School of Informatics

Informatics Research Review
Aspect-based Sentiment Analysis on Online Review Data

Taeseung Hahn
December 2021

Abstract

The field of sentiment analysis, which extracts, analyses and aggregates sentiments from text, has received a lot of attention over the past few years. Due to the nature of online review data, aspect-based sentiment analysis is essential to comprehensively understand and utilise users' opinions and thoughts. In this paper, we examine the overall process of aspect-based sentiment analysis and review various approaches that jointly model the aspect and sentiment.

Date: Friday 17^{th} December 2021

Supervisor: Pavlos Andreadis

Contents

1 Introduction

With the rapid development of e-commerce, the amount of online review data for products and services is also rapidly increasing. People leave reviews and ratings after watching movies, visiting restaurants, or purchasing products. They are made not only through the e-commerce platform, but also through microblogs and community websites. These reviews have important implications for consumers and sellers of goods and services, as well as e-commerce platforms.

In terms of the customers, they are interested in what products are popular, why they are popular, and what are the advantages and disadvantages of each. They also want to make reasonable decisions by comparing and evaluating several products by looking at reviews from other customers. Therefore, e-commerce platforms can improve the quality of information they provide to customers by analysing customer-generated reviews, extracting opinions or emotions about products, and showing them to other customers. When it comes to the product sellers, online reviews are also very important sources because they play an essential role in shaping customers' awareness and perceptions about products [1]. Therefore, it can affect product sales in the future, and thus it is also fundamental for sellers to analyse them to improve their quality of goods and services.

However, due to the vast amount of review data and the nature of text data without a format, it is inefficient or impossible to manually analyse these data. Therefore, studies on sentiment analysis that can systematically extract and analyse sentiment from these data have been conducted for years. Another notable property of review data is that different aspects of a product or service can be evaluated separately. For example, when reviewing a restaurant, you can give different opinions on each aspect, such as positive evaluation of taste and decoration, and negative evaluation of waiting time and interior design of the restaurant. Therefore, in order to properly extract customers' opinions from the review data, opinions and emotions for each aspect must be separately extracted. Accordingly, an aspect-based sentiment analysis on online reviews has been researched as a sub-area of sentiment analysis for recent years.

In the research area of aspect-based sentiment analysis, a variety of models have been proposed that differ in the need for labelled data, adaptability between domains, and the level of aspect detection. However, there is currently no superior model in all respects, and direct comparison between every model is difficult due to the different metrics that measure the performance of each model. The goal of this review is to identify what specific challenges each model has been proposed to overcome, what assumptions they have, what are the advantages and disadvantages of each model.

This review is organized as follows. First, we investigate ordinary sentiment analysis framework. Then, we present the characteristics of online review data and the limitation of ordinary sentiment analysis. Subsequently, we briefly discuss sub-tasks of aspect-based sentiment analysis: Aspect Detection and Identification; Sentiment Classification and Analysis. Afterwards, we review a variety of joint models and their assumptions and advantages/disadvantages. Next, we compare the performance of presented models. Finally, we summarise and conclude the review in section 4.

2 Background

2.1 Outline of Sentiment Analysis

Sentiment Analysis, which is also called as Opinion Mining, is a sub-research area of Natural Language Processing aimed at systematically identifying, extracting, quantifying, and studying emotions expressed in a text [2]. Because the field of emotional analysis exists at the intersection of natural language processing, information retrieval, and artificial intelligence [3], various terms are widely used for similar concepts. The term ‘sentiment analysis’ is be used in this review, where ‘sentiment’ refers to a wide range of emotions such as delight, sadness, happiness, and irritation. In order to simplify the problem of identifying detailed sentiment, sentiment polarity consisting of positive, negative, and neutral is frequently used instead.

In prior studies, sentiment analysis has been categorized into three levels depending on whether the level of analysis is a whole document, a sentence, or an aspect [4, 5, 6]. In this context, ‘aspect’ refers to the detailed attributes that the target, i.e. entity, of the review may have. For example, in the text of ‘The design of the bag is perfect, but it is a little heavy’, ‘design’ and ‘weight’ can be two different aspects of the reviewed entity ‘bag’.

