Comparative Opinion Mining: A Review

Kasturi Dewi Varathan
Department of Information System, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia. Email: kasturi@um.edu.my

Anastasia Giachanou and Fabio Crestani
Faculty of Informatics, Università della Svizzera Italiana (USI), Lugano, Switzerland

Opinion mining refers to the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information in textual material. Opinion mining, also known as sentiment analysis, has received a lot of attention in recent times, as it provides a number of tools to analyze public opinion on a number of different topics. Comparative opinion mining is a subfield of opinion mining which deals with identifying and extracting information that is expressed in a comparative form (e.g., "paper X is better than the Y"). Comparative opinion mining plays a very important role when one tries to evaluate something because it provides a reference point for the comparison. This paper provides a review of the area of comparative opinion mining. It is the first review that cover specifically this topic as all previous reviews dealt mostly with general opinion mining. This survey covers comparative opinion mining from two different angles. One from the perspective of techniques and the other from the perspective of comparative opinion elements. It also incorporates preprocessing tools as well as data set that were used by past researchers that can be useful to future researchers in the field of comparative opinion mining.

Introduction

Social media has changed the way people communicate (Qualma, 2010). Many people are exchanging opinions, emotions, ideas, and critiques on social media platforms because they are more free to express how they feel about a particular aspect or entity. Besides that, studies reveal that they spend a significant amount of time on these media (Buzzetto-More, 2012). This is proven by a vast amount of opinions being posted about movies, music, and views on

certain aspects, people, or events on blogs and forums or review sites (Romero, Galuba, Asur, & Huberman, 2011). The decisions people make are influenced by the opinions and emotions posted on these media (Bar-Haim, Dinur, Feldman, Fresko, & Goldstein, 2011). Brightlo-cal's local online review survey reveals that 88% of customers trust online reviews to make their decisions.[1] The findings reveal the significance of reviews, and every 9 of 10 people read online reviews. This makes current sentiment analysis and opinion mining an important field of research in (Liu, 2012).

The use of these opinions has become a crucial aspect that needs to be addressed (Asur & Huberman, 2010). Besides expressing his or her own positive or negative opinions about an entity or aspect, a person can also express opinions through comparing similar entities or aspects. Take the sentence "picture quality of iPhone 6 is better than Samsung Galaxy Note 3." In this paper, we would like to use these types of comparative opinions. In the past, information on competitors were gathered from formal media such as newspapers, yearly reports, and surveys to gauge the market perspective. Currently, business organizations are relying on real-time customers' reviews when making strategic decisions. This kind of comparative opinion mining contributes to competitive intelligence in helping business organizations identify risk and potential markets at early stages. This helps organizations when making strategic decisions to enhance their sustainability and growth. In addition, it also helps organizations build online reputations and helps individuals during the decision-making process.

A review on comparative opinion mining is very much needed because there is an increasing number of people who rely on online reviews for their decisions. This article reveals the techniques used so far and to what extent the research in this area has progressed. It also identifies the strengths and weaknesses of each method that were explored so far based on the elements (sentence, entity, relation, and

feature) of comparative opinion mining. This survey is also catered for future researchers to focus on some of the research gaps that exist in each of the elements of comparative opinion mining.

Thus, this article focuses on comparative opinion mining in which the first section of the article gives a brief introduction on opinion mining and comparative opinion mining. The main content of this article focuses on comparative opinion mining techniques and mining comparative elements, which include mining comparative sentences, entity detection, relation detection, and feature detection. Resources available for comparative opinion mining are also discussed to help future researchers use these resources. This survey covers research on comparative opinion mining for the past 9 years (2006–2015). We have covered about 38 articles in this survey that focused on comparative opinions. The survey focuses not only on English but also on other languages, especially Chinese and Korean. The articles of other languages that were published in English venues are included in this survey. However, we only concentrate on research articles that focus on comparative opinions obtained from customers' reviews, blogs, or forums. Studies made on comparative statements in general text such as web documents, news, or any other scientific text are not within the scope of this study because they may contain facts rather than opinions (Bos & Nissim, 2006; Fiszman, Demner-Fushman, Lang, Goetz, & Rindflesch, 2007; Huang, Wan, & Xiao, 2011; Huang, Wan, Yang, & Xiao, 2008; Park & Blake, 2012; Sun, Wang, Shen, Zeng, & Chen, 2006; Yang & Ko, 2011, 2009; Zhang & Jin, 2012). Finally, we present the different performance measures that are frequently used in the literature of comparative opinion mining.

