A Concept-based Feature Extraction Approach

Ray R. Hashemi
Department of Computer Science
Armstrong State University
Savannah, GA, USA
Ray.Hashemi@gmail.com

Azita Bahrami
IT Consultation
Savannah, GA, USA
Azita.G.Bahrami@gmail.com

Nicholas R. Tyler
Department of Biology
Armstrong State University
Savannah, GA, USA
Romtinian@gmail.com

Matthew Antonelli
Department of Computer Science
Armstrong State University
Savannah, GA, USA
Mattr.Antonelli@gmail.com

***Abstract*—A concept has a perceived property and a set of constituents. The goal of this investigation is about extraction of meaningful relationships, if any, between the perceived property and the constituent's attributes. Such meaningful relationships (features) may be used as a prediction tool. The presented methodology for extracting the features is based on the concept expansion. To the best of our knowledge, feature extractions based on a concept expansion approach, for use in data mining, has not been reported in the literature. The goal was met by introducing the b-concept, conceptualizing a universe of objects using b-concept, and generating the complete gamma-expansion (CGE) of the b-concepts. The features were extracted from CGEs as anchor prediction (AP) rules. The AP rules were crystalized by a sequence of horizontal-vertical reductions. The prediction powers of the AP rules and their crystalized version were investigated by: (i) using 10 pairs of training and test sets, and (ii) comparing their performances with the performance of the well-known ID3 approach over the same training and test sets. The results revealed that the AP rules and ID3 have similar performances. However, the crystallized prediction rules have a superior performance over the AP rules and ID3. The average of the correct prediction is up by 17%, the average of the false positive is down by 13%, and the average of false negative is up by 3%. In addition, the number of test objects that cannot be predicted is down by 7%.**

***Keywords-b-concept; Concept expansion; Concept Analysis; Data Mining; Prediction Systems; and Crystallizing Prediction Rules*.**

I. INTRODUCTION

A concept is an abstract object possessing a perceived property [1]. For example, a "carcinogen agent" is a concept and its perceived property is that it causes, say, liver cancer. The constituents of a concept are a set of concrete objects described by their own set of attributes. Since a concept has a perceived property, it is considered a proper vehicle for investigation of the possible relationship between its perceived property and its constituents' attributes. Such a feature extraction is more successful when the relationships among the concepts are also established. Building super-concepts and sub-concepts are a part of this effort. Several concepts may create a super-concept and a given concept may serve as a sub-concept of one or more super-concepts [2][3][4]. We introduce the *complete γ-expansion (CGE)* of a concept and provide a methodology to identify a new relationship between a super-concept and its sub-concept(s) using CGE. A concept has a *CGE*, if every sub-expansion of the concept satisfies a given condition set, γ.

The goal of this research effort is three-fold: (i) introducing b-concept, conceptualizing a large dataset of concrete objects (chemical agents) using the new b-concept, and if it is applicable, building the CGE of the concepts, (ii) Extracting the features from the CGEs as the prediction rules and crystalize them by horizontal and vertical reductions, and (iii) compare the prediction power of the prediction rules and crystalized version of the prediction rules against the prediction power of the decision tree approach of ID3 [5].

The remaining organization of this paper is as follows. The Related Works is covered in Section 2. The Methodology is introduced in Section 3. The Empirical Results are discussed in Section 4. The Conclusions and Implications for Future Research is the subject of Section 5.

II. RELATED WORKS

The concept-based analysis is done primarily for building the internal conceptual structure of a given body of objects [6][7], conducting data mining [3][8], and performing image understanding [9]. As a result, the formal concepts [1][4], rough concepts [10][11], fuzzy concepts [12][13] and other forms of concepts [9] have been developed. There are some efforts in learning from the concepts for the purpose of performing a prediction process [3][8]. However, in such efforts number of generated concepts are limited and so the number of perceived properties. One may argue that every object can be considered as a concept with its own perceived property. Thus, it makes more sense that every possible perceived properties participate in the process of extracting features and not a limited number of them. The proposed methodology supports the total inclusion of all the possible perceived properties (inclusion trait).

