



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Extraction of Popular Product Attributes by Using Unsupervised Machine Learning

Kousar Ghalwade, Dr. Mahesh Ghaikwad, Prof. Amruta Sarudkar

PG Student, Department of AIDS, Sanjay Ghodwat University, Kolhapur, India

Associate Professor, Department of CSE, Sanjay Ghodawat University, Kolhapur, Maharashtra, India

Assistant Professor, Department of Computer Engineering, Trinity College of Engineering & Research, Pune, India

ABSTRACT: Marketing of e-commerce networks has low cost, good connections, and other advantages in the worldwide traffic. So, the advantage is that it attracts consumers, decreases maintenance, and increases the revenue of internet marketing. By using popular product attributes from the product specification pages, we create a graphical mode from a variety of new domains and websites. Without using samples of labeling data and current techniques for extracting information by taking account of the popularity of product characteristics, it allows for mapping of the popular product properties and the related product characteristics from several customer reviews. The important thing is that it fills this vocabulary gap between the text on product summary pages and the text on the client's reviews.

KEYWORDS: Machine Learning, Supervised, Unsupervised ML, preprocessing, Extraction Attributes, Features

I. INTRODUCTION

In this project, we are designing an unsupervised framework to draw popular product attributes from the product information from various websites. This framework can not only detect popular product characteristics from a set of customer reviews but also map these popular features to the corresponding product characteristics simultaneously as compared to existing systems. The proposed system aims to build a system that does not experiment with the number of popular product attributes in advance. It can be extracted by automatically recognizing hidden concepts, derived from a set of customer reviews and by addressing the vocabulary from the text in the web-based product attributes and the text in the customer review process. This allows consumers to easily view similarities between the products needed and decide quickly to purchase products online. In today's technology world, data science, machine learning, and artificial intelligence are some of the top trend subjects. There are developments in DATA mining and Bayesian analytics and this raises the need for computer education. Machine learning is a programming discipline to allows them to learn and increase their knowledge automatically. Training in this context includes the identification and appreciation of the data provided and informed decisions based on the data provided. Certain decisions are based on all relevant feedback and are very difficult to take into account.

II. LITERATURE SURVEY

1. Innovation of E-commerce Fresh Agricultural Products Marketing Based on Big Internet Data Platform [Lan Li1, Ying Zhang Miao] E-commerce is suggested as the best method for maximizing the performance of this program. In the context of the electronic commerce setting, research is required to carry out the study procedure for the electronic commerce patterns in a new model of electronic commerce that provides how the invention of the electronic commerce pattern is helpful for us, too, to formulate the electronic trade pattern in an enterprise. Experimental results show that the rate of system data acquisition is high and can boost network traffic efficiency and protection in real-time, through the connections of each module
2. Quantitative Analysis of the Internal Quality Control and Financing Constraints in Electric Power Enterprises based on the SEM model [Daxiang Suo1, Shuibin Jiang2, Ming Qi1, Li Chen3]: The quantitative research approach of internal quality control and funding limitations for electricity companies based on the SEM model proposes. Internal control and corporate governance are two important aspects of the modern enterprise system. The structure of the internal control system and mechanism of corporate governance is rational and efficient to improve the competitiveness of power companies and the necessary precondition for improved electricity industries, especially the scientific and effective internal control system, improving the company's core strength and enhancing the competitiveness of the company sector. The internal control quality is described in this survey as integrated internal control capacity assessment and operational effectiveness.

