IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 11, Issue 3, Aug 2023

Approaches for Extracting Emotions from Customer Reviews Using Data Mining and Natural Language Processing

MANCHIKA SRAVANTHI

ABSTRACT:

There are a few interconnected steps in automatic opinion recognition: locating the limits of an opinion's expression, detecting its polarity, and quantifying its intensity. The research in this field has come a long way, but it still tries to focus on one of the previously mentioned issues at a time. As more and more people share their opinions online, automated information extraction to back up summaries of consumer reviews has become necessary. Most of us have traditionally placed a high value on the opinions of those around us while making important personal decisions. Many internet forums and forums gather and distribute the insights of many people from every part of the world. Viewing the many different perspectives presented on the web is tedious and difficult to make understanding. To make a knowledgeable decision quickly and easily, Opinion Mining collects and analyses feedback from consumers from multiple sources across the Internet. The proposed system's primary function is to synthesize feature-based summaries of product reviews provided by online store customers.

Keywords: NLP, Datamining,Opinin mining, Customer reviews

Introduction

There has been a lot of research and development in recent years into techniques for doing this automatically. The Internet changed communication and interaction among people. They can now provide their opinions on products on merchant sites, as well as share ideas and interact with others, via the use of blogs and discussion forums. User-generated content refers to the information and material published online by users, as opposed to those posted by website owners. User-generated content is now widely acknowledged to provide useful information that may be used for various purposes. down this study, we zero down on product reviews written by actual consumers. Insightful user feedback on items and services can be found in reviews. Potential consumers consult them to learn the tales of others who have utilised the goods. Manufacturers also utilise them to research potential issues with their products and get competition knowledge [1]. E-commerce is growing rapidly, which means more and more individuals will be making purchases through the Internet.

Assistant professor

Holy Mary Institute of Technology and Science

It is now usual practise for online retailers to provide their consumers the option of writing reviews or leaving feedback on the things they have purchased. As the general public gains familiarity with the Internet, more and more of them will be able to and will write evaluations. Because of this, the total number of reviews for any given product increases rapidly.

Human Language Technology (HLT) includes numerous approaches and specialisations. Natural language processing (NLP), speech recognition, automated translation, text creation, and text mining are all part of this broader category. Several decades have passed since NLP's inception. It has created a number of techniques that are often linguistically inspired; for example, it syntactically parses text with the use of information from a formal grammar and a lexicon, and then uses the knowledge gleaned from this process to extract information about what was said.

Any language that is used by humans in their everyday interactions is considered a "Natural Language" (NL), as opposed to a "Man-Made Language" like a programming language. All efforts to use computers to process natural language fall under the umbrella term of Natural Language Processing (NLP). Natural language processing (NLP) is typically used in a way that does not include speech, hence SNLP is the word used to encompass both speech and other forms of natural language processing.

Synthesis of speech, recognition of spoken language, generation of natural language, understanding of spoken language, and machine translation were all components of NLP [36]. Recently, researchers in the fields of Information Retrieval and Computational Linguistics have started concentrating on the opinions expressed in documents rather than the topics they cover. Opinion Mining is a multidisciplinary discipline that borrows heavily from the fields of Information Retrieval and computational linguistics.

Facts and opinions are the two primary types of textual information that exist in the world. Statements that can be proven to be true concerning real-world things and occurrences are called facts. Statements of opinion represent the feelings or perspectives of the speaker on the entities or events being discussed. Information retrieval, web searching, and several other text mining and natural language processing jobs are all examples of areas where the majority of previous research on text information processing has been (almost completely) focused on mining and retrieval of factual information. Prior to the past few years, there was a dearth of research into how opinions are processed. Opinions are so valuable that people often seek them out before making significant decisions. This holds true not only for people, but also for businesses.

The fact that there was already writing containing opinions before the advent of the Internet explains why so little has been studied in this area. Without the Internet, consumers would have to put in a lot more time and energy into making a single purchase decision, while businesses would have to scour the internet for the thoughts of hundreds of thousands of people from dozens of communities. Internet users have access to a wealth of potentially useful data; nevertheless, it is ultimately up to the individual to sift through all of the available data and find the specific pieces of information they need. User-generated content refers to any and all of the web's preexisting information that can be altered by users themselves.

