IJRRECS/October 2013/Volume-1/Issue-6/1423-1428 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

COMPETENT PATTERN RECOGNITION INTENDED FOR TEXT MINING

A.Manasa[1], B.Madhavi Devi[2], Ch.Srinivasulu[3]

[1]Dept of IT, J.B. Institute of Engineering & Technology, Hyderabad, A.P, India

[2]Assistant Professor, Dept of IT, J.B. Institute of Engineering & Technology, Hyderabad, A.P, India

[3]Associate Professor, Dept of IT, J.B. Institute of Engineering & Technology, Hyderabad, A.P, India

ABSTRACT:

In the process of knowledge discovery, data mining is the fundamental measure in the process. The techniques of data mining have been used for the analysis of text by means of extracting the terms of co-occurring as descriptive phrases from the collection of document. A variety of well-organized algorithms has been planned which are determined principally for the development of well-organized mining algorithms for the discovering of patterns from a collection of huge data. The techniques of data mining have been used for the analysis of text by means of extracting the terms of co-occurring as descriptive phrases from the collection of document. The approach of pattern mining-based has been projected, that implements the notion of patters of closed sequential, and pruned non closed patterns to prevail over the drawbacks of approach of phrase-based and reveals the enhancement on terms of efficiency to a convinced extent. The techniques of pattern mining can be used to locate a variety of text patterns, such as recurrent itemsets, sequential patterns, multiple grams, and co-occurring terms in the field of text mining. To progress the efficiency by efficiently using closed patterns in text mining, the model of pattern taxonomy was proposed. An exceptional performance was achieved by the approach of pattern taxonomy models intended for text mining by means of comparing with the modern methods of data mining, the well-known methods of term-based, and the concept models. The proposed

pattern taxonomy models results reveal that the model can generate hopeful growth in efficiency over the other models.

Keywords: Data mining, Pattern mining, Text patterns, Pattern taxonomy model.

1. INTRODUCTION:

The process of nontrivial mining of information from huge databases that is unidentified previously and potentially functional for users is known as the process of knowledge discovery. In the process of knowledge discovery, data mining is the fundamental measure in the process [8]. The detection of interesting information in text documents is known as text mining and is a difficult concern to discover precise knowledge in text documents to help users in locating their desire. The approach of pattern mining-based has been projected, that implements the notion of patters of closed sequential, and pruned non closed patterns to prevail over the drawbacks of approach of phrase-based and reveals the enhancement on terms of efficiency to a convinced extent [1]. Low frequency and misinterpretation are the two basic concerns concerning the efficiency of pattern-based approaches. In data mining communities, the pattern mining has been examined expansively for several years. A variety of well-organized algorithms such as PrefixSpan, Apriori-like algorithms, GST and so on has been planned which are determined principally for the development of well-organized mining algorithms for the discovering of patterns from a collection of huge data [5] [11]. The techniques of data mining have been used for the analysis of text by means of extracting the terms of co-occurring as descriptive phrases from the collection of document. The techniques of pattern mining can be used to locate a variety of text patterns, such as recurrent item sets, sequential patterns, multiple grams, and co-occurring terms in the field of text mining [6]. The notion of closed patterns in text mining was practical and had the potential for recovering the performance of text mining. To progress the efficiency by efficiently using closed patterns in text mining, the model of pattern taxonomy was proposed. To considerably get better the

performance of information filtering, a two-stage model that makes usage of both term-based and pattern based methods was introduced [15]. The strategy of inner pattern deploying model provides an effectual assessment intended for dropping the side effects of noisy patterns for the reason that the assessment of term weights that are in the term space is based on not only statistical properties but also associations of patterns in the equivalent pattern categorization. An exceptional performance was achieved by the approach of pattern taxonomy models shown in fig1 intended for text mining by means of comparing with the modern methods of data mining, the well-known methods of term-based, and the concept models [3] [7]. The proposed pattern taxonomy models results reveal that the model can generate hopeful growth in efficiency over the other models. A modern computational technology that assists to recognize the meaning of text documents is known as natural language processing. A new concept-based representation was presented to overpass the gap between natural language processing and text mining that analyze the terms on the documents and sentence levels [14]. This model contains three components such as the primary component analyzing the semantic arrangement of sentences; the second constituent constructs a conceptual ontological graph in order to explain the semantic structures; and the final component extracts the top perceptions based on the primary two components to construct feature vectors by means of the standard vector space representation [4].