Opinion Mining

In this section, we present an introduction and a brief description for the tasks of opinion and sentiment analysis and comparative opinion mining. It is meant to show the substantial differences between the two areas of research. The section concludes with a description of how this review is organized.

Opinion and Sentiment Analysis

Social media platforms (e.g., Facebook, Twitter, Google+, several blogs) enable people to express and share their thoughts and opinions on the web in a very simple way. User-generated data are important because they contain valuable information about the public view on any topic. However, looking through the vast amount of data to extract useful information is nearly impossible. To this end, researchers have started developing approaches that can automatically mine and analyze opinionated information within a huge amount of data.

Opinion mining, also known as *sentiment analysis*, is one of the emerging fields focused on developing methods that can automatically detect opinionated information and determine the polarity of the opinion toward a specific target. The opinion target is usually a named entity such as an organization, individual, or event. Formally, Liu and Zhang (Liu, 2012, p. 4) define an opinion a "a subjective statement, view, attitude, emotion, or appraisal about an entity or an aspect of an entity from an opinion holder." The same definition states that "an entity is a concrete or abstract object such as product, person, event, organization which can be represented as a hierarchy of components, sub-components, and their attributes" (p. 3). For example, in the sentence "The photos quality of my new Nikon camera is excellent," the entity for which an opinion is expressed is the Nikon camera, the aspect of the entity is the photos quality, and the polarity of the opinion is positive.

Opinion mining is a valuable tool for a variety of applications that require understanding public opinion. One typical example is business intelligence. This refers to enterprises that can use opinion mining to capture the views of customers about their services or products and use this information to improve their quality (Inui et al., 2008). In addition, potential customers of a product can consult past experiences of other users to decide whether to buy the product (Bickart & Schindler, 2001; Morinaga, Yamanishi, Tateishi, & Fukushima, 2002). Another application, known as government intelligence, refers to using opinion mining by government to understand what citizens really need and want and act promptly (Arunachalam & Sarkar, 2013). Finally, opinion mining can be used in politics for understanding what voters think (Laver, Benoit, & Garry, 2003).

Opinion mining has been studied on many media including reviews, forum discussions, blogs, and microblogs. There are numerous articles that have been published for opinion mining and sentiment analysis. Two long and detailed surveys on opinion mining were presented some time ago by Pang and Lee (2008) and Liu and Zhang (Liu, 2012).

Comparative Opinion Mining

Comparison plays an important role in evaluating products, services, people, action, and so on. To achieve the objective of being able to compare objects especially from web 2.0 technologies that incorporate user-generated data, we first need to identify and classify which are the statements that belong to the comparative and the noncomparative categories. Mining opinion and sentiments alone is insufficient because this will only show how much people talk and how they feel about certain products or services. This leads to incorrect judgments because these reviewers are likely not to have any experience on other products or services. On the other hand, comparative opinion is mainly provided by people who have experience related to the products or services they are comparing and it certainly reveals information that is more precise on the similarities and differences between products or services. These kinds of evaluations are important in gauging

customers' views because they manage to reveal precise rather than vague information that could be otherwise retrieved from purely direct opinion–based reviews. Many people analyze reviews to help them make decisions. For example,

> Example 1: "Samsung Note 3 is good in terms of performance."
> Example 2: "Samsung Note 3 is good in terms of performance compared to iPhone."

From these 2 sentences, it clearly shows that Example 2 will give more precise information to the future purchaser for helping him or her to choose one product compared to another. This is the kind of statement that needs to be retrieved and analyzed. Comparative opinion mining has also become a niche area of opinion mining because we are overloaded with a huge amount of user-generated data. Comparative opinion statements comprise only 10% of the total opinionated text (Kessler & Kuhn, 2013), and these comparative statements provide precise information to users. Thus, information could be derived in which business organizations and individuals are expected to be the key players who will benefit from comparative opinion mining.