Sentiment analysis can be leveraged for a variety of purposes. Business organizations can utilise it to extract customers’ emotions from various sources such as customer feedback and customer centre messages. Furthermore, it can also be used to check whether the response of company-related news and comments on social media promotions is positive or negative. It is also frequently utilised to improve the quality of products and services by reflecting customers’ opinions.

2.2 Online Review Data

More and more Internet users are expressing their opinions and thoughts through microblogs and reviews on e-commerce platforms. The reviews are online user-generated data from various anonymous users, which essentially has the following characteristics.

1. Massive and Noisy

Due to the characteristic that it is generated simultaneously through the Internet from a large number of users, review data is inherently noisy and the amount is massive. Therefore, it is almost impossible to manually analyze these data, and even if possible, it is very inefficient.

2. Irregular Format

   Online review data can have various forms. Review data may contain both types of information or only one of the following: the numerical rating and the review text. The numerical rating is a quantitative summary of the reviewer's experience, attitude, opinion, or emotion about a product or service, usually expressed in numbers. The review text is an open-ended textual description of the reviewer's opinion on the product or service [7, 8]. In the case of numerical rating, only one rating for the overall experience may exist, or each rating for a detailed aspect may exist separately. In this review, studies that analyze only numbers will be excluded. This review focuses on approaches analyzing ratings and text data together, or extracting and analyzing customer opinions and sentiments from text.

3. Rich information on various aspects and the sentiment polarization of the user [9].

   In general, one review does not deal with only one aspect of the product. For example, when a customer writes a review of a hotel, one can write an experience of various aspects such as 'price', 'location' and 'room condition' along with overall satisfaction with the hotel. Satisfaction with each aspect may not be the same, and text expressions for representing each satisfaction may also be used in various ways.

3 Literature Reviews

3.1 Aspect-based Sentiment Analysis

In earlier works, many studies only focused on extracting topics from text data. However, the impact of a reviews can vary depending on the sentiments from them. For example, even if reviews for a particular product surge, the impact of the reviews on future sales will completely vary depending on whether the sentiment in the review is positive or negative. To capture positive and negative sentiments from text data many studies were conducted [2, 3, 4, 5, 6].

Another part to consider when analyzing sentiments for review data is that one review usually deals with several aspects of the product, and sentiments for each aspect may be different. For example, when reviewing a specific movie, people can evaluate the aspects of 'Story', 'Running Time', 'Visual Beauty' and 'Sounds', and emotions for each aspects may contradict each other. As another example, reviews of restaurants may include evaluations of aspecs for 'taste', 'decoration', 'waiting time' and 'interior design', while reviews of mobile phones may include evaluations of aspects for 'screen size', 'battery', 'design' and 'price'. Therefore, it can be useful to extract not only the overall sentiment but also different emotions for each aspects in the review.

Aspect-based sentiment Analysis takes as input a set of user reviews for a specific product or service and produces a set of relevant aspects, the aggregated sentiment for each aspect, and supporting textual evidence [10]. Standard aspect-based sentiment analysis consists of following two main tasks:

1. Aspect Detection

The goal of aspect detection is to find the set of relevant aspects for a rated entity and extract all textual mentions that are associated with each.

2. Sentiment Classification

   The goal of sentiment classification is to identify and aggregate sentiment over each aspect to provide the user with an average numeric or symbolic rating.

In the early works, many approaches conduct these two tasks separately. In contrast, many recent studies have shown attempts to perform two tasks simultaneously with an integrated model, and this review focuses on this approach.

3.2 Aspect Detection and Identification

Aspect Detection or Aspect Identification is to find an aspect set for an entity evaluated in a review and extract all text references related to each. According to Mukherjee et al., Aspect Detection, which extracts aspects of the entity mentioned in the review, again consists of two sub-tasks [?]. The first sub-task is to extract aspect terms from the opinion corpus, and the second sub-task is to cluster synonymous aspect terms into categories in which each category represents a single aspect.