The concept expansion has been heavily investigated in information retrieval for the purpose of query expansions to retrieve more relevant or pseudo-relevant documents (objects) from a corpus of documents [14][15]. In general, the concept expansion is done by changing the "bag of terms (features)" that are relevant to the query to a new larger bag of features that seems more relevant. In fact, such expansion tries to include more relevant features to improve the retrieval of more relevant objects. Such methodology does not have any application in mining data for prediction. In contrast, the concept expansion that we propose includes more relevant objects to improve the extraction of more relevant features (inducing trait) and it has a great potential to serve as a prediction approach.

To the best of our knowledge, there is not any existing concept-based prediction methodology that supports both inclusion and inducing traits.

III. METHODOLOGY

First, some terminologies need to be defined. Second, the expansion of the super-concepts is presented. Third, the extraction of features is explored, and finally, the crystallization of the extracted features is investigated.

Definition 1: Let U be a universe of objects and c = {O_1, . . ., O_s, . . ., O_n} ⊂ U. The subset c makes a b-concept, if $f(A_j^i, A_j^s) \le$ b, (for i =1 to n, i≠ s, and j =1 to m) where, A_j^i is the j-th attribute of O_i, m is the number of attributes for O_i, $f(A_j^i, A_j^s) = \sqrt{(A_j^i - A_j^s)^2}$, and b is a constant value. O_1, . . ., O_s, . . ., O_n are the members of the b-concept c, c = b-concept(O_s), and O_s is the concept's anchor—G(c) = O_s.

Definition 2: If G(c_k) ∈ c_m, then c_k is a sub-concept of c_m ($c_k \le c_m$, where ≤ is a binary relation) and c_m is a super-concept of c_k.

Definition 3: Let L be all the concepts of U. L is a partial ordered set with binary relation of ≤ and (L, ≤) is a compete lattice.

As an example, let us consider the set of objects in Table 1 and b = 3. Using definition 1, the concept c_4 with the anchor of O_4 is composed of the following objects c_4 = {O_4, O_5, O_6, O_7}. The concept c_6 with the anchor of O_6 includes objects of O_4, O_6, and O_7, c_6 = {O_4, O_6, O_7}. Since the anchor of c_6 is a member of c_4, then c_6 is the sub-concept of c_4 and c_4 is the super-concept of c_6—using definition 2.

TABLE I. A SET OF OBJECTS.

Object	A1	A2	A3	A4	A5
O_1	-1	1	2	3	5
O_2	2	-2	4	6	2
O_3	5	-3	2	3	3
O_4	6	3	5	1	8
O_5	7	4	3	-2	10
O_6	5	4	2	2	7
O_7	6	4	2	2	8

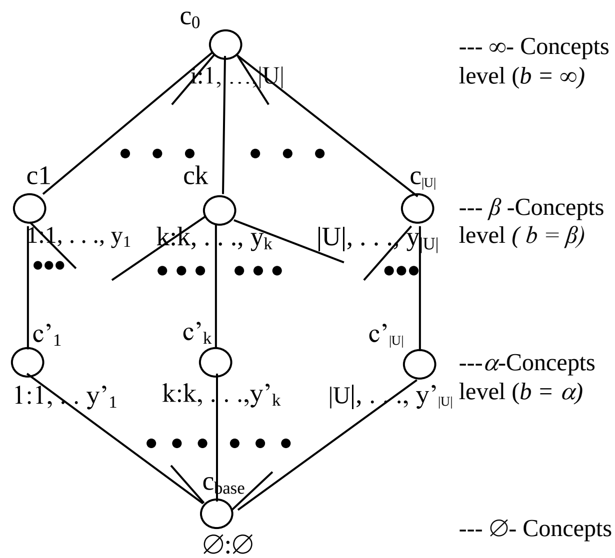

Figure 1. A lattice of concepts for the universe of object U and the two values of b = β and b = α.

The constant b is in the range of [0, +∞]. Let us build the lattice for those b-concepts of U generated for two different values of b (α and β, where α < β)—using definition 3. The lattice has four levels. The first level contains apex, concept c_0. The second level includes all the concepts for which b = β, (β-concepts). At the third level, all the concepts for which b = α (α-concepts) are included. The last level contains the base. At each level, there are |U|, not necessarily distinct, concepts, such that every object of U serves as the anchor of one concept; see Figure 1.