It is still a major burden for a company to regularly update all of their products, which requires them to find and monitor online opinion sources. Many online resources exist for a single product, but it is up to the consumer to sift through them and find the relevant data. An automated approach that facilitates more effective opinion mining is required to address the overall issue. Feelings analysis is another name for opinion mining. [2]

Literature review

Academic and business sectors have shown growing interest in opinion mining in reviews during the past few years. The majority of the prior research, however, has concentrated on summarising and extracting opinions from reviews [3, 4, 5, 6]. Opinion summarising relies heavily on the proper categorisation of feature expressions, which are essentially domain synonyms. It's incredibly time-consuming and tedious for humans to categorise the hundreds of feature expressions that can be discovered from text for an opinion mining application. We need some mechanical help here. The most obvious method for addressing this issue is unsupervised learning, often known as clustering. The parallels

Similarity in distribution is the foundation for most clustering measures [8,9,10,11,12,13,14,15]. Topic modelling was also applied in recent publications [16,17].

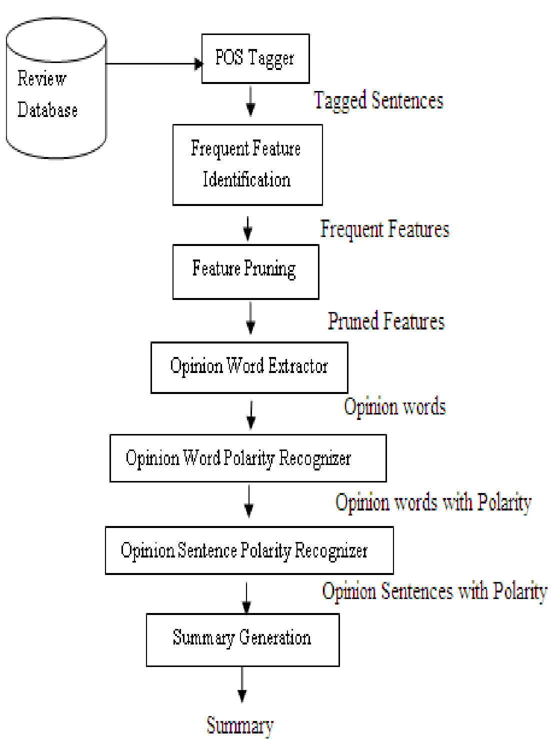
Finding people's feelings and thoughts about a subject, as well as their thoughts on its features, is the goal of opinion mining [18]. Recent studies have concentrated on discovering consumer feedback in online product reviews ([19]). The main difference of reviews is that they each focus on one product for analysis. Few people engage with one another, and some of the content is irrelevant. Online comments and conversations, however, are an exception to this rule. Finding terms like "great," "amazing," "wonderful," "bad," and "poor" that convey an opinion (or mood) was the starting point for opinion mining studies. The process of mining such words and determining their semantic orientations (i.e. positive or negative) has been the focus of numerous scholars. The authors of [23] discovered many linguistic rules that might be utilised to isolate opinion words and their polarities in a massive corpus. In [21, 26], [30], and other places, this technique has been used, expanded, and

ISSN 2347–3657
Volume 11, Issue 3, Aug 2023

refined. A bootstrapping method is proposed in [24, 27] to identify related words in WordNet (http://wordnet.princeton.edu/) by using a small collection of given seed opinion words.

Sentiment classification of product reviews at the document level is the next big thing [20, 29, 31]. The goal here is to determine whether a given review is favourable or negative, and to label it as such. This may be for anything from a movie to a camera to a car. Sentence-level sentiment classification, wherein each sentence is categorised as reflecting a favourable or negative view, has also been explored by a number of researchers [27, 32, 33]. Opinion mining and summarization using features is proposed as a model in [24,28]. The opinion mining problem is better defined by this paradigm. It explains how to extract useful data from unstructured texts and provides a summary of opinions in an organised style. In [22, 25], the issue of opinion mining from comparison sentences is presented. User reviews are frequently published on various online platforms, including retailer sites, forums, blogs, and message boards. User-created content or user-made media describes this type of content. These user-generated contents or media can be mined with one of three mining tasks. 'Web Data Mining Exploring Hyperlinks, Contents, and Usage Data' by Bing Liu [32] explains these three mining jobs.