Aspect Detection previously studied can be classified into two types. The first type is an approach that extracts only aspect without clustering, and the second type is an approach that extracts aspect using statistical topic modeling and clusters it with unsupervised model. Both approaches are useful and the standard models for this task are PLSA and LDA. However, there is a problem in LDA and PLSA that these models tends to only extract global topics rather than detailed local topics [11]. For example, if the topic is extracted from the review of the restaurant, it tends to be extracted as the topic of 'London Restaurant' and 'Paris Restaurant' rather than the aforementioned 'taste', 'decoration', 'waiting time' and 'interior design'. Since these global topics are not what the user wants to evaluate, the need for an approach capable of modeling local topics has been raised.

To solve this problem, Titov and McDonald presents a model called MG-LDA that can model not only the global topics but also the local topics, aspects, of the entity [11]. This model is based on the expansion of standard models such as LDA and PLSA. MG-LDA is allowed to generate terms from either a global topic or a local topic. They are chosen based on the document level context and sliding window context over text, respectively. Local topics more faithfully model aspects evaluated throughout the review corpus. Since this model is an unsupervised model, it has the advantage of having few restrictions on use, and it has been improved in the future to become the basis for the MAS model.

On the other hand, Jo proposed Sentence-LDA(S-LDA), a stochastic generation model that assumes that all words in a single sentence are generated in one aspect [12]. S-LDA is proposed to overcome the property of LDA that positions of individual words are neglected, which is not always appropriate. They noted the fact that words about an aspect tend to co-occur within close proximity to one another. This model became the basis for the Aspect and Sentiment Unification(ASUM) model that jointly discovers a pair of aspects and sentiment in a sentence.

3.3 Sentiment Classification and Analysis

The second task, primarily called sentiment Classification, is to identify whether the semantic orientation of the text given in the each aspect is positive, negative, or neutral. The models for this task are largely categorized into a supervised model and a non-supervised model depending on whether a positive/negative labelled corpus is required and the sentiment classification by supervised learning has been studied in various ways [13, 14, 15, 16, 17, 18]. However, supervised learning approach has drawbacks in that labelled corpus itself cannot be easily obtained or is costly.

One big challenge in sentiment classification is that the same sentiment textual expression can have completely different meanings in different domains or topics. For example, the adjective 'unpredictable' can be used in a negative sense, such as 'unpredictable steering' in car reviews, but can also be used in a positive sense, such as 'unpredictable plot' in movie reviews [19]. In other words, sentiment polarities is aspect-dependent and domain-dependent which is supported by the report from A. Aue and M. Gamon. They reported that the accuracy performance of in-domain Support Vector Machines(SVMs) classifier trained on the movie review data dropped from 90.45 percent to 70.29 and 61.36 percent respectively when it is directly tested on book review and product support services data [20].

3.4 Joint Aspect Detection and Sentiment Analysis

Due to the dependency between sentiment polarities and aspect, there have been various attempts to model aspect and sentiment simultaneously in consideration of interplay of them.

Mei et al. proposed a Topic-Sentiment Mixture(TSM) model that simultaneously processes background words, sides, and emotions [21]. However, the model does not directly model sentiments and requires post-processing to calculate the positive/negative coverage in a document in order to identify its polarity [22].

As mentioned in Section 2.2, some review data may have ratings for evaluation aspects along with text at the same time. In consideration of these characteristics, Titov and Mcdonald proposed a new statistical model called Multi-Aspect Sentiment(MAS) that utilizes additional information by adding aspect rates as an observation variable [10]. The model makes the following two assumptions. First, the ratable aspects generally represents coherent topics that can be potentially found in the cooccurrence information in the text. Secondly, they assume that the most predictable feature of the aspect rating are features derived from the text segment discussing the corresponding aspect [10]. The assumption that at least one rated aspect should exist in one review is somewhat impractical. In addition, MAS model lacks the flexibility to apply to other domains because it requires additional data, aspect ratings. In other words, the model cannot be said to be a fully unsupervised model, but has some limitations as a supervised model.