3. Research on the Growth Mechanism of Agricultural Production Enterprises from the Perspective of Resource Dynamic Supply [Yan Xu¹, Hong Chen¹, Chang Liu²]: Agricultural production is experiencing a marked change in the United States as consumer demands, input costs, food safety issues, and environmental consequences move rapidly. Multidimensional components and drivers that work in a complex way to influence production sustainability compose agricultural production systems. In a mixed method approach, we combine quantitative and qualitative data to create and simulate a system dynamic model that examines the structural relationship between these drivers with the economy, climate, and agricultural production society
4. Product aspect ranking using sentiment analysis and TOPSIS [Saif A. Ahmad Ali Alrababah¹, Keng Hoon Gan², Tien-Ping Tan³]: The rapid expansion of client evaluations on e-commerce sites motivated many scientists to explore the issue of recognizing the product aspects listed in online reviews. Many studies the study are based on three main criteria: 1) the drawing up of aspects that have been reported on regularly in online reviews, 2) the determination of important aspects that many consumers in their reviews have defined positively and negatively, and 3) the relation between the product aspect of the domain (e.g. the camera) and other aspects. This paper uses sentiment analysis and TOPSIS (Technique for Order Quality by Similarity to Ideal Solution) to propose a new product element ranking system in its response to these proposals. The proposed work is divided into two stages: the extraction of aspects and the classification of aspects. In the extraction aspect point, sentiment analysis is applied to understand the product aspects of customer reviews based on the 3 extraction criteria. In the second stage, TOPSIS simultaneously included the product aspects derived from the preceding parameters to generate the most representative product aspects. The analytical assessment of the research proposed using four items online tests indicates its efficacy in defining representative aspects
5. Unsupervised Extraction of Popular Product Attributes from E-Commerce Web Sites by Considering Customer Reviews [Lidong Bing, Taklam Wong, Wailam]: An unregulated learning system for removing attributes from product description pages was developed at various Web sites for e-commerce. Unlike current extracted information approaches that take into account the popularity of product attributes, our proposed system can detect popular product features from a collection of customer reviews as well as map popular product features. One of our new features is to address a language gap between the product overview page text and the customer feedback site. We are theoretically constructing a segregated graphic model based on secret random conditional fields. As an unsupervised model, it is easy to apply our system to various new domains and websites without requiring marking samples.
6. A Sentiment Classification Model Based On Multiple Classifiers [Cagatay Catal¹]: Customer reviews have become a critical factor in consumer decisions through the extensive use of social networks, forums, and blogs. Since the early 2000s, researchers have been focusing on such comments to categorize them automatically as polarities, such as positive, negative, and neutral. This problem is known as the description of sentiments. The goal of this study was to investigate and propose a new classification methodology for the potential advantage of the multi-classified system definition for the Turkish sentiment classification. For 3 classifiers, viz. Naive Bayes, the SVM, and Bagging, voter algorithms were used. SVM parameters were optimized when utilized as an individual classifier. Experimental results demonstrated that multiple classifier systems increase the efficiency of individual classifiers in Turkish sentiment classifying datasets and that these multiple classifier systems contribute to their strength. The method suggested achieved better efficiency as Naive Bayes, the best individual classification for these datasets, and Support for Vector Machines. A strong sentiment classification method is offered by multiples classifier systems (MCSs), which should be considered when designing an MCS-based prediction system for the parameter optimization of individual classifier modes.
7. Extracting Attribute: Value Pairs from Product Specifications on the Web [Petar Petrovski, Christian Bizer]: The author provided the system of derived value pairs from specifications of products on the Internet. The author. Supervised learning is used to separate or not the HTML table and HTML list into a webpage. We again use controlled learning to identify columns as columns of attribute or value columns to extract attribute-value pairs from the HTML fragments found in the specification detector. We add several new features for specification detection compared with DEXTER, the current state-of-the-art method for extracting value-paired attributes of product specification, and support the extraction by pairs of specifications having more than two columns. The developers of the report suit a Bing search engine dataset. Developers use historical knowledge in their approach to build attributes and to fit schemas.
8. Aspect extraction in sentiment analysis- comparative analysis and survey [Toqir A. Rana¹, Yu-N Cheah¹]: Sentiment analytical (SS), due to the expansion of the World Wide Web (WWW), has become one of the most important and increasingly common areas of information and text mining. SA deals with the measurement of the looks, thoughts, and emotions of users concealed in the text. Extraction of elements is SA's most critical and thoroughly studied process to accurately identify sentiments. Such methods were graded according to the method adopted. A thorough comparative analysis is carried out among various approaches of aspects extraction, despite

being a traditional survey, which not only elaborates the performance of any technique but also leads readers to compare precision to other state-of-the-art and most recent approaches.