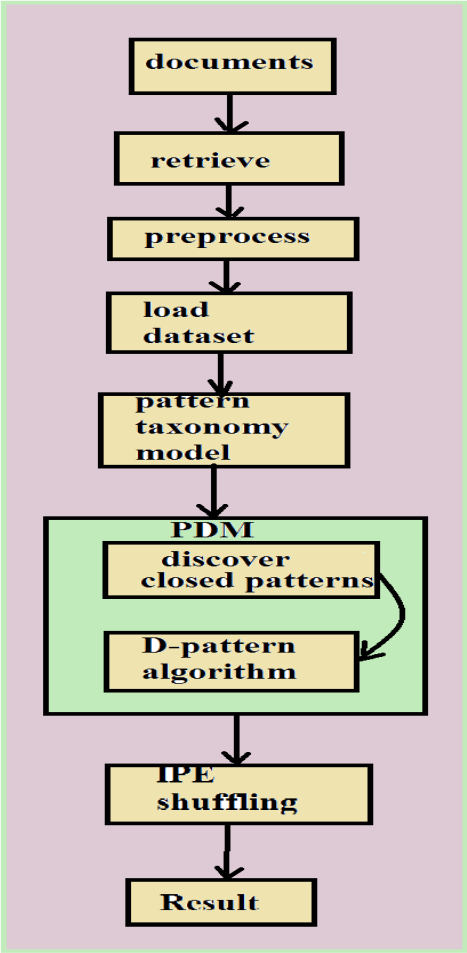

Fig1: An overview of pattern discovery

2. METHODOLOGY:

To get better the performance of closed patterns in text mining, sequentially to employ the semantic information in the pattern organization we necessitate interpreting discovered patterns by means of summarizing them as d-patterns in order to precisely assess term weights [9] [13]. The d-patterns comprise more semantic sense than terms that are particularly based on a technique of term-based technique is the justification following this motivation. A term with a senior term-based technique assessment could be worthless if it has not cited by several d-patterns [2]. The evaluation of term weights is dissimilar to the normal approaches of term-based. In the approach of term-based, the assessment of term weights is based on the allocation of terms in documents [10]. The technique that is used to diminish the side effects of noisy patterns because of the problem of low-frequency is known as inner pattern evolution since it alters merely a pattern's term supports inside the pattern [12]. Documents are classified by the threshold into appropriate or inappropriate categories. The general notion of pattern updating is that initially total conflict offenders are disconnected from d-patterns. Their term supports are reorganized consecutively to decrease the effects of noise documents for partial conflict offenders.

3. RESULTS:

An exceptional performance was achieved by the approach of pattern taxonomy models (inner pattern evolution) intended for text mining by means of comparing with the modern methods of data mining, the well-known methods of term-based, and the concept models. The proposed pattern taxonomy models results reveal that the model can generate hopeful growth in efficiency over the other models such as CBM and SVM. The use of the deploying method is capable for resolving the misinterpretation difficulty and the results can be elucidated. The strategy of inner pattern deploying model provides an effectual assessment intended for dropping the side effects of noisy patterns for the reason that the assessment of term weights that are in the term space is based on not only statistical properties but also associations of patterns in the equivalent pattern categorization.

4. CONCLUSION:

The approach of pattern mining-based has been projected, that implements the notion of patters of closed sequential, and pruned non closed patterns to prevail over the drawbacks of approach of phrase-based and reveals the enhancement on terms of efficiency to a convinced extent. The techniques of pattern mining can be used to locate a variety of text patterns, such as recurrent itemsets, sequential patterns, multiple grams, and co-occurring terms in the field of text mining. To progress the efficiency by efficiently using closed patterns in text mining, the model of pattern taxonomy was proposed. An exceptional performance was achieved by the approach of pattern taxonomy models intended for text mining by means of comparing with the modern methods of data mining, the well-known methods of term-based, and the concept models. The proposed pattern taxonomy models results reveal that the model can generate hopeful growth in efficiency over the other models. The strategy of inner pattern deploying model provides an effectual assessment intended for dropping the side effects of noisy patterns for the reason that the assessment of term weights that are in the term space is based on not only statistical properties but also associations of patterns in the equivalent pattern categorization.

REFERENCES:

[1] M. Sassano, "Virtual Examples for Text Classification with Support Vector Machines," Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '03), pp. 208-215, 2003.

[2] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Prefixspan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," Proc. 17th Int'l Conf. Data Eng. (ICDE '01), pp. 215-224, 2001.

[3] K. Sparck Jones, S. Walker, and S.E. Robertson, "A Probabilistic Model of Information Retrieval: Development and Comparative Experiments—Part 2," Information Processing and Management, vol. 36, no. 6, pp. 809-840, 2000

[4] M. Seno and G. Karypis, "Slpminer: An Algorithm for Finding Frequent Sequential Patterns Using Length-Decreasing Support Constraint," Proc. IEEE Second Int'l Conf. Data Mining (ICDM '02), pp. 418-425, 2002.

[5] D.D. Lewis, "An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task," Proc. 15th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '92), pp. 37-50, 1992.

[6] N. Jindal and B. Liu, "Identifying Comparative Sentences in Text Documents," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '06), pp. 244-251, 2006.

[7] H. Ahonen, O. Heinonen, M. Klemettinen, and A.I. Verkamo, "Applying Data Mining Techniques for Descriptive Phrase Extraction in Digital Document Collections," Proc. IEEE Int'l Forum on Research and Technology Advances in Digital Libraries (ADL '98), pp. 2-11, 1998.

[8] K. Sparck Jones, S. Walker, and S.E. Robertson, "A Probabilistic Model of Information Retrieval: Development and Comparative Experiments—Part 1," Information Processing and Management, vol. 36, no. 6, pp. 779-808, 2000.

[9] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," Proc. European Conf. Machine Learning (ICML '98),, pp. 137-142, 1998.

[10] Y. Li, C. Zhang, and J.R. Swan, "An Information Filtering Model on the Web and Its Application in Jobagent," Knowledge-Based Systems, vol. 13, no. 5, pp. 285-296, 2000.

[11] T. Rose, M. Stevenson, and M. Whitehead, "The Reuters Corpus Volume1—From Yesterday's News to Today's Language Resources," Proc. Third Int'l Conf. Language Resources and Evaluation, pp. 29-31, 2002

[12] Y. Xu and Y. Li, "Generating Concise Association Rules," Proc. ACM 16th Conf. Information and Knowledge Management (CIKM '07), pp. 781-790, 2007.

[13] S. Shehata, F. Karray, and M. Kamel, "A Concept-Based Model for Enhancing Text Categorization," Proc. 13th Int'l Conf. Knowledge Discovery and Data Mining (KDD '07), pp. 629-637, 2007.

[14] T. Joachims, "Transductive Inference for Text Classification Using Support Vector Machines," Proc. 16th Int'l Conf. Machine Learning (ICML '99), pp. 200-209, 1999.

[15] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB '94), pp. 478-499, 1994