On the other hand, Lin and He proposed Joint Sentiment-Topic(JST) model which is a fully unsupervised model that does not require labelled data such as labeled corpus or user rating at all [22]. This model is an extended form of standard LDA, which is added sentiment layer. It is a joint model can model and detect sentiment and topics simultaneously. However, this work has limitations in that it did sentiment analysis at the document level without considering several detailed aspects within one review. They overcame the limitations of JST in separate aspects and opinions based on sentiment vocaburary [23].

Meanwhile, Jo and Oh proposed Sentence-LDA(S-LDA), a stochastic generation model that assumes that all words in a single sentence are generated in one aspect [12]. They also extend it to Aspect and Sentiment Unified Model(ASUM) to modelling the sentiments about a variety of aspects. ASUM finds a pair of aspect, sentiment called senti-aspect. ASUM is as same with JST in terms of that the sentiment is integrated with a topic in a single language model.

However, there is a difference between them in that ASUM has a constraints that the words in a single sentence to come from the same language model. Threfore, compare to JST, each of the inferred language models is more focused on the regional occurrences of the words in a document in ASUM.

JIANG et al. tried hybrid approach with Max Entropy. They proposed MaxEnt-LDA, a hybrid LDA model that jointly discover both aspects and aspect-specific opinion words [19]. This model is a topic modeling approach that introduces MaxEntropy, assuming that one sentence conveys one opinion about one aspect.

Kim et al. noted that the levels of aspects of interest of people seeking information from reviews are not the same [9]. For example, some users are interested in general aspects such as 'screen size' and 'battery capacity' in laptop reviews, while others may be interested in more detailed aspects such as 'CPU frequency' and 'cash size'. From this point of view, the study emphasized the need for a hierarchical structure of sentiment analysis. They proposed Hierarchical Aspect-Sentiment Model(HASM) using tree structure to explain the hierarchy of aspect-based sentiments. In this model, each aspect or sentiment polarity is modeled as a distribution of words.

He et al. noticed the attribute that online data is continuously generated by a large number of users [24]. They proposed a Dynamic Joint Sentiment-Topic(dJST) model that can detect and track the relationship between aspect and sentiment that can change over time. They tried to capture the most recent textual expressions represent sentiment by dynamically updating the sentiment analysis model according to dynamically changing data.

Tang et al. proposed Joint Aspect-Based Sentiment Topic(JABST) model [25]. JABST extracts multi-grained aspects simliar to MG-LDA. In addition, it also jointly extracts opinions through modelling aspects, opinions, sentiment polarities and granularities at the same time. In addition, they also proposed MaxEnt-JABST model which is a supervised model, combining JABST and max entropy. When there is labelled data available, MaxEnt-JABST can improve the accuracy and performance of aspect extraction using the additional information.

3.5 Evaluation of the Models

Author	Model	Review Domain	Measures	Performance
Mei et al.	TSM	Weblogs	KL-Div.	21/19 (Pog/Neg)
Titov and Mcdonald	MAS	Hotel	Avg. Prec.	74.5% - 87.6%
Lin and He	JST	Movie	Accuracy	78%
Jo and Oh	ASUM	Product/Restaurant	Accuracy	84% - 86%
JIANG et al.	MaxEnt-LDA	Restaurant/Hotel	nDCG@5	76.4% - 82%
Kim et al.	HASM	Laptops/DSLR	Accuracy	85%
Tang et al	JABST	Product/Restaurant	Accuracy	80%-81%
Tang et al	MaxEnt-JABST	Product/Restaurant	Accuracy	81%-83%

It is not easy to make a direct comprehensive comparison across all models due to the various evaluation metrics used in different approaches. This is because the selection of evaluation metrics would depend on the domain that sentiment analysis is applied on. In addition, various approaches define emotions in different ways, which makes the appropriate evaluation metrics for each model different. For example, some methods classify sentiment in binary or ternary, whereas others use a 0-10 rating. However, despite the difficulty of direct comparison across various models, it can be seen that LDA-expanded models generally show good performance.

4 Summary & Conclusion

In this review, we firstly examined the outline of standard sentiment analysis and the characteristics of online review data. Due to the nature of the review data, one review may have evaluation of several aspects of the reviewed entity, which cannot be extracted by ordinary sentiment analysis model. Therefore, we could identify the need for Aspect-based sentimental analysis.