Each concept has a concept name, c_i, anchor, G(c_i), and members. The notation G(c_i): O_i, . . ., O_y is used to display the anchor and members of the concept c_i. The concept at the apex, c_0, includes all the objects of the universe as its members (b = ∞). Thus, any member can be designated as the anchor of the concept of c_0. Therefore, G(c_0) = O_i. The concept at the base includes no objects.

Reader needs to keep in mind that because of the huge range of values for constant b, the resulting lattice may have infinite number of levels. Building such an extremely large lattice is unnecessary because: (i) once the value of b reaches to the point that forces all the objects of U into one concept, then all the levels of the lattice beyond that b value have exactly the same concepts and (ii) turning the entire

 ISBN: 978-1-61208-364-3

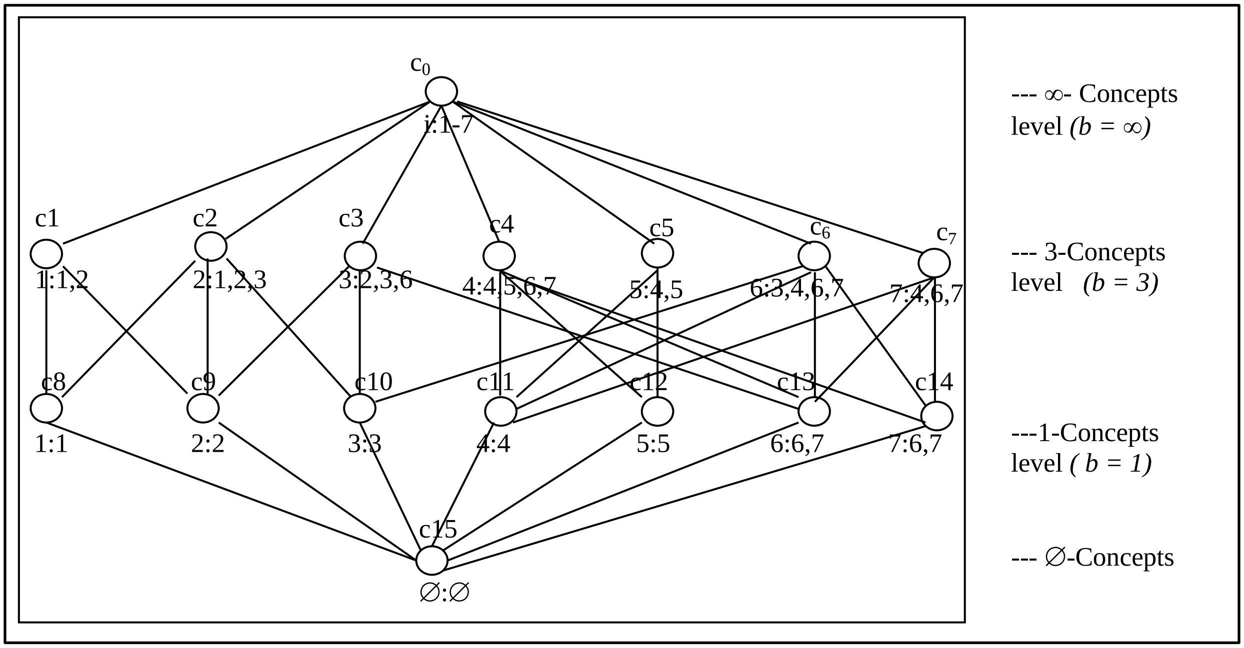

Figure 2. The lattice for the objects of Table I, α=1 and β=3.

TABLE II. THE γ-EXPANSION OF C4: (a) C4 AND ITS MEMBERS AND (b) THE COMPLETE γ-EXPANSION OF C4.

