

Survey on Data Mining Method for Finding Competitors from Large Unstructured Data

Mahendra Eknath Pawar

Research Scholar, RIMT University (Punjab)
Email : mahendraepawar@gmail.com

Dr.Satish Saini

Professor (ECE), RIMT University (Punjab)
Email : satishsainiece@gmail.com

Article Info

Volume 83

Page Number: 9716– 9725

Publication Issue:

May - June 2020

Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 18 May 2020

Abstract:

In present world Compleitive business, the achievement is indeed in light of the capacity to make more engaging clients of things than the opposition. Along the lines of research has demonstrated the strategic importance of identifying and monitoring company competitors. The marketing and management community, motivated by this issue, is focused on empirical methods for competitor identification and methods for analyzing known competitors. Such an expression can indeed be an indicator of competitiveness, but in many areas it does not exist.

Keywords: Competitors, Data mining, Electronic commerce, Information Search and Retrieval, Web mining.

I. INTRODUCTION

The world is becoming flat, and competition in various fields is increasing. On the other hand, in order to survive in such a competitive world, need to fully understand the position of your competitors. In the study of Fortune 500 companies about 55% use conflicting information about the management strategy. For example, when creating a prudent premise, when Microsoft entered a strategic decision new electricity. To speculation should include at least who is the leader in this area, and what is the key to the leader to success. For example, careful assumptions had to be made when Microsoft strategically decided to enter a new area electrical entertainment. Speculation should include at least who is the leader in this field and what a leader's key

to success. On the other hand, comparison with similar products with the same characteristics, such as price, functionality and quality, is also absolutely necessary for the consumer who buys a particular product. In this paper, we will consider such tasks related to the process of mining of competitive entities, domain and evidence for decision-making, as well as the tasks of mining and competitors.

Monitoring and identifying the company's competitors studied at an earlier job. Data mining is the best way to process vast amounts of information for mining competitor's product review format online can be enriched with information about your opinion. However, it is generally difficult to understand all the reviews on

different websites for competitive products and to obtain the insightful suggestions manually. In the early works of literature, many authors intelligently and efficiently analyzed such large customer data.

Recently, The researchers are pointed out that the use of artificial intelligence (AI) and data mining developed e-commerce applications. This study helps to understand the huge data and online reviews of the users. These studies help designers to understand the large volumes of customer requests and online reviews of products for improvement. However, these arguments are not enough and are far from potential problems. These include a thorough review of the product, an online review, and a thorough analysis of what competitors do. In fact, a typical scenario for a customer-driven new product design (NPD) requires a thorough analysis of its strengths and weaknesses, as it can be successful in a fierce market competition.

This analysis gives different philosophies actualized for my rival's link to life in respect of customer relations, conclusion and use of information mining [1].

Unstructured information administration:

The information gathered from the web is now semi-organized or unstructured and then. Semi-organized information can be found in XML, JSON and other configurations. Unstructured sources are in an

alternative organization that does not fall into a predefined class. While overseeing a huge number of clients, the business will have problems managing the increased expenses made by the connections between individuals. Nevertheless, if all the client information is embedded in the database, subsequent records can be used to identify the client by adding business insight to the mining procedure, by adding a competitor exchange information to pick up a better understanding of these clients. The creators claimed that information mining is a way of coping with the help an organization that grows more powerful technology to meet a confrontation in the market. Information warehousing is a valuable and accurate way to accumulate disparate and scattered pieces of business information and pass on detained useful data to procedures. Information Mining Innovation wrapped up in change to be leveraged learning to show information. Candidate mining from the Web information framework is very adaptable.

Therefore, amongst other concentrated procedures is outstanding and effective the use of web information for convenient selection help. Client information for the mining of competitors is collected through several strategies that are generally not structured.

An unstructured framework is one that does not have a foreordained shape or structure and is usually full of printed information. Unstructured framework included in Figure 1 indicates unstructured and organized the framework. The extraction of data from the pages of the site is a dynamic range of research.

