B\} \quad (2)$$

(the set of objects common to the attribute set in B)

A formal concept of the formal context (G, M, I) is a pair (A, B) where $A \subseteq G$ and $B \subseteq M$, satisfying $A' = B$ and $B' = A$. We also have a property $A'' = A$ and $B'' = B$ where $(.)''$ is the closure operator. Thus, a formal concept consists of a set A (extent of the concept) and a set B (intent of the concept) such that all the objects in A share a common attribute set B .

A partial order relation \leq , showing the subconcept–superconcept hierarchy is present in the set $\beta(G, M, I) = \{(A, B) | A' = B \text{ and } B' = A\}$. It is defined by $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$. A complete lattice is formed by $\beta(G, M, I)$ satisfying \leq .

2.2 Stability

Stability was first introduced in [5, 6], it was later revised in [7, 13] and is defined as follows:

Definition 1 Let $\Gamma = (G, M, I)$ be a formal context and (A, B) be a formal concept of Γ ; then the stability of the concept $\gamma(A, B)$ is defined as

$$\gamma(A, B) = \frac{|\{X \subseteq A \mid X' = B\}|}{2^{|A|}} \quad (3)$$

Stability measures how much a concept intent depends on objects (extent) of the concept. It is to be noticed that stability index is between 0 and 1.

As enumerating stability is an NP-complete problem, estimation to compute stability was done by Buzmakov et al. in their paper [8]. They obtained an estimated lower bound and upper bound on the value of the stability index and merged both together. This amalgamation was termed is ‘*bounding method*’. But the tightness of the bound cannot be ensured. Monte Carlo approximation was first discussed in [9] but it was found to be less efficient than the bounding method.

It is to be mentioned that the state of art methods exploit the ordering property of the lattice where the lattice is created at the outset, and then the stability of a superconcept is estimated with the help of the subconcepts. It would be better if only the stable concepts are projected at the onset, and only then, the lattice is generated with only those stable concepts.

Our approach to estimate stability involves the reduction of the stability function, i.e. the set $\{X \subseteq A \mid X' = B\}$ of a concept (A, B) to a graph problem to estimate the stability index of that concept without taking the subconcepts into consideration.

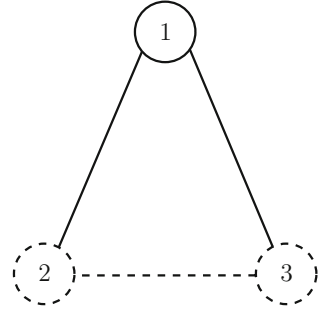
3 Our Approach to Estimate Stability

The stability index is mapped to a graph F . A reduced graph \bar{F} of F is taken by deleting some particular vertices and edges. More than one connected component could be a result of the action. Our goal is now to find the total number of cliques in each connected component and finally summing them to get the total number of cliques in \bar{F} .

Before proceeding, we would like to give symbols which we will be frequently using. For a given concept (A, B) in a context $\gamma = (G, M, I)$:

1. A is a set of objects $\{a_1, a_2, \dots, a_n\}$ and A_k is the subset of k objects of A .
2. $|A|$ denotes the cardinality of A .
3. 2^A denotes the power set of A .
4. B is the set of attributes/properties common to the set A .

Fig. 1 Mapping of the stability index of concept $(\{1,2,3\} \{a,b,g\})$ of the context ‘Living Beings and Water’ [12] to a graph is shown, where ‘solid circles’ represent V_{stb} , ‘dashed-circles’ represent \overline{V}_{stb} , ‘solid lines’ represent E_{stb} and ‘dashed-lines’ represent \overline{E}_{stb}



Lemma 1 Let $\Gamma = (G, M, I)$ be a formal context and (A, B) be a formal concept of Γ ; then the stability of the concept (A, B) ; can also be expressed as

$$\gamma(A, B) = 1 - \frac{|\{X \subseteq A | X' \neq B\}|}{2^{|A|}} \tag{4}$$

We now map our problem of computing the stability of a concept (A, B) to a graph problem.

Let $F = (V, E)$ be a graph having the ordered pair (V, E) as the set of vertices and edges, respectively. For any concept (A, B) , each object a_i of A correspond to a vertex V of the graph F . Therefore vertex set of F is $V = \{a_1, a_2, \dots, a_n\}$. All the vertices of F are mutually connected to form a complete graph.

