International Journal of Academic Multidisciplinary Research (IJAMR)
ISSN: 2000-006X
Vol. 2 Issue 3, March – 2018, Pages: 1-4

A Survey on Idea Mining: Techniques and Application

Nicholaus J. Gati[1], Lusekelo Kibona[2]
[1]Information Systems department
University of Dodoma, Dodoma, Tanzania
nicholaus.gati@udom.ac.tz

[2]Computer Science department
Ruaha Catholic University, Iringa, Tanzania
lusekelo2012@gmail.com

Abstract: *Idea mining is an interesting field in the area of information retrieval and it is increasingly becoming important asset for decision makers. Huge volumes of high quality data from various sources such as scanners, mobile phones, loyalty cards, the web, and social media platforms presents enormous opportunity for organization to achieve success in their businesses. It is possible to achieve this by properly analysing data to reveal feature patterns; hence decision makers can capitalize upon the resulting ideas from wealth of information available. Idea mining helps managers to use their time efficiently by grasping ideas from a huge amount of unstructured data with text mining, instead of reading through page by page, it also helps to avoid the danger of overlooking data with useful ideas. The main objective of this survey article is to provide assessment on a number of successful text mining and text classification tools combined with idea mining measures to extract the idea from texts.*

Keywords: *Text mining; Idea mining; decision making; Feature learning*

1. INTRODUCTION

An idea is the thought of human, can be expressed in various ways such as opinion, plan, image, prospection, prediction and evaluation on a particular theme[1]. Also Kruse et al. [2] described an idea as the large amounts of data in human mind that can be represented in the form of images. In the information era, automatic detection and analysis of meaningful ideas holds key on growth and development of organizations and businesses. Enormous amounts of heterogeneous data have become available on hand to decision makers. It is very difficult for decision makers to read all the available structured, semi structured and unstructured data without loss of information as huge amount of these data are in textual formats. In order to gain advantage of the rapid growth of such data and help decision makers to formulate proper decision making, solutions are required to handle and extract value and knowledge from these datasets. Text mining or knowledge discovery is that sub process of data mining that is widely used to discover hidden patterns and significant information from the huge amount of unstructured written material. This text mining uses techniques of different fields like machine learning, case based reasoning, visualization, database technology statistics, text analysis, knowledge management, natural language processing and information retrieval. Idea mining is a natural extension of text mining as it is based on getting valuable information from texts relying on generated patterns, models or other rules for interesting and useful unstructured text. To achieve the goal of extracting valuable knowledge from the available text documents, text mining uses automated methods. Text Mining represents a step forward from text retrieval. Text mining process involves five steps which includes document gathering, document pre-processing, text transformation, feature selection and data mining pattern selection. To realize the idea mining process, methods from text mining and text classification (tokenization, term filtering methods, Euclidean distance measure etc.) are combined with a new heuristic measure for mining ideas. As a result, the idea mining approach extracts automatically new and useful ideas from user given text. Therefore, there is a need for efficient idea mining techniques to take advantage of the available data. This paper provides an overview on different methods and tools which can be applied to mine ideas, as well as the opportunities provided by the application of idea mining in various decision domains.

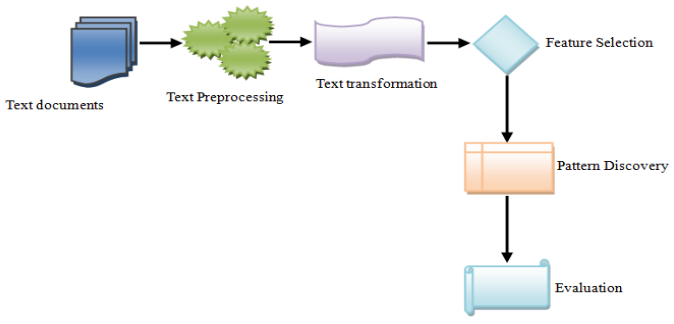

Figure 1**:** Text Mining Process flow

2. RELATED WORK

[3] Proposed the concept of idea mining as a method for obtaining new and useful ideas form unstructured data. They mainly focused on the idea definition that could solve

International Journal of Academic Multidisciplinary Research (IJAMR)
ISSN: 2000-006X
Vol. 2 Issue 3, March – 2018, Pages: 1-4

technological problems. Amandi [4] based research on determining factors that allow users of micro blogs to act as good source of information.

[5] suggested that, there are different domains of application which have different properties for an idea. Various studies from the past identified idea properties for different domain such as in medical domain, social domain and technological domain [5]. Even though there are various methods and techniques in collecting and analysing data and results, idea mining is flexible to be applied in any domain using any method for data collection and analysing results.

