Aggregate Customer Products Reviews on machine learning techniques using ontology process

S Suma Christal Mary1*, R Josphineleela2, K Sundaramoorthy3, E Thenmozhi4, K. Mamatha5

1,2,4Department of Information Technology, Panimalar Institute of Technology, Chennai, Tamil Nadu, India
3Department of Information Technology, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
5Assistant Professor, Department of Management Studies, Bharat institute of Technology, Hyderabad, Telangana, India

Email: *sumasheyalin@gmail.com

Abstract: Online customer reviews are the best advertisement and/or communication tools for product consistency. Consumers often prefer to do some research before agreeing to buy. Customer ratings are a significant source of knowledge for influencing consumer buying choices in the present. In this paper, we discussed SigmaCR’s aim to explore machine learning techniques and skills, natural language processing and information technology to create an intelligent device that analyses consumer product feedback via the Internet and provides comprehensive product summaries and consumer feelings. It would also be a very practical method for evaluating the power of many recent research projects, including ontology-based knowledge extraction and sentimental analysis.

Keywords: customer reviews, SigmaCR’s, natural language, sentimental analysis and product summaries.

1. Introduction

In natural language processing and information retrieval, the application of ontology and knowledge architecture principles is an emerging approach to computational linguistics and natural language processing methods for real-time applications. In natural language processing, scientists have often been difficult to obtain fair precision. [3] The SigmaCR’s project is a hybrid methodology in different science fields, including artificial language, machine learning, data processors and information engineering, to create a smart framework for the extraction and collection of online user feedback[4].

The methods of information extraction can differ, depending on the type of text, from system to system. Three kinds of documentation are available: open, semi-structured and structured. It is also very straightforward to derive information from these documents by setting the rules. Structured documents are in a predefined format. Half-structured documentation does not have a fixed format such as structured documentation but has known data types. [10] It helps certain information to be retrieved from records using the frameworks for the rule base. Free texts, for example, in structured or semi-structured documents, contain no framework within the document. It is not easy to obtain information from free texts because of random behavior [1]. Structured data or knowledge from unstructured texts is a daunting challenge and has required a great deal of study over the decades. The task of named entity extraction [2] was focused on early study activities. The research
emphasis of early history was on rule-based IE efforts [5]. Due to the necessary knowledge of linguistics and the information field, regulatory structures are easier to understand and create. For such domains or closed domains, they are effective in IE. [7] The data owner (DO), for saving local data storage expenses, normally stocks vast volumes of data in the cloud. However, the cloud service provider (CSP) may provide direct access to all customer data without any data security process [9].

A Soft logic for knowledge extraction may be viewed as statistical/machine learning-based approaches. Laws had complicated predicates, but as messages became more casual and organized, they became delicate. However, the mathematical approaches [15] had a smoother rationale when defining pertinent evidence. Statistical approaches [8] are algorithms that must be qualified for supervised learning. When they are supplied with greater training sets, they become more reliable, and this could be increased by simply extending the training set to maximize accuracy and recall [6]. Three main techniques have been established according to Sunita Sarawagi [3] statistical learning. It was,

- Generative trends (based on Hidden Markov Models)
- Custom versions (based on maximum entropy)
- Comprehensive conditions model (popularly called Conditional Random Fields)

2. Methods And Materials

The information collection system's functionality is to search on the Internet for product feedback to provide raw information for research, which is used to locate and retrieve feedback through web services and spiders.

2.1. Ontology Product:
It is the subsystem of information. A customized ontology reflects an awareness of the commodity domain. It includes goods, characteristics, parts, etc. The overall architecture has been shown in fig (1)

2.2. Language improvement ontology:
Knowledge Retrieval-While utilizing WorldNet, the linguistic enrichment module linguistically strengthens the lexical layer of ontology by discovering related terms for each ontological definition. The basis for matching related words was a modern semantic similarity measure explored in our work.

2.3. Classifier:
The classifier classifies the feedback into product groups referred to in Ontology. This method considers and derives product features from the ontology- and ontology-based jape laws.

2.4. Sentiment Analysis:
In this process, consumer emotions/feedback are defined and related to recognised characteristics.

2.5. Aggregation:
This module collects and aggregates data from function and feeling analyses to provide results.

2.6. Persistence and submission:
The persistence module is responsible for the permanent findings evaluated into NO-SQL data layer.

2.7. Model mining feature:
An experimental subsystem using techniques, including frequent itemset mining, depends on apriority and classifies new characteristics that are not used in ontology.
3. Deployment
This section explains the methods and instruments we used in SigmaCR implementation. Different management tools for handling code, building, and code consistency are included in the templates we used.

3.1. Types of processes:
We have implemented two different development processes in the planning process, which present two of the main process models for viewpoints. It is Scrum and the production guided by research. The following articles [13] give comprehensive information on the two process models we have used and the factors behind the choices made in selecting models.

3.2. Phase for Scrum Development:
Scrum is a product development methodology (Figure 2) that is an agile development type of software. It is an iterative and systematic approach to growth that proceeds with multiple iterations known as sprints. The team sets goals and expectations for the sprint at the outset of each sprint, [14] ushering in two-four weeks. The scrum team would then have meetings every day (usually starting simultaneously and every day) to discuss the project status. Day scrums normally last about 15 minutes.