Concept	**Members**	**α-concepts of Member**	**Cover of α-concept**
c_4	O_4	c_{11}	c_4, c_5, c_6, c_7
	O_5	c_{12}	c_4, c_5
	O_6	c_{13}	c_3, c_4, c_6, c_7
	O_7	c_{14}	c_4, c_6, c_7

(a)

Concept	**Members**	**α-concepts of Member**	**Cover of α-concept**
c_4	O_4	c_{11}	c_4, c_5, c_6, c_7
	O_5	c_{12}	c_4, c_5
	O_6	c_{13}	c_3, c_4, c_6, c_7
	O_7	c_{14}	c_4, c_6, c_7
	O_3	c_{10}	c_2, c_3, c_6

(b)

Using the same analogy, the partial expansion for c_{12} and c_{14} do not change c_4 either. However, the $c_4 \cup$ cover(C_{13}) changes c_4 by adding a new object O_3 to c_4, Table 2.b. The new c_4 is the total γ-expansion of the original c_4. Because of the object O_3, only one new α-concept of c_{10} is added to the list of α-concepts of c_4 which has the cover of {c_2, c_3, c_6}.

objects of U into one concept clearly makes the conceptualization process of U a moot one.

As an example, the lattice for the set of objects of Table 1, for b = α = 1 and b = β = 3 is shown in Figure 2.

A. *Super-Concept Expansion*

A sub-concept within a lattice of concepts may have several super-concepts that are collectively referred to as the *cover* of the sub-concept. For example, the cover for the α-concept of c_{11} in Figure 2 is: cover(c_{11}) = {c_4, c_5, c_6, c_7}

Definition 4: Let γ be a set of conditions that is used to discriminate against the β-concepts. Let also c_j be a β-concept satisfying γ. In addition, let c_j have q sub-concepts of (α_1-concept, . . , α_q-concept). Furthermore, let c_j be expanded by all the covers of one of its sub-concepts (α_p-concept), $c_j = c_j \cup$ Cover(α_p-concept). If the expanded c_j also satisfies γ, then the new c_j is the *partial γ-expansion* of c_i over α_p-concept. The concept c_j is *totally γ-expanded* over its members when all the possible partial γ-expansions of c_j are done. A totally γ-expanded c_j may have a new set of sub-concepts (α-concepts). The concept c_j reaches its *complete γ-expansion (CGE)* when it cannot have any more partial expansions.

As an example, let us assume that concept c_4 satisfies the condition set of γ. (The condition set of γ is explained in detail in the next subsection.) The c_4 includes objects O_4, O_5, O_6, and O_7. The α-concepts of c_4 along with their covers are shown in Table 2.a. The $c_4 \cup$ cover(c_{11}) is the first partial γ-expansion of c_4. Let us assume that the expanded c_4 also satisfies γ. The partial expansion does not add to the members of c_4. That is, the cover of c_{11} includes concepts c_4, c_5, c_6, and c_7 that collectively contain objects of O_4, O_5, O_6, and O_7 which is the same as the objects in c_4 prior to expansion.

ISBN: 978-1-61208-364-3

Further expansion of c_4 for creation of its CGE starts with the $c_4 \cup \text{cover}(c_{10})$. This partial expansion changes c4 by adding object O_2 to c_4. Let us assume that the expansion does not satisfy γ. As a result, the CGE of c_4 includes the objects of O_3, O_4, O_5, O_6, and O_7.

After the CGE of a β–concept is obtained the following two steps take place:

a. Removing those concepts along with all the dangling edges from the lattice that their anchors are found in the complete γ-expansion of the β-concept.

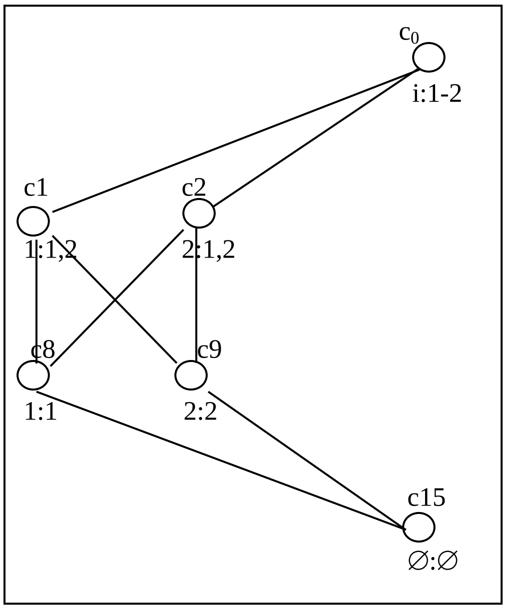

Figure 3. The adjusted lattice after removing the CGE of c4.

b. Deriving a new object as the anchor of the β–concept. The attribute A_i of the new object has the value equal to the average of all the A_i values of its members. Since the expanded β–concept is different from its original, calculation of the new anchor is necessary.