There are many idea processing models as concluded by[6], all of them model ways to characterize, extract, create and evaluate new ideas. Some methods consider the existing problem then create some ideas related to the problem using text mining techniques [7].

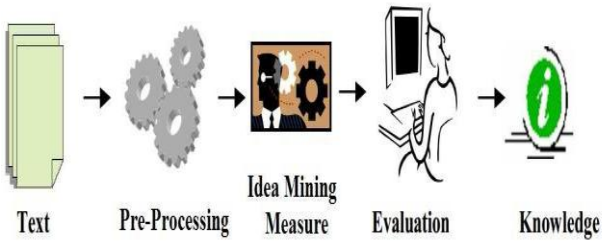

Figure 2**:** Idea Mining Process

Different idea processing models may vary in their details, but according to [6], they consists three necessary steps as follows:

• Preparation of collected document: In the first step, the idea mining approach focused on the provided textual information by the user that contained description of the problem.

• Extraction of text patterns (ideas) from a new text and collection document: In the second step, the user has to provide further textual information which may contain an idea that probably can solve the problem. Therefore, with an automatic process, text patterns appeared in a very large number of overlapped texts should be extracted. Text patterns are then compared to the problem description by using a specific idea mining measure. With this measure, text patterns can be classified as an idea.

• Evaluation of text patterns: In the third step, all extracted text patterns are evaluated for their usability and usefulness.

Thorleuchter in [6] introduced a new approach for extracting ideas from unstructured text automatically. The extraction of a very large number of the text patterns from the overlapped text phrases was done based on the proposed equation (1).

$$T_i = \omega_{i-\min(\sum_i^{k=1}(\omega_{i+k}))} \cup \omega_i \cup \omega_{i+\min(\sum_i^{k=1}(\omega_{i-k}))} \ldots (1)$$

According to [6]building text patterns depends on the length *l* and term weights of stop words and non-stop words.

Figure 3, depicts an example on how to create text patterns by using equation (1).

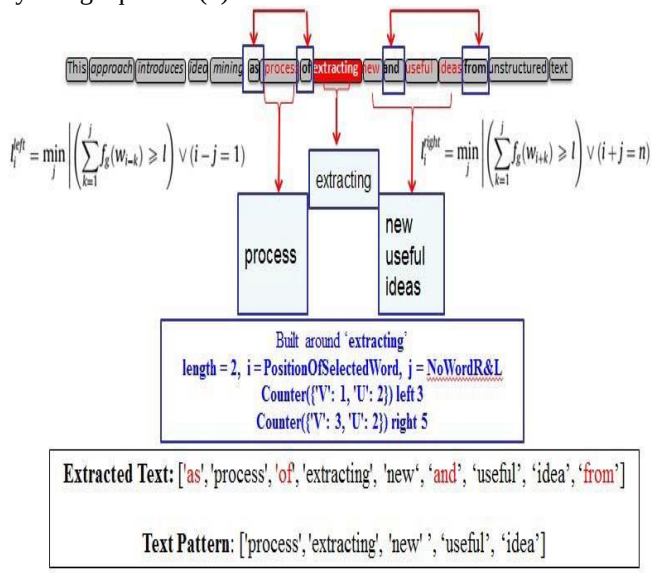

Figure 3**:** This example shows how text pattern is created from text

According to [8], similarity measurement is significant in organizing documents, hence measuring similarity is one of the most fundamental document analysis technique especially in information retrieval.

Idea mining measure uses different similarity measurement techniques, such as cosine similarity, Euclidian distance to mention a few. Given two documents $\vec{t_a}$ and $\vec{t_b}$, their cosine similarity is as follows:

$$SIM_C(\vec{t_a}, \vec{t_b}) = \frac{\vec{t_a}.\vec{t_b}}{|\vec{t_a}| \times |\vec{t_b}|}$$

Where $\vec{t_a}$ and $\vec{t_b}$ are m-dimensional vectors over the term set $T = \{t_1 \ldots \ldots t_m\}$. Each dimension represents a term with its weight in the document, which is non-negative. As a result, the cosine similarity is non-negative and bounded between [0, 1] [9]. In idea mining applications the most preferred and frequently used distance measure is the Euclidian distance[6], as it consumes less time. [6] Proposed a new idea mining measure which compares the similarities between vectors from new text to similar vectors in the text collection. In more recent studies for mining ideas [10], two step classification is done, by comparing term vector from new text to all vectors from text collections using Euclidean distance measure or any other similarity measure, and then comparing each vector from the new text to its most similar vectors using the proposed idea mining measure.