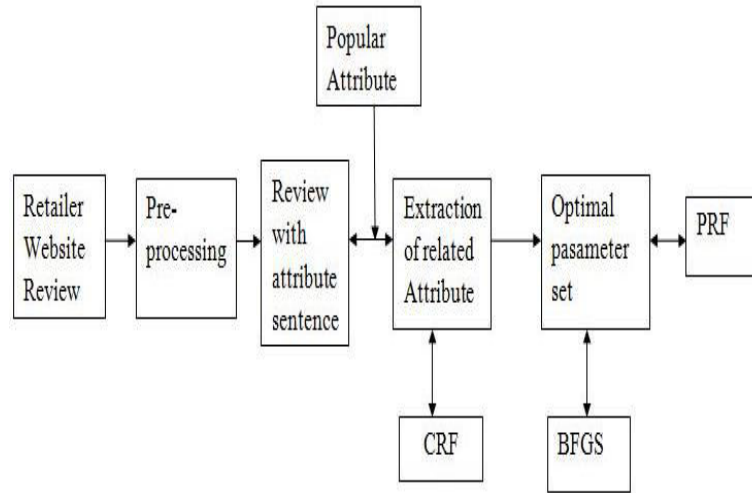
9. A Feature Terms Extraction Process based on Customer User Recommendation Polarity Analysis [Tomofumi Yoshida, Daisuke Kitayama]: Paper provides a method for the extraction of feature conditions that reflect the feelings over the use of a customer reviews product on websites a recommendation based on content. Given previous research, which indicates that negative and impressive experiences have more effect than positive ones, we describe the conditions about factors on which consumers disagree in comments about the advantages and disadvantages of sensations relating to product usage. Our methodology consists of extracting sentences from consumer reviews that express opinions and considers each assessed word as a candidate for product features. Using each candidate's positive opinion, we extract feature words for the chosen product by looking at a feature score based on a positive evaluation ratio, to determine how divided the opinions of the reviewers are. We are presenting an experiment to determine the usefulness of function words extracted using our method.
10. Innovation of E-commerce Fresh Agricultural Products Marketing Based on Big Internet Data Platform [Lan Li1*, Ying Zhang Miao2]: In practice, agricultural production in our country was a family as a small production unit, the rural household often relies not on old prices to choose the project and decides the size of the production between agricultural products. E-commerce is expected to increase efficiency. In the sense of electronic trade, the research into an examination of electronic business trends is relevant, as the new electronic trade trend is very important to us because it offers a way for electronic commerce innovation to also be useful in formulating a specific e-commerce pattern Beginning with this impetus and using the large Internet network, this paper proposes the innovative idea of the marketing of electronic commercial fresh agricultural products.
11. Quantitative Analysis of the Internal Quality Control and Financing Constraints in Electric Power Enterprises based on the SEM model [Daxiang Suo1, Shuibin Jiang2, Ming Qi1, Li Chen3]The requirement to increase the competitiveness of power companies is a sensible and effective internal control system. The authors examine the internal quality control and funding constraints in electricity companies based on the models of SEM in this paper. The empirical analysis has revealed that corporate governance has an important influence on internally controlled quality and has an influence of 0.163, and the impact on the efficiency of investment of the funding constraints is 0.702. As a result, improving the quality of internal control by the electricity companies has a positive effect on the performance of business while conducive to overcoming financing constraints.
12. Product aspect ranking using sentiment analysis and TOPS [Saif A. Ahmad Ali Alrababah1, Keng Hoon Gan2, Tien-Ping Tan3]: The explosive growth of e-commerce customer reviews has prompted many researchers to investigate the problem of recognizing the product issues mentioned in online reviews. Many research studies are based on 3 key criteria: 1) on the aspects that were consistently addressed in online review; 2) on the important aspects that many clients consider positive and negative in their assessment; and 3) on the relationship between the domain product factor (such as 'camera') and the other expectations. In the extraction process, feeling analyses are used to identify product aspects of customer reviews based on the three extraction criteria. In the second phase, TOPSIS concurrently included the product aspects taken from the previous criterion for the preparation of a list of the most important product aspects.
13. Study on the Enterprise Financial Management System Design and Accounting Optimization based on E-commerce Background [Ming Yang]: The market data and financial information are centralized comparatively in the e-commerce world so that the financial reports system can easily understand the current financial data activities and quickly produce financial information. The author analyzes the architecture and accounting implementation of the financial management system of the company in this article.
14. Unsupervised Extraction of Popular Product Attributes from E-Commerce Web Sites by Considering Customer Reviews [Lidong Bing, Taklam Wong, Wilam] Unsupervised learning process for processing popular product attributes from product interpretation. Compared to conventional knowledge extraction methods that do not take into consideration the importance of products, our proposed approach is capable not only of identifying popular product range features but also of comparing these popular features to the product attributes in consideration.