The result of lattice reduction using the CGE of c_4 is shown in Figure 3. Considering Table I, the new anchor for the CGE of c_4 has the attribute values of: (5.8, 2.4, 2.8, 1.2, 7.2).

Following the same process, another complete-γ-expansion is produced, from Figure 3, that belongs to c_1. The reduction of the lattice using the CGE of c_1, causes the deletion of the entire lattice. This indicates the end of the process of the concept expansion. The attribute values for the anchor of the CGE of c_1 are: (0.5, -05, 3, 4.5, 3.5).

B. *Feature Extraction*

The driven forces behind the feature extraction from the universe of objects, U, are the conceptualization of U, and γ-condition set. The former one contributes to the size, depth and cost of building the lattice and the later one contributes to the expansion of the qualified concepts.

The lattice size and depth are influenced by the number of objects in U the values for b, respectively. Since b values are too many, the building of the lattice is prohibitively expensive. Thus, building of only a two-level lattice (excluding apex and base levels) is preferred. The small values for b are more attractive because they relax the forced conceptualization of objects in U. As a result, the concepts are more organic and so their internal characteristic (features).

The extracted features are influenced by the γ-condition set. Let us assume that we are after extracting features that can be used for prediction (a set of prediction rules). To complete such extractions, a γ-condition set is introduced for the purpose of discriminating against concepts such that the concepts with a weak set of features be filtered. To explain it further, a decision attribute is assigned to every object in U. This attribute does not participate in conceptualization of U. The prediction of the value of the decision attribute for a set of new objects is the ultimate goal. If the minimum 2/3 of the members of a concept have the same decision as the anchor of the concept, the concept satisfies the γ-condition set—(i.e., the concept is a qualified one). Therefore, the concept's members collectively own a set of internal characteristics that support the decision of the concept's anchor with the strength of 2/3 out of one.

During the complete γ-expansion of a concept, the γ-condition set filters the unwanted partial γ-expansions to protect the strength of the internal characteristics of the concept. Since the anchor of a complete γ-expansion of a concept represents the entire members of the concept, so it represents their internal characteristics of them too.

The extracted features from CGE of a concept in form of prediction rules are referred to as the *anchor prediction (AP) rules* and presented in the production rules format. The AP rules extracted from the objects of Table I, are shown below using the two new anchors of the CGEs for concepts of c_4 and c_1. Let us assume that the decision for the anchors of c_4 and c_1 are d_1 and d_2, respectively. The AP rules are:

$(A_1 = 5.8, A_2 = 2.4, A3 = 2.8, A_4 = 1.2, A_5 = 7.2) \rightarrow d_1$
$(A_1 = 0.5, A_2 = -0.5, A_3 = 3, A4 = 4.5, A_5 = 3.5) \rightarrow d_2$

C. *Crystallization of the Extracted Features*

Let us assume that the extracted set of AP rules is applied on a given test set, TE, and the quality of the prediction outcome is measured, Q. The crystallization of the AP rules is started by applying first the horizontal and then the vertical reductions. The details for both reductions are covered in the following two subsections.

1. *Vertical Reduction of the AP Rules:* The goal of this reduction is to remove as many rules as possible from the set of AP rules without lowering the prediction ability of the set. To meet the goal, one rule is randomly removed from the set, the new set of AP rules is applied against the

 ISBN: 978-1-61208-364-3

test set of TE, and the quality of the prediction is measured, Q'. If Q' ≥ Q then: (a) the new set of AP rules replaces its predecessor and (b) Q' becomes the new Q. This process continues until no more rules can be removed from the set. There is a chance that none of the rules can be removed. This means the prediction rules cannot be vertically reduced.