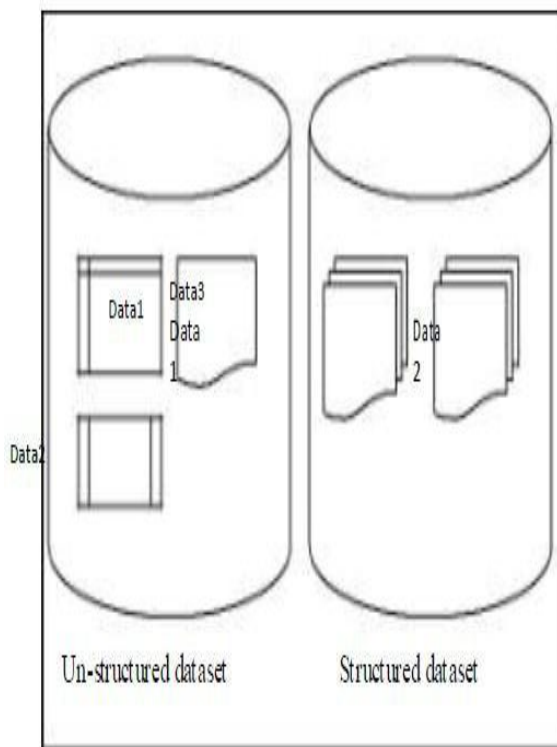


Fig. 1. The diagram of structured and Unstructured system [1]

Researchers were making different arrangements, from a broad perspective, to give the same report. Many web data extraction processing frameworks ask the customer to mark the purpose of the test; the principle of information extraction can be learned. For a controlled learning process, most self-loader

frameworks are more accurate than fully programmed ones that do not have human intermediaries. The self-loader technology is not suitable for large web applications where it is necessary to remove information from a large number of sites. In addition, the site tends to change a lot in the design of the site pages, the extraction rules created in the past have been disabled, and the ease of use of the self-loader strategy is even more that it is a job after a large number of reasons fully or completely focused on the programmed placement.

Web data extraction allows you to record all the best updates and information at the unit level. As a solitary information unit, every datum record last as a solitary information unit can be used to remove the point by the point Information Unit inside the information record and above the point. According to the record level extraction technology, the organized information extraction from the Web page included after recognizing the information area including every one of the large-scale records and distributing the information locale to the singular record has been widely thought. Early to be discovered in the rapper who developed the body with great deals to connect with various Web locales, because they are very serious work. To address this issue, proposed the acceptance of rappers as a self-loader strategy and

examined their effects. These strategies require a few name pages in the objective area as a contribution to playing enlistment. In this way they have confinement for extensive scale applications. To beat the above disadvantages, a fully programmed strategy was created. The creators were prone to the problem of unsupervised Web information extraction utilizing a fully programmed data extraction device called viper. The device allows you to concentrate on individual information display, repeating the structure of a lonely Web page with a high degree of accuracy and recognition of the pair with the use of video set data. Even so, this procedure requires the execution of a number of data sets [2] [3].

II. LITERATURE REVIEW

S.Babu, T.S.umamaheswari and P.Sumathi [1], the researchers showed, the formal meaning of intensity between the two things which approved both quantitative and subjective. Formalization is related to the cross on space, overcoming the weaknesses of past methodologies. They take into account the various variables that have been ignored so far, for example, the location of things in multidimensional part Space, customer trends and conclusions. O. R. L. Sheng and Z. Ma, G. Pant [2], propose an approach to classify competitive enterprises which are important for enterprises. Based on the structure of the intercompany system derived from the quotation of the enterprise in the online news fact,

the author presents an approach of the technique of graph theory and machine learning to conclude the correlation of the competition enterprise.

Ping Ji, Jian, Y.Jin, and RuiGu [3], author select a few representative sentences, the following were examined: expression of information, comparison of information, diversity of information. Three greedy algorithms are proposed to formulate the optimization problem and to analyze the problem. David Molnar, Benjamin Livshits and Saxena, Prateek [4], they suggest the application of Scriptgard, it detects and repairs the wrong placement of the disinfectant. ScriptGard can be useful as a test aid to developers and as a run-time mitigation technique. AtliBenediktsson, Johannes R. and Ghamisi [5], the author proposes hyper spectral images classification framework. The spatial information is extracted by using the Extended Multivariate Attribute Profile. The author overcome the shortcomings of EMAP to introduce the fully automatic scheme of the proposed method. Petrucci and Giulio [6], this paper proposes a semi-supervised ontological learning method. This methodology integrates a new axiomatic extraction technique utilizing some features of the text corpus. Anna Lisa, Isabelle Augenstein, Fabio Ciravegna and Gentile [7], in this work he uses the link data for the scale of Web information extraction and display, prompting the

results of the task of the wrapper. The main contribution of this work is that linked Data, which is an inaccurate, redundant and large-scale knowledge resource, effectively and efficiently, supports the extraction of information on the Web scale.