III. PROPOSED SYSTEM

PROBLEM DEFINATION

Features and attributes both refers to the characteristic of a specific domain object. Use "attribute" in the product description is used to refer to such element and use "popular feature" used to refer to the concepts of the feedback that have been uncovered. This model generates associations between the attributes and popular features automatically to classify popular attributes. Therefore, it is supposed that every feature discovered in the reviews is a popular feature.

IV. SYSTEM ARCHITECTURE



V. OVERVIEW OF THE FRAMEWORK

In this proposed work, two main components are included. The first element is the common extractor attribute feature to extract text remains from the product description Web pages that match the typical attributes. Web pages are considered to be a type of semi-structured text document that includes a mixture of considered materials such as HTML tags and free text that can be either ungrammatical or short sentences. Given the unique tokens page in W ($tok1, \dots, tokN(W)$) within the domain, our objective is to classify all $tok1, k$ text fragments such that $V(tok1, k) = A_j$ and $A_j = tokN(W)$ where $ApoP$ is known as the $APOP$. This role can be described as a problem with sequence labeling. We are marking each token with two sets of labels, specifically in $(tok1, -tokN(W))$.

The first set includes the "B," "I" and "O" labels that signify an attribute's beginning, both inside and outside an attribute. $A_j \in APOP$, that is, the form of popular features, is the second set of labels. To address sequence marking issues CRFs have been introduced as the latest model. However, the current standard CRF models are unsatisfactory. First, each token is labelled with two types of labels at the same time, whereas standard CRFs only recognize one type of label. The second reason is because, by the secret principles developed in customer reviews, the common characteristics are related and unknown in advance.

It fails to use supervised instruction in traditional CRFs. We have created a graphic model based on secret CRFs to address this problem. The graphical model proposed can use the hidden concepts derived and the clues from the features of the layout and text contents. There is also an unregulated learning algorithm to remove the common attributes. The second part is designed to extract $APOP$ automatically from a client analysis set (R). This portion produces a set of documents derived from R . A popular feature generates the terms "panel," "resolution," "screen," and so on while a popular feature produces "camera," "speaker" etc. Our graphical model can extract the text fragments associated with common attributes by using such details on words.

There are several contributions to this work. First, we have developed an inexperienced approach based upon hidden conditional random fields (CRFs) to extract product attributes from the product description pages by considering terms related to popular features as a problem with unknown attributes. The first contribution to this is that we are using the common attribute extraction as an extraction issue with unknown features. The second goal is to close the language gap between the features found in feedback and the features found in the product description. In that context, even common attributes can be extracted. Thirdly, on a wide range of product description pages obtained from 13 different fields we have carried out detailed studies. We also compared some current models that can fairly solve this unregulated common extraction problem. The experimental results will demonstrate our framework's efficiency and solidity.

Two major components form this conceptual structure. The first component is the popular extraction attribute, which extracts text fragments from the web pages that are popular attributes. Websites are considered to be a type of semi-structured text documents with a mix of structured contents such as HTML tags and free texts, ungrammatical or

composed of only short sentences. Firstly, every token is simultaneously branded with two types of labels, whereas standard CRFs recognize only one type of label. The second reason for this is that the common attributes are related to and unknown in advance by secret principles derived from customer reviews. The second component seeks to extract APOP automatically from a customer feedback set. This portion produces several documents derived from R. In addition to the hints of the interface features, the proposed graphic models will manipulate derived secret concepts. An unregulated learning algorithm for the removal of common attributes is also created. This graphical model can extract fragments of the text relating to popular attributes by using such information on terms.