3.3. Development powered by testing (TDD):
We used to test [11] led development as our development approach for Scrum sprints to use Scrum as a development process. The developer first writes a failed test case for a desired feature/task in a test-driven development. The developer first writes code to pass the checks effectively and then matches the code established by the team's coding principles & criteria [12].
Figure 2: Development process of Scrum

4. Result and Discussion

The enrichment ontology paradigm has been used in product and services and analysed with a Product Ontology. Product ontology included a mix of 15 items (including "Car," "bike" and "TV" as well as 40, including "acceleration" and "fuel efficiency" and "image quality" of these products as well as 40 features (including "power consumption"). The paradigm enhanced ontology. The overall accuracy, memory and F calculation were calculated in compliance with the following parameters. Precise = TP/(TP+FP) = Number of senses Predicted to be matching/Total number of senses predicted to match (as judged by humans) correctly. Recall = TP/P=Sense number Right / Sense number Properly + Sense number Properly not expected not to fit. = number of senses expected to correlate correctly / total number of corresponding senses. Table-1 represent predictions.

| System                  | Precision | Recall   | F Measure |
|-------------------------|-----------|----------|-----------|
| Naive system            | 0.720     | 0.947    | 0.818     |
| Wang et al.(Recall)[37] | 0.711     | 0.657    | 0.688     |
| Wang et al.(Laezecick-Chodorova)[37] | 0.711 | 0.845 | 0.772 |

4.1. Model of testing:

For OBNER (Ontology-based Named Entity Recognition), the Precision-Recall and F1 acts have been tested using an assessment corpus of 112 main attribute annotations. We did not take the anaphora resolution (word it refers to) into account in the assessment process because anaphora resolution is a different step where function and product references are calculated. During the assessment, there were 138 classes of Substance Ontology, 75 of which were Ten3.

Thirteen classes of calculation (like Time Measures etc.). The product classification comprised 50 classes, including big classes such as "Electronic consumer product." The system was compared to a naive system which only mentions morphological analysis and a mark function. Table-2 represents OBNER model analysis.
5. Conclusion
To create an intelligent framework to extract and aggregate user feedback on the Internet, SigmaCR's has incorporated the integrated methodology of numerous science fields including estimation linguistics, machine learning, information mining and knowledge engineering. We developed an "Ontology" to keep awareness of the domain of goods during this process. We developed an ontology-based knowledge extraction method to extract product from a research corpus written in natural language.' We have also developed a sentiment analysis framework that ties customer views on the characteristics and product comparisons that have been established. In this process, we have also implemented a model that enriches ontology lexically and a new semantic resemblance to the actual ontology tool collection. We should infer that although the whole scheme has been very good in its goal the test of the strength of many new research initiatives including ontology-based knowledge extraction and emotion analysis would serve as a very realistic implementation.

Reference
[1] Agrawal, R., Ho, H., Jacquenet, F., & Jacquenet, M. (2005, June). Mining information extraction rules from datasheets without linguistic parsing. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 510-520). Springer, Berlin, Heidelberg.
[2] Biega, J., Kuzey, E., & Suchanek, F. M. (2013, May). Inside YAGO2s: A transparent information extraction architecture. In Proceedings of the 22nd International Conference on World Wide Web (pp. 325-328).
[3] Sarawagi, S. (2008). Information extraction. Now Publishers Inc.
[4] "Message Understanding Conference," Wikipedia, 2013. [Online]. Available: http://en.wikipedia.org/wiki/Message_Understanding_Conference. [Accessed 08 April, 2013].
[5] Appelti, D. E., Hobbs, J. R., Bear, J., Israel, D., & Tyson, M. (1993, August). FASTUS: A finite-state processor for information extraction from real-world text. In IJCAI (Vol. 93, pp. 1172-1178).
[6] Precision and recall," Wikipedia, [Online]. Available: http://en.wikipedia.org/wiki/Precision_and_recall. [Accessed 19 April 2013].
[7] Grishman, R. (1997, July). Information extraction: Techniques and challenges. In International summer school on information extraction (pp. 10-27). Springer, Berlin, Heidelberg.
[8] "Illinois Part of Speech Tagger," Cognitive Computation Group, 2010. [Online]. Available: http://cogcomp.cs.illinois.edu/page/software_view/3. [Accessed 17 April, 2013].
[9] Subha, K., Reddy, P. K., Teja, D. S., Reddy, O. S., & Reddy, P. S. T. Data Sharing in Cloud Computing Based On Attribute-Based Encryption System.
[10] Pandey, P., & Soni, N. (2019, February). Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 320-322). IEEE.
[11] Li, S., & Li, Y. (2018, May). A Sentiment Analysis of Online Reviews Based on the Word Alignment Model: A Product Improvement Perspective. In 2018 2nd IEEE Advance Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 2226-2231). IEEE.
[12] Wladislav, S., Johannes, Z., Christian, W., André, K., & Madjid, F. (2018, December). Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Reviews. In 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 270-273).
IEEE.

[13] Gao, B. J., & Medjo, F. (2019, December). Statistical Correction of Average Customer Ratings for Product Ranking. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 6052-6054). IEEE.

[14] Kasturia, V., Sharma, S., & Sharma, S. (2020, January). Automatic Product Saleability Prediction using Sentiment Analysis on User Reviews. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 102-106). IEEE.

[15] Dharaiya, S., Soneji, B., Kakkad, D., & Tada, N. (2020, July). Generating Positive and Negative Sentiment Word Clouds from E-Commerce Product Reviews. In 2020 International Conference on Computational Performance Evaluation (ComPE) (pp. 459-463). IEEE.