To investigate the issue of feature-based opinion summary of online product evaluations, Minqing Hu and Bing Liu [33] presented a method. There are three steps to this task, which begin with receiving a set of customer reviews for a product and ending with a summary based on the information found therein:



(1) identifying product features about which customers have expressed opinions (called product features); (2) for each feature, identifying review sentences that give positive or negative opinions; and (3) synthesising the information found. Based on data mining and NLP techniques, they proposed a set of methods for mining and summarising product reviews.

An efficient approach for determining the semantic orientations of reviewers' judgements on product features was proposed by Xiaowen Ding, Bing Liu, and Philip S. Yu [34]. It solves the issues of (1) using opinion words whose semantic orientations vary depending on context and (2) aggregating several opinion words inside the same sentence, both of which are significant shortcomings of the current techniques. A comprehensive method is proposed that can reliably

In honour of Helena Ahonen and Oskari Heinonen A approach for extracting descriptive phrases from digital document collections was proposed by Mika Klemettinen and A. Inkeri Verkamo [37]. The basis for text mining framework. The data mining approach that will be used is predicated on generic episodes and rules for generic episodes. (1) Overarching structure of text mining; data gathered may contain time-ordered information or observations. The input is textual data, and the output is a description of recurrent phenomena in the database. Sequential data can be analysed using a modified version of association rules known as "episode rules," along with the concept of "episode." Two, data preparation accounts for about 80% of the total effects. Experiments with real-world data validated the usefulness of our method by demonstrating that the episodes and episode rules they generate may be used to distinguish between different types of texts. Results trimming and weighting rely heavily on both pre- and post-processing.

Bing Liu, Minqing Hu, and Junsheng Cheng [35] zeroed emphasis on one subset of opinion sources: product reviews written by actual buyers. When comparing consumer feedback on numerous products, we suggested a revolutionary visual analysis technique. We developed a supervised pattern discovery technique to automatically recognise product categories to aid in visual analysis.format reviews (2) featuring pros and cons from Pros and Cons. It is far more efficient than manual tagging, and a user-friendly interface is provided so that the analyst may interactively correct any errors made by the automatic system.

ISSN 2347–3657

Volume 11, Issue 3, Aug 2023

Review Database: Product reviews from various online review sites have been gathered to form the review database. Sites like www.ebay.com, www.eopinion.com, www.imdb.com, and www.amazon.com can all be mined for information. Some review data sets can be found at http://www.cs.uic.edu/liub.

POS Tagger: Retail Point-of-Sale (POS) Tagging is performed with a POS Tagger. Tagging each word in a sentence as a noun, verb, pronoun, adverb, adjective, or some other lexical class marker is known as part-of-speech (POS) tagging. Stanford Tagger[5] is one example of a POS tagger that can be used for this purpose.. For example consider following sentence-

"The battery life is not long enough."

After tagging this sentence with Stanford Tagger we get following output.

Battery/NNP life/NN of/IN this/DT handset/NN is/VBZ good/JJ

Frequent Feature Identification: Repeated features are discovered after POS tagging. A product's frequent features are those about which many individuals have voiced an opinion, whereas its infrequent features are those about which few voices have been heard. Reviewer comments and ratings on specific product features are obtained in the previous phase, POS tagging. This phase involves classifying product attributes as either common or rare. In this system, we will simply look at the most common characteristics. In order to identify all recurring characteristics, association mining might be used. Association mining is used mostly as a result of an observation. A customer review will often include issues that have nothing to do with the benefits of the product. The tales of various clients tend to vary. However, when discussing particular product aspects, they typically use the same vocabulary. Therefore, it is appropriate to employ association mining to identify common item sets, as these sets are likely product features.

Feature Pruning: Association mining can produce a large number of candidate frequent features, but not all of them are legitimate. Those improbable characteristics can be cut away by one of two pruning methods.