VI. EXISTING SYSTEM

We describe a proposed structure with two important components in this existing system. This first aspect is the common element extraction function that finds to eliminate textual segments from those in the product description web pages. Web pages are considered to be semi-structured text documents with a mixture of standardized content such as HTML tags and free phrases that are either ungrammatical or comprise short sentences. The buying rate of e-commerce firms is lower. E-commerce company finding is performed on the network with actual benefits compared to the buying activity of other firms. The electronic commerce purchasing enterprise is founded in network contact, negotiation, and order completion contracts and will not be limited to other constraints on completion, transaction costs, saving time, and energy by using an online banking payment Process to complete the transaction on the network. However, The conventional supply is communicating by telephone, fax, and so on, and at last both sides need to meet and discuss communicating, and trading, and this is a slow mechanism rather than rapid contact between businesses and electronic commerce firms. The benefits of e-commerce are not only reflected in the remainder of the inventory but they can be monitored on a timely basis through the network itself, effectively reducing inventory costs, reducing product cost and cost. E-commerce enterprises based on network infrastructure use reflect the regional logistical strengths, create a broad and rational logistics system, make it more convenient to transport the product between the company and upstream providers and downstream customers, directly on corporate performance, speed up corporate capital activities.

VII. CONCLUSION AND FUTURE SCOPE

we are proposing that the e-commerce increase would create new opportunities. Financial management research is more attentive and developable in the new environment. There is therefore a very significant practical value to exploring financial management dependent on e-commerce. Electronic commerce can bring together pre-natal agricultural production, production, and various post-natal links to a solution, agricultural production, and information on the markets, not as a symmetrical issue which enables agricultural producers to promptly understand the information on the market and produce a fair organization that avoids the price h according to market requirements. The proposed model will address current challenges and provide the world with a new marketing scenario for new agricultural products. In this project, we can learn about the strategic climate, strategic goals, strategic material, and method of financial management for electronic commerce.

REFERENCES

1. Akash U. Suryawanshi, P. D. N. K. (2018). Review on Methods of Privacy-Preserving auditing for storing data security in the cloud. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCCE)*, ISSN, 7(4), 247–251.
2. Archana, R. C., Naveenkumar, J., & Patil, S. H. (2011). Iris Image Pre-Processing And Minutiae Points Extraction. *International Journal of Computer Science and Data Security*, 9(6), 171–176.
3. Ayush Khare, D. N. J. (2017). Perspective Analysis Recommendation System in Machine Learning. *International Journal of Emerging Trends & Technology in Computer Science*, 6(2), 184–187.
4. AyushKhare Nitish Bhatt, Dr Naveen Kumar, J. G. (2017). Raspberry Pi Home Automation System Using Mobile App to Control Devices. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(5), 7997–8003.
5. AyushKhare, J. G., Bhatt, N., & Kumar, N. (2017). Raspberry Pi Home Automation System Using
6. Mobile App to Control Devices. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(5), 7997–8003.
7. Bhore, P. R., Joshi, S. D., & Jayakumar, N. (2016). A Survey on the Anomalies in System Design: A Novel Approach. *International Journal of Control Theory and Applications*, 9(44), 443–455.