G. Lausen and K. Simon [8], in this paper, the problem of the extraction of the unsupervised web data is described. The paper shows that they can extract the web data without the teacher by assuming the page which consists of the repetition pattern, such as the result page of the search engine, and so on. Zelenko, Oleg Semin and Dmitry [9], the author presents an automated system to discover competing companies from public sources. Dimitrios Gunopulos, Theodoros, George Valkanas and Lappas [10], in this paper, the formal definition of competitiveness between the two items is described. They present an efficient way to assess the competitiveness of large data sets, and this method to the natural problem of finding top-k competitors for a given item is different areas. Olivia RL Sheng, Pant and Gautam [11], The authors believe that through rigorous analysis, including the contents of the company's website, more than 2,600 companies to measure the presence of on-line isomorphism, as well as to reveal the usefulness in predicting the relationship of competitors. Margaret A. Peteraf, Bergen and Mark [12], In this paper, gathering

insights from the areas of Strategic Management and marketing, the manager is a powerful tool to overcome this common problem, the authors take into account a wide range of competitors, including potential, competitor, alternative, and indirect competitors, the two-stage frame for the identification and analysis of competitors (in particular, Petafer in 2001). x. liu, B. yu and p.s .ding [13], the author propose a comprehensive dictionary-based approach to solve this problem using external evidence of natural language and language expressions. H. salem, Abbasi and A. chen [14], the paper study about the stylistic and syntactic features for the sentiment classification of English and Arabic content and mining the contents. qi, l, Wang, f, l and chen[15], in this paper, the methods are compared not only experimentally on the basis of real world review data sets, but also adopted a conditional random-field (CRF) model, and their performance is evaluated in comparison with related algorithms.

III. COMPARATIVE ANALYSIS

Table 1.Comparative Analysis Table

Sr. No	Paper Name	Author	Study	Conclusion
1.	MINING COMPETITORS FROM LARGE UNSTRUCTURED DATASETS USING CMINER	S.Babu, T.S.umamaheswari and P.Sumathi	The researchers showed, the formal meaning of intensity between the two things which approved both quantitative and subjective.	They had a computationally difficult problem of finding the best k candidate for a given product.
2.	MINING COMPETITOR RELATIONSHIPS FROM ONLINE NEWS: A NETWORK-BASED APPROACH	Z. Ma, G. Pant, and O. R. L. Sheng	The author proposes a neutral language method. The method can not use natural language processing for news.	The author concludes the conflict between the pairs of companies participating in the network, and the performance of competitors from various metrics.
3.	IDENTIFYING COMPARATIVE CUSTOMER REQUIREMENTS FROM PRODUCT ONLINE REVIEWS FOR COMPETITOR ANALYSIS	Ping Ji, Jian, Y.Jin, and RuiGu	Study a framework for selecting a pair of Representative comparative sentences with specific product features from competitive product reviews is exemplified with the help of emotion analysis techniques.	Three greedy algorithms are proposed to formulate the optimization problem and to analyze the problem.
4.	SCRIPTGARD: AUTOMATIC CONTEXT-SENSITIVE SANITIZATION FOR LARGE-SCALE LEGACY WEB APPLICATIONS	David Molnar, Benjamin Livshits and Saxena, Prateek	Proposes a Scriptgard application, it detect and repair the wrong placement of disinfectants.	There has been a significant increase in server and browser context.
5.	AUTOMATIC SPECTRAL SPATIAL CLASSIFICATION FRAMEWORK BASED ON ATTRIBUTE PROFILES AND SUPERVISED FEATURE	AtliBenediktsson, Johannes R. and Ghamisi	The classification framework is proposed for hyper spectral images.	They introduce a fully automatic scheme of the proposed method to overcome the shortcomings of EMAP.