Compactness pruning: We term features with at least two words "feature phrases," and this process eliminates useless feature phrases. The placement of items (or words) in a sentence is irrelevant to the association mining algorithm. But in a sentence, phrases formed by words that come together in a particular order are more likely to convey the intended meaning. It follows that not all of the terms that appear often in feature output from association mining constitute actual features. The goal of compactness pruning is to eliminate candidate features whose words do not occur in a predictable sequence.

Redundancy pruning: In this step, we focus on removing redundant features that contain single words.

Opinion Word Extractor: Opinion Word Extractor is used to pull out opinion words once pruning has been completed. An opinion sentence is a sentence that describes a product and includes at least one opinion word. The following is how the opinion word extractor finds opinion words:

Each sentence in the database is checked for the presence of opinion words, and if any are found, they are isolated as adjectives.

It uses the nearest adjective to describe each feature.

Opinion Words Polarity Recognizer: Each opinion word in a sentence is identified and assigned a polarity. The steps to identify an opinion word's polarity are as follows:

Assign the polarity from the database if words indicating an opinion are found.

Find a synonym for the opinion term if it is not in the database. Assign the polarity of the synonym to the opinion word, and save the result if a synonym exists.

If a term of opinion neither has a synonym in the database nor is it otherwise represented there, its antonym should be sought out. If an antonym exists, the opinion word should be given the opposite polarity of the antonym and added to the database.

.

Opinion Sentence Polarity Recognizer: Opinion Phrase Polarity Recognizer takes note of the polarity of each review's individual opinion sentences. It adheres to the procedures listed below for determining the polarity of

ISSN 2347–3657

Volume 11, Issue 3, Aug 2023

an opinion:

If an opinion statement contains an odd number of negation words, you should provide it an opinion word with the opposite polarity.

Assign the polarity of the opinion word to the sentence if there are an even number of negation words.

Assign the polarity of the opinion word to the opinion sentence if the negation word is absent.

Feature-Based Summary Review GenerationThe steps that follow are used to create the final feature-based review summary:

According to the opinion sentences' orientations, relevant opinion sentences are categorised as either positive or negative for each feature that has been detected. To demonstrate how many reviews are positive or negative towards the feature, a count is calculated.

According on how frequently they appear in the reviews, all features are ranked. Feature phrases are displayed before single word features because users typically find phrases to be more appealing. Rankings of other kinds are also feasible**.**

Conclusion

In this research, we present a system that utilises data mining and natural language processing (NLP) methods to extract and summarise client feedback. The goal is to summarise many customers' experiences with a product supplied online by focusing on its particular features. Both regular consumers and the businesses that make the products would benefit greatly from an overview of the reviews.

We plan to further advance methods in future work, as well as address the remaining problems we've identified above, such as pronoun resolution, opinion strength assessment, and the study of adverb, verb, and noun expressions of opinion. Finally, we'll also be researching review monitoring.

References

[1] K. Dave, S. Lawrence & D. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. WWW'2003.

[2] Bing Liu. "Opinion Mining." Invited contribution to Encyclopedia of Database Systems, 2008.

[3]. M. Hu & B. Liu. Mining and summarizing customer reviews. KDD'2004.

[4]. A. Ntoulas, M. Najork, M. Manasse & D. Fetterly. Detecting Spam Web Pages through Content Analysis. WWW'2006.

[5]. B. Pang, L. Lee & S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. EMNLP'2002.

[6]. M. Sahami and S. Dumais and D. Heckerman and E. Horvitz. A Bayesian Approach to Filtering Junk {E}-Mail. AAAI Technical Report WS-98-05, 1998.

[7] Liu B, Hu M, and Cheng J. Opinion Observer: Analyzing and Comparing Opinions on the Web. in Proceedings of WWW. 2005.342-351.

[8] Bollegala D, Matsuo Y, and Ishizuka M. Measuring semantic similarity between words using web search engines. In Proceedings of WWW. 2007.757-766

[9] Chen H, Lin M, and Wei Y. Novel association measures using web search with double checking. in ACL. 2006.1016

[10] Lin D. Automatic retrieval and clustering of similar words. 1998: Proceedings of ACL.768-774

[11] Lin D and Wu X. Phrase clustering for discriminative learning. in Proceedings of ACL. 2009.1030-1038

[12] Pantel P, Crestan E, Borkovsky A, Popescu A, and Vyas V. Web-scale distributional similarity and entity set expansion. in Proceedings of EMNLP. 2009.938-947.