Aspect-based sentimental analysis consists of two tasks: Aspect Detection and Sentiment Classification. We briefly investigated both tasks and reviewed a variety models for joint aspect detection and sentiment analysis. Because the goal of this review was identify the assumptions and advantages/disadvantages of each model, we focused on the assumptions behind the model and similarities and differences between models.

It was clear that models has its own characteristics and limitations according to the required data (e.g. text, text with ratings, sentiment labelled data), the text level of analysis (e.g. whole document, sentence, aspect), the level of extraction of topics (e.g. global/local topic) simplifying assumptions (e.g. one sentence has only one aspect, at least one rated aspect should exist in one review).

We can also confirm that model have been improving in various aspects over time, but models that are superior in all aspects do not currently appear to exist. In addition, the fact that it is difficult to compare model performances because the domains to which each model is applied are different and the evaluation measure used are different remains a challenge.

References

[1] Xinxin Li and Lorin M. Hitt. Self-Selection and Information Role of Online Product Reviews. *https://doi.org/10.1287/isre.1070.0154*, 19(4):456–474, 7 2008.

[2] Erik Cambria, Soujanya Poria, Alexander Gelbukh, and Mike Thelwall. Sentiment Analysis Is a Big Suitcase. *undefined*, 32(6):74–80, 11 2017.

[3] B Pang, L Lee Comput. Linguist, and undefined 2009. Opinion mining and sentiment analysis. *Citeseer*, 2(2):1–135, 2008.

[4] Abinash Tripathy, Abhishek Anand, and Santanu Kumar Rath. Document-level sentiment classification using hybrid machine learning approach. *Knowledge and Information Systems*, 53(3):805–831, 12 2017.

[5] Bing Liu. Sentiment analysis and subjectivity. *Handbook of natural language*, 2010.

[6] Kim Schouten and Flavius Frasincar. Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830, 3 2016.

[7] Susan M. Mudambi and David Schuff. What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly: Management Information Systems*, 34(1):185–200, 3 2010.

[8] Do Hyung Park and Sara Kim. The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7(4):399–410, 12 2008.

[9] Suin Kim, Jianwen Zhang, Zheng Chen, Alice Oh, and Shixia Liu. A Hierarchical Aspect-Sentiment Model for Online Reviews. *Proceedings of the AAAI Conference on Artificial Intelligence*, 27(1):526–533, 6 2013.

[10] Ivan Titov and Ryan Mcdonald. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. *Proceedings of Association for Computational Linguistics (Proceedings)*, pages 308–316, 2008.

[11] Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models. *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*, pages 111–120, 2008.

[12] Yohan Jo and Alice Oh. Aspect and sentiment unification model for online review analysis. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011*, pages 815–824, 2011.

[13] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. 12 2002.

[14] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. 5 2002.

[15] Bo Pang and Lillian Lee. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. 9 2004.

[16] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. *International Conference on Information and Knowledge Management, Proceedings*, pages 625–631, 2005.

[17] Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125, 5 2006.

[18] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. pages 440–447, 2007.

[19] Xin ZHAO Jing JIANG, Hongfei LI YAN Xiaoming, Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid. *MIT*, pages 9–11, 2010.

[20] Anthony Aue and Michael Gamon. Customizing Sentiment Classifiers to New Domains: a Case Study. 2005.

[21] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and Chengxiang Zhai. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. *Proceedings of the 16th international conference on World Wide Web*, 2007.

[22] Chenghua Lin and Yulan He. Joint Sentiment/Topic Model for Sentiment Analysis. *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09*, 2009.

[23] Chenghua Lin, Yulan He, Richard Everson, and Stefan Rüger. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):1134–1145, 2012.

[24] Yulan He, Chenghua Lin, Wei Gao, and Kam Fai Wong. Dynamic joint sentiment-topic model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1), 1 2014.

[25] Feilong Tang, Luoyi Fu, Bin Yao, and Wenchao Xu. Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*, 488:190–204, 7 2019.