2. *Horizontal Reduction of the AP Rules:* The goal of this reduction is to make the list of attributes for the entire AP rule set as short as possible. To meet the goal, one attribute, A_i, is kept in the AP rule set and the rest of the attributes are removed. The new AP rules are applied on the TE and the quality of the prediction is measured. This process is repeated for every attribute and at the end the attribute with the highest prediction quality, Q', is the winner. If Q' ≥ Q, then the winner attribute is the smallest subset of attributes representing the horizontal reduction of the set. If this is not the case, then another attribute is added to the winner attribute and the quality of prediction is checked for the pair. This process is repeated for every possible pair and at the end the pair with the highest prediction quality, Q'', is the winner. If Q'' ≥ Q, then the winner pair is the smallest subset representing the horizontal reduction of the set. If the condition of Q'' ≥ Q is not true, the winner pair grows to three attributes and this process continues until the minimum subset of the attributes is found with the prediction quality, at least, as good as Q. There is a chance that such subset cannot be found. This means the prediction rules cannot be horizontally reduced.

IV. EMPIRICAL RESULTS

An object set describing the properties of 1018 chemical agents was provided by a team of bio-chemists. Each chemical agent had eight attributes. One of the attributes is the decision and indicates whether the agent is carcinogen or not. The ten percent of the objects with decision zero along with the ten percent of the objects with decision one are randomly selected to make the test set. One may create 10 different test sets randomly such that the test sets do not have any objects in common. Let us consider one of the test sets. After creating the test set the remaining records are used as a training set. However, the training set must include equal number of objects for both decisions and include the largest number of objects as possible. As a result, we created 10 pairs of training and test sets such that the test sets did not have any objects in common.

For each pair, (i) the conceptualization of the training set was done for b = α = 0 and b = β = 1 and (ii) the AP rule set was generated and used to predict the decision for the objects of the test set and the quality of predictions was measured. We compared the prediction performance of the AP rule set, and the reduced AP rule set. The comparisons are shown in Table 3. All the training sets had both horizontal and vertical reductions. We have used both sequence of horizontal-vertical reductions and vertical-horizontal reductions of the AP rule set and the former one produced better prediction results and they are the ones shown in Table 3.

TABLE III. THE COMPARISON OF THE PREDICTION POWER OF THE ID3, AP RULE SET, AND REDUCED AP RULE SET FOR 10 PAIRS OF TRAINING AND TEST SETS.

Pair	Method	% correct prediction	% False (+)	% False (-)	% Not predicted
1	**ID3**	66.7	24	0	9.6
	AP Rules	66.7	28.5	4.8	0
	Crystalized AP Rules	90.5	2.4	7	0
2	**ID3**	76.2	19	0	4.8
	AP Rules	76.2	17	7	0
	Crystalized AP Rules	88.1	2.4	9.5	0
3	**ID3**	66.7	26	1	7
	AP Rules	81.0	9.6	4.8	4.8
	Crystalized AP Rules	92.9	0	7	0
4	**ID3**	71.4	17	1	12
	AP Rules	78.6	9.6	12	0
	Crystalized AP Rules	88.1	0	12	0
5	**ID3**	71.4	21	1	7
	AP Rules	61.9	24	12	2.4
	Crystalized AP Rules	83.4	0	17	0
6	**ID3**	76.2	12	2	12
	AP Rules	69	9.6	14	7
	Crystalized AP Rules	88.1	2.4	7	0
7	**ID3**	69	14	0	17
	AP Rules	69	12	12	0
	Crystalized AP Rules	90.5	4.8	4.8	0
8	**ID3**	66.7	24	0	9.6
	AP Rules	66.7	17	17	0
	Crystalized AP Rules	88.1	2.4	9.5	0
9	**ID3**	64.3	24	0	24
	AP Rules	66.7	17	9.6	7
	Crystalized AP Rules	85.7	0	14.3	0
10	**ID3**	73.8	12	2	14
	AP Rules	69	14	17	0
	Crystalized AP Rules	88.1	2.4	9.5	0
Avg	**ID3**	**70.2**	**12**	**2**	**11.7**
	AP Rules	**71.2**	**15.8**	**11.02**	**2.12**
	Crystalized AP Rules	**88.4**	**1.7**	**9.8**	**0**

One may raise the question of how good is the performance of AP rules in reference to other algorithms used for prediction. One of the well-known algorithms used for classification and prediction is ID3. We use ID3 to extract rules from each training set and measure the quality of the extracted rules in predicting the decision for the objects of the corresponding test set. The results are also shown in Table 3.

V. CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

The results presented in Table 3 revealed that the AP rule set performs as good as the ID3 algorithm. That means the proposed concept analysis has a high potential to serve as a prediction tool. The Crystalized AP rule set has a superior performance in compare with both ID3 and AP rule set.

We have observed that the anchor of a concept represents all of the concept members. Based on this observation, one may replace a large set of objects with a much smaller set of their anchors. Reduction of the size of the universe of objects may be useful in reducing the size of a Big Data. By doing so, one may be in a better position to analyze a very large object set. As part of the future research, this investigation is in progress.

Through the entire conceptualization process, we assumed that all the attributes of an object have the same strengths. This assumption is quite true for some universe of the objects such as the one used for obtaining the results shown in Table 3. However, the assumption is false for some other universe of objects. As another part of the future research, we revisit the creation of the b-concepts using varying strengths for the attributes of the objects.

REFERENCES

[1] R. Wille, “Restructuring Lattice Theory: An Approach based on Hierarchies of Concepts, Ordered Sets,” Reidel Dordecht-Boston Publisher, 1982.

[2] B. Ganter, B. and R. Wille, “Formal Concept Analysis: Mathematical Foundations,” Springer-Verlag Publishing, Berlin, 1999.

[3] J. S. Deogun, V. V. Raghavan, and H. Sever, “Association Mining and Formal Concept Analysis,” Proc. the Joint Conference in Information Science, Research Triangle Park, NC, 1998, pp. 335-338.

[4] R. Hashemi, L. LeBlanc, and T. Kobayashi, "Formal Concept Analysis in Investigation of Normal Accidents,” the International Journal of General Systems, vol 33, no. 5, October 2004, pp. 469-484.

[5] J. R. Quinlan, “Induction of Decision Trees,” Machine Learning”, vol. 1, no. 1, 1986, pp. 81-106.

[6] V. Evans, “Lexical concepts, cognitive models and meaning-construction,” Journal of Cognitive Linguistics, vol. 17, 2006, pp. 491-534.

[7] L. Barsalou, “Continuity of the conceptual system across species,” Journal of Trends in Cognitive Sciences, vol. 9, 2005, pp. 309-311.

[8] R. Hashemi, L. Le Blanc, A. Bahrami, M. Bahar, and B. Traywick, “Association Analysis of the Alumni Giving: A Formal Concept Analysis,” the International Journal of Intelligent Information Technologies, vol. 5, no. 2, April-June, 2009, pp. 17-32.

[9] R. Hashemi, L. Sears and A. Bahrami, “An Android Based Medication Reminder System: A Concept Analysis Approach,” Proc, the 19th International Conference on Conceptual Structures (ICCS'11), Sponsored and proceedings published by Springer-Verlag as Lecture Notes in Artificial Intelligence, (LNAI 6828) Derby, UK, July 2011, pp. 315-322.

[10] L. Yu and H. Liu, “Efficient Feature Selection via Analysis of Relevance and Redundancy,” The ACM Journal of Machine Learning Research, vol. 5, Dec. 2004, pp.1205-1224.

[11] Y. Lei and M. Luo, “Rough concept lattices and domains,” Annals of Pure and Applied Logic, Joint Workshop Domains VIII: Computability over Continuous Data Types, Y. L. Ershov, K. Keimel, U. Kohlenbach and A. Morozov (eds.), Novosibirsk, Rusia, Published by Elsevier, vol. 159, no. 3, Sept. 2009, pp. 333-340.

[12] R. Dietz and S. Moruzzi (editors), “Cuts and clouds. Vagueness, Its Nature, and Its Logic.” Oxford University Press, 2009.

[13] A. Markusen, "Fuzzy Concepts, Scanty Evidence, and Policy Distance: The Case for Rigor and Policy Relevance in Critical Regional Studies," In: Regional Studies, vol. 37, no. 6-7, 2003, pp. 701-717.

[14] D. Metzler and W. B. Croft, “Latent Concept Expansion Using Markov Random Fields”, SIGIR’07, Amsterdam, The Netherlands, July 2007, pp. 311-318.

[15] Y. Qui and H. P. Frei, “Concept Based Query Expansion”, Proc. of the 16th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR’93), 1993, pp. 160-169.

 ISBN: 978-1-61208-364-3