We now have the following definitions:

Definition 2 A vertex $a_i \in V_{stb}$, iff $a'_i = B$. If $a'_i \neq B$ then $a_i \in \overline{V}_{stb}$ (refer Fig. 1).

Definition 3 An edge $(a_i, a_j) \in E_{stb}$, iff $\{a_i, a_j\}' = B, i \neq j$. Otherwise; $(a_i, a_j) \in \overline{E}_{stb}$ (refer Fig. 1).

The set of \overline{V}_{stb} vertices can further be classified into two sets- \overline{V}_{stb_e} and \overline{V}_{stb_n} . \overline{V}_{stb_e} are incident to at least one \overline{E}_{stb} edge. And \overline{V}_{stb_n} are the \overline{V}_{stb} vertices which are not adjacent to any \overline{E}_{stb} edge.

We state a proposition from Hui-lai Zhi et al. in [14].

Proposition 1 [14] Let $\Gamma = (G, M, I)$ be a context, and (A, B) be a concept of the Γ . For any $s \subseteq A$, if $s' = B$ then $g' = B$ for all $s \subseteq g \subseteq A$.

Observation 1 Edges between a V_{stb} vertex and any other vertex are always E_{stb} edge.

It should be noted that an E_{stb} edge may also exist between two \overline{V}_{stb} vertices.

The vertices and edges can also be termed as 1-cliques and 2-cliques, respectively.

A clique can be defined as a maximal complete subgraph of a given graph [15]. A clique of size k is called a k-clique. 1-cliques corresponds to vertices, 2-cliques to edges and 3-cliques to 3-cycles (triangles).

Observation 2 Let (A, B) be a concept of Γ and A_k is a subset of k objects of A . If there is at least one E_{stb} edge in the k -clique formed from A_k , then $A'_k = B$.

Observation 3 Let the graph F be reduced to \bar{F} by deleting all V_{stb} vertices and E_{stb} edges. For any subset A_k of the reduced graph, $A'_k \neq B$.

Observation 4 After deletion of V_{stb} vertices and E_{stb} edges; the stability index computed from the reduced graph $\bar{F} \subseteq F$ will be same as the stability index calculated from the complete graph.

Now, in the subgraph \bar{F} consisting of \bar{V}_{stb} and \bar{E}_{stb} as the set of vertices and edges, respectively, there could be more than one connected components. A *connected component* of an undirected graph is a subgraph, satisfying the criteria where edges connect any two vertices of the subgraph and the same is not connected to any other vertices belonging to the supergraph.

There is a linear time algorithm to compute the connected components in a graph [16]. And for each concept, we have to find the \bar{V}_{stb} vertices and \bar{E}_{stb} edges which will form one or more connected components graph and for each such component, we have to find the total number of cliques which will finally result in the total number of cliques in the whole graph. But finding the total number of cliques in a graph takes exponential time [15]. Bollobás and Nikiforov [17], Nikiforov [18] have provided an estimate of the largest eigenvalue of the adjacency matrix of a graph based on the number of cliques using the idea of number of walks.

We have the following result by Nikiforov [17]:

Result 1 [17] Let \bar{F} be a graph, where \hat{Z}_i denote the count of all i -cliques of \bar{F} and λ be the largest eigenvalue of the adjacency matrix of \bar{F} . Then for every graph \bar{F} and $r \geq 2$

$$\lambda^{r+1} \leq (r + 1)\hat{Z}_{r+1} + \sum_{s=2}^r (s - 1)\hat{Z}_s \lambda^{r+1-s} \tag{5}$$

Lemma 2 Given any formal context $\Gamma = (G, M, I)$ where (A, B) is a concept of the Γ , then the set $(\{X \subseteq A | X' \neq B\})$ of the concept (A, B) can always be reduced to a unique graph which will consist of one or many n -vertex, connected component graph.

Theorem 1 Given any formal context $\Gamma = (G, M, I)$ and (A, B) is a concept of the Γ ; the stability index $\gamma(A, B)$ is at most $1 - \frac{\{|\hat{Z}(\min)| + |\bar{V}_{stb_n}| + 1\}}{2^{|A|}}$ where $|\hat{Z}(\min)|$ is the minimum number of cliques of the graph \bar{F} .