[13] Pereira F, Tishby N, and Lee L. Distributional clustering of English words. in Proceedings of ACL. 1993.183-190

[14] Resnik P. Using information content to evaluate semantic similarity in a taxonomy. in IJCAI. 1995.448-453

[15] Sahami M and Heilman T. A web-based kernel function for measuring the similarity of short text snippets. In Proceedings of WWW. 2006.377-386

[16] Guo H, Zhu H, Guo Z, Zhang X, and Su Z. Product feature categorization with multilevel latent semantic association. In Proceedings of CIKM. 2009.1087-1096

[17] Titov I and McDonald R. Modeling online reviews with multi-grain topic models. in WWW. 2008.111-120

[18] Hu M. and B. Liu. 2004. Mining and Summarizing Customer Reviews. Proc. of KDD

[19] Liu B. 2010. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing N. Indurkhya and F. J. Damerau.

[20] Dave, D., Lawrence, A., and Pennock, D. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. Proceedings of International World Wide Web Conference (WWW'03), 2003.

[21] Ding, X., Liu, B. and Yu, P. A Holistic Lexicon-

Based Approach to Opinion Mining. Proceedings of the first
ACM International Conference on Web search and Data Mining (WSDM'08), 2008.
[22] Ganapathibhotla, G. and Liu, B. Identifying Preferred Entities in Comparative Sentences. To appear in Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08), 2008.
[23] Hatzivassiloglou, V. and McKeown, K. Predicting the Semantic Orientation of Adjectives. ACLEACL' 97, 1997.
[24] Hu, M and Liu, B. Mining and Summarizing Customer Reviews. Proceedings of ACM SIGKDD International
Conference on Knowledge Discovery and Data Mining (KDD'04), 2004.
[25] Jindal, N. and Liu, B. Mining Comparative Sentences and Relations. Proceedings of National Conference on Artificial Intelligence (AAAI'06), 2006.
[26] Kanayama, H. and Nasukawa, T. Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP'06), 2006.
[27] Kim, S. and Hovy, E. Determining the Sentiment of Opinions. Proceedings of the 20th International Conference on Computational Linguistics (COLING'04), 2004.
[28] Liu, B., Hu, M. and Cheng, J. Opinion Observer: Analyzing and Comparing Opinions on the Web. Proceedings
of International World Wide Web Conference (WWW'05), 2005.
International Journal of Computational Intelligence and Information Security, July 2012 Vol. 3, No. 6 ISSN: 1837-7823
58
[29] Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), 2002.
[30] Popescu, A.-M. and Etzioni, O. Extracting Product Features and Opinions from Reviews. Proceedings of the
2005 Conference on Empirical Methods in Natural Language Processing (EMNLP'05), 2005.
[31] Turney, P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of
Reviews. ACL'02, 2002.
[32] Minqing Hu and Bing Liu Mining Opinion Features in Customer Reviews KDD'04, August 22–25, 2004
[33] Bing Liu,Minqing Hu,Junsheng Cheng, Opinion Observer: Analyzing and Comparing Opinions on the Web,WWW 2005, May 10-14, 2005
[34] Bing Liu, Web Data Mining Exploring Hyperlinks, Contents, and Usage Data', Springer, 2007.
[35] Standard Tagger Version 3.0 (http://www-nlp.standfold.edu/software/tagger.html)
[36] Rune Saetre, "GeneTUC Natural Language Understanding in Medical Text", Thesis of Department of
Computer Science and Technology, June 2006, Norwegian University of Science and Technology, (NTNU),
Norway.
[37] Helena Ahonen, Oskari Heinonen Mika Klemettinen, A. Inkeri Verkamo "Apply Data Mining Techniques for
descriptive Pharse extraction in digital document collections", 24 april 1998, IEEE International Forum,
Santa Barbara, CA.