- | | | | | |
|-----|--|--|--|---|
| 6. | EXTRACTION
INFORMATION
EXTRACTION
FOR LEARNING
EXPRESSIVE
ONTOLOGIES | Petrucci and
Giulio | Propose several specific
research directions for
designing effective
methodologies for semi-
supervised ontological
learning. | This methodology integrates a new
axiomatic extraction technique utilizing
some features of the text corpus. |
| 7. | UNSUPERVISED
WRAPPER
INDUCTION

USING LINKED
DATA | Anna Lisa,
Isabelle
Augenstein,

Fabio
Ciravegna
and Gentile | In this work he uses the
link data for the scale of
Web information

extraction and display,
prompting the results of
the task of the wrapper.
This method is very
flexible for different
domains and does not
require training
materials, but uses
linked data as a
background knowledge
source to construct
essential learning
resources | The main contribution of this work is that
linked Data, which is an inaccurate,
redundant and large-scale knowledge
effectively
resource, and efficiently,
supports the extraction of information on
the Web scale. |
| 8. | VIPER:
AUGMENTING
AUTOMATIC
INFORMATION
EXTRACTION WITH
VISUAL
PERCEPTIONS | G. Lausen
and K. Simon | The problem of the
extraction of the
unsupervised web data
is described. | The system for high-precision extraction,
which is combined with regular internal
structure of the complex web page. |
| 9. | AUTOMATIC
COMPETITOR
IDENTIFICATION
FROM
PUBLIC INFORMATION
SOURCES | Zelenko,
Oleg Semin
and Dmitry | The author presents an
automated system to
discover competing
companies from public
sources. | Validate the results of the system and
deploy it on the web as a powerful
analytical tool for retail and institutional
investors. |
| 10. | EFFICIENT
AND DOMAIN-
INVARIANT
COMPETITOR
MINING | Dimitrios
Gunopulos,
Theodoros,
George
Valkanias and
Lappas | Describe the definition
of competitiveness. | Present an efficient way to assess the
competitiveness of large data sets. |
| 11. | WEB
FOOTPRINTS OF
FIRMS: USING
ONLINE
ISOMORPHISM | Olivia RL
Sheng, Pant
and Gautam | Proposed predictive
models can be
effectively applied to the
focused company. | Find the models to finding the
competitiveness. |

- | | | | | |
|-----|--|--------------------------------------|---|---|
| | COMPETITOR IDENTIFICATION | | | |
| 12. | COMPETITOR IDENTIFICATION AND COMPETITOR ANALYSIS: A BROAD BASED MANAGERIAL APPROACH | Margaret A. Peteraf, Bergen and Mark | In this paper, gathering insights from the areas of Strategic Management and marketing. | From the framework for competitor identification to develop a hierarchy of competitor recognition, it is used in conjunction with resource equivalence to generate hypotheses about competitive analysis. |
| 13. | HOLISTIC LEXICON-BASED APPROACH TO OPINION MINING | x. liu, B. yu and p.s .ding | In particular, to study the issue of determining the semantic orientation(positive, neutral negative or) of opinions on the product features of the review. | A system called opinion observer is implemented. |
| 14. | A SENTIMENT ANALYSIS IN MULTIPLE LANGUAGES: FEATURE SELECTION FOR OPINION CLASSIFICATION IN WEB FORUMS | H. salem, Abbasi and A. chen | The study proposes the use of sentiment analysis techniques for the classification of opinions of Web forums in multiple languages. | The proposed features and technologies will be evaluated in benchmark movie review data sets and in US and Middle East web forum posts. |
| 15. | COMPARISON OF FEATURE-LEVEL LEARNING METHODS FOR MINING ONLINE CONSUMER REVIEWS | qi, l, Wang, f, l and chen | The methods are compared notonly experimentally on the basis of real-world review data sets, but also adopted a conditional random-field (CRF) model, and their performance is evaluated in comparison with related algorithms. | It was revealed that the CRFs-based method is more accurate. |
| 16. | GATHER CUSTOMER CONCERNS FROM ONLINE PRODUCT REVIEWS – A TEXT SUMMARIZATION APPROACH | liu, y, j, loh, zhan and j, loh | They propose an automated summary approach based on an analysis of the internal topic structure of the review article to assemble customer concerns. | They show that the user's responsiveness and ability to discover the most important topics are superior to the peer approach, namely the summary of opinion mining and clustering. |
| 17. | IDENTIFYING COMPARATIVE | ping ji, Ruigu. Jin | In this study, a framework for selecting | The actual large amount of data from Amazon as a result of the com, the |