Organization of This Review

This review focuses on comparative opinion mining. It is divided into two main parts: comparative opinion mining techniques and comparative opinion mining elements. This is similar to looking at two sides of the same coin. The Comparative Opinion Mining Techniques section talks about comparative opinion mining techniques exploring research from the perspective of the techniques. On the other hand, the Mining Comparative Opinion Elements section reviews research from the perspective of the opinion elements (sentence, entity, relation, feature). There, research work based on each of the elements is covered in detail. The Resources for Comparative Opinion Mining section instead covers the resources that are used and available to be used by the future researchers for researching in this area, including available data sets and tools. Conclusions and future work are given in the Conclusions and Future Work section.

Comparative Opinion Mining Techniques

Different techniques have been employed to address the task of comparative opinion mining. These techniques can be roughly divided into the following classes:

- Machine Learning
- Rule Mining
- Natural Language Processing

A machine learning approach employs a machine learning method and a set of different features to address different tasks of comparative opinion mining. The machine learning methods can be further divided into supervised and unsupervised learning methods. A rule mining approach considers methods that are based on association rules and sequential patterns. The rules can be either manually defined or automatically mined from text. The natural language processing (NLP) approaches include methods that apply semantic or syntactic analyses of the text.

In the rest of this section, we briefly describe the methods that have been applied to address the different tasks of comparative opinion mining. This section is organized based on the approach used regardless of the task being studied.

Machine Learning Approaches

Machine learning is a subfield of computer science that studies methods and algorithms that can learn from data and make predictions (Mohri, Rostamizadeh, & Talwalkar, 2012). In comparative opinion mining, approaches from the field of machine learning are usually employed to address different tasks. For example, classifiers are developed and trained with the aim to distinguish comparative sentences from noncomparative sentences. Support vector machines (SVM), naïve Bayes, and conditional random fields (CRFs) are some of the most popular approaches for comparative opinion mining.

Supervised learning. Supervised learning includes methods that learn a function given a set of training data. The goal of those methods known as classifiers is to predict the class attribute for the unlabeled data. The classifiers are trained based on a set of features extracted from the training data. Two of the most popular features are the opinion terms (love, great, hate, etc.) and the part-of-speech tags (POS) that are used to annotate the words with their syntactic behavior (i.e., verbs, adjectives, adverbs).

One of the most popular text classifiers is the SVM. According to this method, training data are marked as belonging to one category and mapped into the space so that the categories are as wide as possible. SVM has been used in many applications including mining comparative sentences. Wang, Zhao, Xin, and Xu (2015) employed SVM to build a model that could classify sentences as comparative or noncomparative. Keywords, sequence patterns, and manual rules were used as features to train the model. Their approach was applied to Chinese comparative customer reviews.

Liu, Huang, Zhang, Chen, and Chen (2013) also employed SVM to identify Chinese comparative sentences. The SVM classifier was trained using class sequential rules, comparative words, and statistical feature words. SVM was also used by Li et al. (2011) to classify the polarity of the review sentences. Tkachenko and Lauw (2014) used SVM for comparative sentence and aspect detection.

A multiclass SVM was proposed by Xu, Liao, Lau, Tang, and Wang (2009) for identifying and categorizing the comparative relations. In contrast to the typical SVM classifiers

that are only applied for two classes, the multiclass SVM can be applied for more than two classes. Xu et al. (2009) proposed to train a classifier for every category and then to assign the classifier with the higher score as the prediction category. Linguistic features including entity types, words, and POS tags were used as features to train the model.

The CRFs proposed by Lafferty, McCallum, and Pereira (2001) are a type of statistical modeling method. In contrast to a typical classifier, CRFs also consider the labeling context to identify relationships between data. CRFs were used by Xu, Liao, Li, and Song (2011) to extract comparative relations from customer reviews. Specifically, they proposed a graphical model based on a two-level CRF to extract and visualize the comparative relations between products. Complicated dependencies between relations, entities, and words and the interdependencies among relations were used as features. Their algorithm was applied on Amazon customer reviews, epinions.com, blogs, Short Message Service, and emails. One interesting conclusion from this study was the usefulness of the comparative relation map for enterprise risk management and decision making.

CRFs were also applied by Liu et al. (Liu, Xu, et al., 2013) for extracting comparative elements from the sentences that were already identified as comparative. In their study, they applied CRFs to identify the product name and the target attribute in the comparative sentences. Words, POS, distance from the comparative word, and domain-specific words were used as features for the element extraction. Another work that applied CRF was presented by Feldman, Fresko, Goldenberg, Netzer, and Ungar (2007). They employed CRFs to detect terms that frequently appeared with a specific product. For this reason, they built one CRF model for POS tagging and one for chunking. The proposed CRF model was trained on the CoNLL-2000 (Tjong Kim Sang & Buchholz, 2000) corpus.