8. Bhore, P. R., Joshi, S. D., & Jayakumar, N. (2017a). A Stochastic Software Development Process Improvement Model To Identify And Resolve The Anomalies In System Design. *Institute of Integrative Omics and Applied Biotechnology Journal*, 8(2), 154–161.
9. Bhore, P. R., Joshi, S. D., & Jayakumar, N. (2017b). Handling Anomalies in the System Design: A Unique Methodology and Solution. *International Journal of Computer Science Trends and Technology*, 5(2), 409–413.
10. Desai, P. R., & Jayakumar, N. K. (2017). A Survey on Mobile Agents. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 5(XI), 2915–2918.
11. Divyansh Shrivastava Amol K. Kadam, Aarushi Chhibber, Naveenkumar Jayakumar, S. K. (2017). Online Student Feedback Analysis System with Sentiment Analysis. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(5), 8445–8451.
12. R.Udayakumar, R.Sugumar, Weighted Particle Swarm Optimization Algorithms and Power Management Strategies for Grid Hybrid Energy Systems, *MDPI Engineering Proceedings*, Volume 59, Issue 123, pp.1-11, December 2023.
13. Gawade, M. S. S., & Kumar, N. (2016). Three Effective Frameworks for semi-supervised feature selection. *International Journal of Research in Management & Technology*, 6(2), 107–110.
14. Jayakumar, D. T., & Naveenkumar, R. (2012). SDjoshi,“. *International Journal of Advanced Research in Computer Science and Software Engineering*,” *Int. J.*, 2(9), 62–70.
15. Jayakumar, N. (2014). Decreases and Discretization Concepts, tools for Predicting Student’s Performance. *Int. J. Eng. Sci. Innov. Technol*, 3(2), 7–15.
16. Jayakumar, N. (2015). Active storage framework leveraging processing capabilities of the embedded storage array.
17. Jayakumar, N., Bhardwaj, T., Pant, K., Joshi, S. D., & Patil, S. H. (2015). A Holistic Approach for Performance Analysis of Embedded Storage Array. *Int. J. Sci. Technol. Eng*, 1(12), 247–250.
18. Jayakumar, N., Iyer, M. S., Joshi, S. D., & Patil, S. H. (2016). A Mathematical Model in Support of Efficient Offloading for Active Storage Architectures. In *International Conference on Electronics, Electrical Engineering, Computer Science (ECS) : Innovation and Convergence (Vol. 2, p. 103)*.
19. Jayakumar, N., & Kulkarni, A. M. (2017). A Simple Measuring Model for Evaluating the Performance of Small Block Size Accesses in Lustre File System. *Engineering, Technology & Applied Science Research*, 7(6), 2313–2318.
20. Jayakumar, N., Singh, S., Patil, S. H., & Joshi, S. D. (2015). Evaluation Parameters of Infrastructure Resources Required for Integrating Parallel Computing Algorithm and Distributed File System. *IJSTE-Int. J. Sci. Technol. Eng*, 1(12), 251–254.
21. B Yamini, K Sudha, M Nalini, G Kavitha, R Siva Subramanian, R Sugumar, Predictive Modelling for Lung Cancer Detection using Machine Learning Techniques, 2023 8th International Conference on Communication and Electronics Systems (ICCES), pp.1220-1226.
22. Khare, A., & Jayakumar, N. (2017). Perspective Analysis Recommendation System in Machine Learning. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 6(2), 184–187.
23. Kumar, N., Angral, S., & Sharma, R. (2014). Integrating Intrusion Detection System with Network Monitoring. *International Journal of Scientific and Research Publications*, 4, 1–4.
24. Namdeo, J., & Jayakumar, N. (2014). Predicting Student's Performance Using Data Mining Technique with Rough Set Theory Concepts. *International Journal*, 2(2).
25. Naveenkumar, J., & Joshi, S. D. (2015). Evaluation of Active Storage System Realized Through Hadoop. *Int. J. Comput. Sci. Mob. Comput*, 4(12), 67–73.
26. Gaikwad Mahesh Parasharam, R. B. Dhumale.” Tumour Diagnosis Using Gray Level Co-occurrence Matrix and Artificial Neural Network”. *International Journal Of Advance Science and Technology VOL.29.No.5,(2020),pp.4255-4264*
27. KORE, MS AMRUTA, DR PROF DM THAKORE, and DRAK KADAM. "Innovation of E-commerce Fresh Agricultural Products Marketing Based on Big Internet Data Platform." (2019).
28. Kore, A. A., D. M. Thakore, and A. K. Kadam. "Unsupervised extraction of common product attributes from E-commerce websites by considering client suggestion." *Int. J. Innov. Technol. Explor. Eng* 8.11 (2019): 1199-1203



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details