CUSTOMER REQUIREMENTS FROM PRODUCT ONLINE REVIEWS FOR COMPETITOR ANALYSIS	and jian	a pair of Representative comparative sentences with specific product.	category of large-scale experiment was carried out, and the final promising result was realized, and the effectiveness of the proposed approach was prove.
18. MINING COMPARATIVE OPINIONS FROM CUSTOMER REVIEWS FOR COMPETITIVE INTELLIGENCE	y. song, k. xu, j. li, and s. s. liao	The study of opens the door for companies to analyze the rich consumer generated data for risk management.	From the verious sources the functions frameworks are limited.

IV. CONCLUSION

The study shows that the importance of identifying and monitoring corporate competitors. The proposed competitive definition is a feature of the space where the user requests a uniform distribution of values. This approach is based on the assumption that such comparative evidence is found in abundance in the available data. This work will present an end-to-end way to mine such information from large data set in customer reviews.

REFERENCES

- [1] S.Babu, T.S.umamaheswari and P.Sumathi, "Mining Competitors from Large Unstru-ctured Datasets Using CMiner", IJCRT 2017.
- [2] Z. Ma, G. Pant, and O. R. L. Sheng, "Mining Competitors from Large Unstructured Datasets", 2016.
- [3] Ping Ji, Jian, Y.Jin, and RuiGu. "Identifying comparative customer requirements fromproduct online reviews for competitor analysis" Engineering Applications of ArtificialIntelligence 49, 2016.
- [4] David Molnar, Benjamin Livshits and Saxena, Prateek, "SCRIPTGARD: automaticcontext-sensitive sanitization for large-scale legacy web applications." ACM 2011.
- [5] AtliBenediktsson, Johannes R. and Ghamisi, "Automaticspectral aspatial classification framework based on attribute profiles and supervisedfeature extraction." IEEE 2014.
- [6] Petrucci and Giulio, "Information extraction for learning expressive ontologies" , Springer, Cham, 2015
- [7] Anna Lisa, Isabelle Augenstein, Fabio Ciravegna and Gentile, "Unsupervised wrapper induction using linked data." ACM, 2013.
- [8] G. Lausen and K. Simon, "ViPER: Augmenting Automatic Information Extractionwith Visual Perceptions," ACM 2009
- [9] Zelenko, Oleg Semin and Dmitry, "Automatic competitor identification from publicinformation sources" International Journal of Computational Intelligence andApplications 2.03 (2002)
- [10] Dimitrios Gunopulos, Theodoros, George Valkanas and Lappas, "Efficient anddomain-invariant competitor mining." ACM, 2012.
- [11] Olivia RL Sheng, Pant and Gautam, "Web footprints of firms: Using online isomorphism for competitor identification" Information Systems Research26.1 (2015)
- [12] Bergen, Mark, and Margaret A. Peteraf. "Competitor identification and competitor analysis: a broad based managerial approach" 2002.
- [13] x. liu, B. yu and p.s .ding, "a holistic lexicon-based approach to opinion mining", [2008]
- [14] H. salem, Abbasi and A. chen, "a sentiment analysis in multiple languages: feature selection for opinion classification in web forums", [2008]
- [15] qi, l, Wang, f, l and chen, "comparison of feature-level learning methods for mining online consumer reviews", [2012]
- [16] liu, y, j, loh, zhan and j, loh, "gather customer concerns from online product reviews – a text summarization approach", [2009]

- [17] ping ji, Ruigu. Jin and jian, “identifying comparative customer requirements from product online reviews for competitor analysis”, [2016]
- [18] y. song, k. xu, j. li, and s. s. liao “mining comparative opinions from customer reviews for competitive intelligence”, [2011]