Theorem 2 Given any formal context $\Gamma = (G, M, I)$ and (A, B) is a concept of the Γ ; the stability index $\gamma(A, B)$ is at least $1 - \frac{\{|\hat{Z}(\max)| + |\bar{V}_{stb_n}| + 1\}}{2^{|A|}}$ where $|\hat{Z}(\max)|$ is the maximum number of cliques of the graph \bar{F} .

Theorems 1 and 2 give new upper and lower bounds to the stability index. We now present the procedure formally in Algorithm 1.

Algorithm 1 Stability Index

 Compute stability indexes of all concepts of a given formal context $F = (G, M, I)$

Input: Formal Context

Output: Stability Indexes of all concepts

```

1: procedure CALCULATE-STABILITY
2:   for each concept  $\in F$  do
3:     Find the set of  $\bar{V}_{stb}$  ( $\bar{V}_{stb_e}$  and  $\bar{V}_{stb_n}$ ) and  $\bar{E}_{stb}$  at the generation time of each concept.
4:     Find the connected components (CC) of the graph taking  $\bar{V}_{stb_e}$  and  $\bar{E}_{stb}$  as the set of
       vertices and edges, respectively [16].
5:     for each connected component (CC) do
6:       Create adjacency matrix  $D$  of the vertices.
7:       Compute  $\lambda$ .
8:       Compute  $\hat{Z}(max_i)$  and  $\hat{Z}(min_i)$ 
9:     end for
10:    Compute  $\hat{Z}(max)$  and  $\hat{Z}(min)$ 
11:    Compute the upper bound and lower bound of stability index as in Theorems 1 and 2.
12:  end for
13: end procedure

```

4 Time Complexity and Performance Analysis

The asymptotic time complexity of creating a lattice of formal concepts using FCA algorithms like Fast Close by One (FCbO) is $O(|G|^2|M|L)$ [19], where $|G|$ is the number of objects, $|M|$ is the number of attributes and L is the number of concepts. The time complexity of finding connected components in a graph is $O(|V| + |E|)$ [16]. Also, the time complexity of finding the largest eigenvalue of an adjacency matrix is $O(|V|^2)$ where $|V|$ and $|E|$ are the number of vertices and edges in a (V, E)-graph.

Thus, our proposed method has a time complexity (finding the stability of all concepts in the lattice) of $O(|G|^2|M|L + L|A|^2)$ where $|A|$ is the number of vertices in the reduced graph F of a concept. And the time complexity of finding stability of a single concept is $O(|A|^2)$.

The bounds were tested (Figs. 2 and 3) against the ‘*bounding method*’ [8] discussed earlier. It was found that our algorithm results in stricter bounds in case of small concepts having high stability whereas the bounding method is not giving good estimation. With the increase in the number of concepts (large concepts), it is seen that both bounds are portraying relatively similar results. But in case of concepts having very low stability or as in our case after the graph formation, the eigenvalue is large ($\lambda = n - 1$) being a complete graph (K_n); our bounds do not give good results (because of the characteristic polynomial). So if ($\lambda = n - 1$), we do not compute the stability, as it will be a very low and of little use.

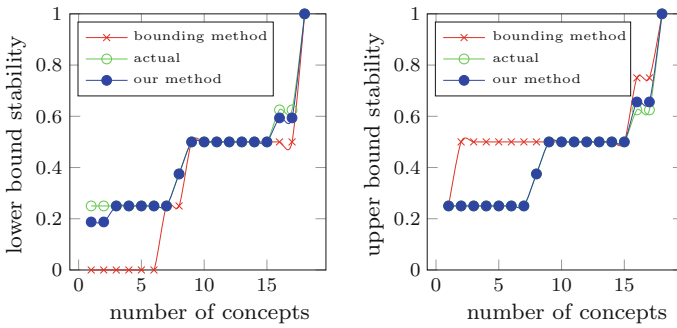


Fig. 2 Lower bound and upper bound comparison of the context ‘Living Beings and Water’ [12]