3. METHODS AND MATERIAL

The methodology adopted by this study was 'Internet Search'. The study consulted different sources on the Internet to establish evidence and facts about the claimed issues. Where possible the websites of the specific resource were visited, for example website of some journals which only put materials in html format rather than pdf or documents. The reviewed literatures are mostly available on the Internet. Another means employed is observations and where possible in some areas algorithm were developed to facilitate the discussion. So generally secondary source of data were mainly used in a large part to come up to conclusion.

4. DISCUSSIONS

i. IDEA EVALUATION

Intuitively, the successfulness of the information retrieval is accomplished when all relevant documents are retrieved through the given query that meets the human satisfaction. Human satisfaction is judged by the relevance of retrieved documents. {Relevant} represent a set of all relevant documents relating to a particular query.{Retrieved} represent a set of all retrieved documents. When some conditions are met, {Relevant}∩{Retrieved} represents relevant with retrieved documents

The two most frequent and basic measures for information retrieval effectiveness is precision and recall [11].

1. **Precision: -** Precision is calculating percentage of retrieved documents which are relevant to the query.
2. **Recall: -** Recall is calculating percentage of relevant documents which are relevant to the query.

ii. IDEA MINING APPLICATION

Idea Mining is used to discover ideas which are innovative by nature. [6] Implemented an application that enables an automated identification of ideas

Idea mining application allows extracting and discovering unstructured information sources to identify useful ideas for supporting strategic decision makers. Idea mining application is used for idea identification in several areas: product development, signal detection, technological, social and medical areas

5. CONCLUSION

Idea mining help managers to save time by analysing huge amounts of unstructured data, without any concern of losing data. Consequently, analysing documents and extracting ideas will result in a competitive advantage for a firm, hence firms has to utilize the advantages of idea mining in order to make good decisions because there is important hidden knowledge in textual datasets. Text mining technique is basically used for extracting pattern from unstructured data. Various techniques for efficiently performing idea mining are discussed in this survey. The main focus on this survey is basically to provide an overview about idea mining which has gained increasing attention in recent years. In the era of massive data availability idea mining is of great significance, and can provide unforeseen insights and benefits to decision makers in various areas. If properly exploited and applied, idea mining has the potential to provide a basis for advancements, on the scientific, technological, and humanitarian levels.

6. FUTURE WORK

Proposed methods mainly focus on extracting ideas from textual data, therefore missing out on extracting ideas from images, video, voice to mention a few. In the future the existing methods may be expanded to create a method that is capable of extracting meaningful ideas from textual data and other types of data by employing Information retrieval techniques.

7. ACKNOWLEDGEMENT

Parts of this paper benefited from related discussions with colleagues, notably Samwel Tarus.

Also we would like to thank friends Neema, Nelvin and Nelson Lusekelo Kibona for being there all the time when we needed their supports.

REFERENCES

[1] T.-Y. Lee, "'A study on extracting ideas from documents and webpages in the field of idea mining,'" *J. Korean Soc. Inf. Manag.*, vol. 29, pp. 25–43.

[2] and E. S. P. Kruse, A. Schieber, A. Hilbert, "'Idea mining–text mining supported knowledge management for innovation purposes,'" 2013.

[3] D. Thorleuchter, D. V den Poel, and A. Prinzie, "Mining ideas from textual information," *Expert Syst. Appl.*, vol. 37, pp. 7182–7188, 2010.

[4] M. G. Armentano, D. Godoy, and A. A. Amandi, "Followee recommendation based on text analysis of micro-blogging activity," *Inf. Syst.*, vol. 38, pp. 1116–1127, 2013.

[5] D. Thorleuchter and D. Van den Poel, "Extraction of ideas from microsystems technology," *Adv. Comput. Sci. Inf. Eng. Springer*, pp. 563–568, 2012.

[6] D. Thorleuchter, D. V. den Poel, and A. Prinzie, "Mining ideas from textual information," *Expert Syst. Appl.*, vol. 37, no. 10, pp. 7182–7188, 2010.

[7] A. Osborn, "Your Creative Power: How to Use Your Imagination to Brighten Life, to Get Ahead," *Univ. Press Am.*, 2008.

[8] S. M. Weiss, N. Indurkhya, and T. C. et Al, "TM: Predictive Methods for Analyzing Unstructured Information, USA," *Springer*, 2005.

[9] A. Huang, "Similarity measures for text document clustering," in *Proc. New Zealand Computer Science*

Research Student Conference (NZCSRSC), 2008, pp. 49–56.

[10] D. Thorleuchter and D. Van den Poel, "Companies website optimising concerning consumer's searching for new products," in *International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE)*, 2011, pp. 40–43.

[11] B. Lamiroy and T. Sun, "'Computing precision and recall with missing or uncertain ground truth,' in Graphics Recognition. New Trends and Challengesuth," *Springer*, p. 149–162., 2013.