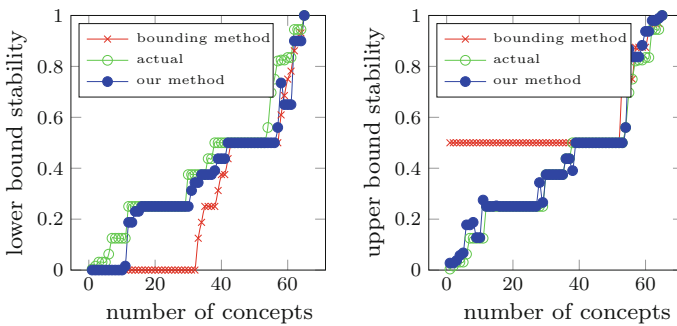


Fig. 3 Lower bound and upper bound comparison of the context ‘Tea Ladies’ [20]

5 Conclusion

Estimation of stability index for a formal concept is a hard problem. We have put forward an algorithm exploiting results from spectral graph theory to estimate stability. Its strength lies in the fact that it does not take the ordering property of the lattice into consideration and each concept is treated individually. Our algorithm has a polynomial time complexity of $O(|A|^2)$ where $|A|$ is the number of vertices in the reduced graph F of a concept.

As the estimation of the stability index is found to be dependent on the estimation of total clique counts, so in future, better estimation of the latter will result in even better bounds for stability.

References

1. Belohlavek, R., Trnečka, M.: Basic Level in Formal Concept Analysis: Interesting Concepts and Psychological Ramifications. In *IJCAI*, pp. 1233–1239, June, 2013
2. Wille, R.: Formal concept analysis as mathematical theory of concepts and concept hierarchies. *Formal Concept Anal.* **3626**, 1–33 (2005)
3. Carpineto, C., Romano, G.: *Concept Data Analysis: Theory and Applications*. Wiley (2004)
4. Kuznetsov, S.O.: On computing the size of a lattice and related decision problems. *Order* **18**(4), 313–321 (2001)
5. Kuznetsov, S.O.: Stability as an Estimate of the Degree of Substantiation of Hypotheses Derived on the Basis of Operational, Similarity (1990)
6. Kuznetsov, S.O.: On stability of a formal concept. *Ann. Math. Artif. Intell.* **49**(1), 101–115 (2007)
7. Roth, C., Obiedkov, S., Kourie, D.G.: On succinct representation of knowledge community taxonomies with formal concept analysis. *Int. J. Found. Comput. Sci.* **19**(02), 383–404 (2008)
8. Buzmakov, A., Kuznetsov, S.O., Napoli, A.: Is concept stability a measure for pattern selection? *Procedia Comput. Sci.* **31**, 918–927 (2014)
9. Babin, M., Kuznetsov, S.: Approximating concept stability. *Form. Concept Anal.* 7–15 (2012)
10. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: *Ordered Sets*, pp. 445–470. Springer, Netherlands (1982)
11. Wolff, K.E.: A first course in formal concept analysis. *StatSoft* **93**, 429–438 (1993)
12. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer Science & Business Media (2012)
13. Roth, C., Obiedkov, S., Kourie, D.: Towards concise representation for taxonomies of epistemic communities. *Lect. Notes Comput. Sci.* **4923**, 240–255 (2008)
14. Zhi, H.L.: On the calculation of formal concept stability. *J. Appl. Math.* (2014)
15. Moon, J.W., Moser, L.: On cliques in graphs. *Israel J. Math.* **3**(1), 23–28 (1965)
16. Hopcroft, J., Tarjan, R.: Efficient algorithms for graph manipulation (No. STAN-CS-71-207). Department of Computer Science, Stanford University, California (1971)
17. Bollobás, B., Nikiforov, V.: Cliques and the spectral radius. *J. Comb. Theory Ser. B* **97**(5), 859–865 (2007)
18. Nikiforov, V.: Some inequalities for the largest eigenvalue of a graph. *Comb. Probab. Comput.* **11**(2), 179–189 (2002)
19. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. *J. Exp. Theor. Artif. Intell.* **14**(2–3), 189–216 (2002)
20. Wille, R.: Concept lattices and conceptual knowledge systems. *Comput. Math. Appl.* **23**(6–9), 493–515 (1992)