Another classifier that is frequently used by researchers in comparative opinion mining is naïve Bayes. This classifier is one of the most popular classifiers used for text classification (Mohri et al., 2012). The model is based on the Bayesian theorem and performs well when there is a large number of dimensions. In text classification, naïve Bayes follows a bag-of-words approach to compute the posterior probability of a class. In other words, the text is considered as a bag of words without considering grammar and word order. Naïve Bayes was employed by Jindal and Liu (2006a) to classify sentences as comparative. Jindal and Liu (2006a) used class sequential rules as features. The naïve Bayes classifier was also used by Tkachenko and Lauw (2014) to detect comparative sentences and entities of interest.

Unsupervised learning. Unsupervised learning is a type of machine learning algorithm used to detect hidden structure in unlabeled data. Clustering is one of the most popular unsupervised approaches. Clustering groups a set of data in such a way that similar data will be in the same group. K-means clustering is one of the most widely known clustering methods. K-means groups the observations into k clusters in which each observation belongs to the cluster with the nearest mean.

In the literature of comparative opinion mining, unsupervised learning is underexplored. To the best of our knowledge, applying unsupervised learning on comparative opinion mining has only been considered by Tkachenko and Lauw (2014), who proposed a generative model for comparative sentences. Their proposed model called CompareGem (COMParative RElation GEnarative Model) jointly modeled comparative directions at the sentence level and ranked at the entity level. Gibbs sampling (Casella & George, 1992), a Markov chain Monte Carlo method, was used for hidden variables inference. CompareGem was evaluated on both supervised and unsupervised settings for relation and entity detection. K-means was used as a baseline in the unsupervised setting.

Rule Mining Approaches

Association rule mining is a popular data mining method for discovering interesting relations between data. When applied to text mining, its main objective is to discover strong rules and patterns between terms. Sequence pattern mining is another type of rule mining. In contrast to association rule mining, it considers the order of terms within a sentence. According to the definition given by Agrawal, Imieliński, and Swami (1993) the rules in association rule mining are defined as $X \Rightarrow Y$, where X, Y are subsets of items.

This mining model is general and can be applied in many domains. For example, it is used in market analysis for a more profitable product promotion, in medical diagnosis to assist physicians to cure patients, and in bioinformatics for analyzing DNA sequences. Association and sequential rules have also been extensively applied in Web and text documents.

Associations were considered by Ganapathibhotla and Liu (2008) with the aim to identify the preferred entity in the comparative sentences. Pros and cons reviews were used to determine the sentiment orientation of the comparative word and the feature. Comparative words, compared features, compared entities, and a comparison type were used as features in their method. They also proposed one-side association measures to compute the association between a comparative word and a feature.

Class sequential rules (CSR) is a type of sequential pattern mining in which each data sequence is labeled with a class. CSR was proposed by Jindal and Liu (2006a) to discover frequent patterns in comparative sentences. The sequence patterns, which were used as features, were based on POS tags and comparative key phrases. The sentiment of each pattern was then determined by the sentiment of the manually identified key phrases. Their experiments on news articles, consumer reviews, and forum discussions showed
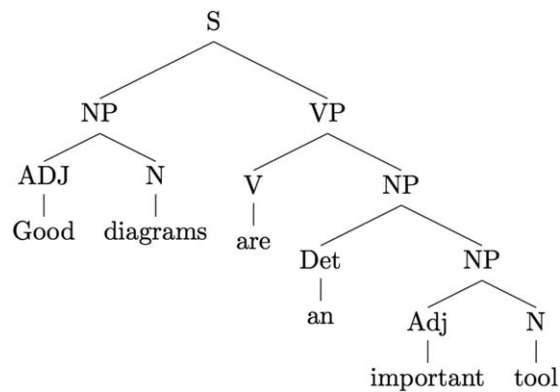
FIG. 1. The parse tree of the sentence "Good diagrams are an important tool."

FIG. 2. The dependency tree of the sentence "Nick bought a camera."

that combining CSR with manual rules could effectively identify comparative sentences.

A CSR-based approach was also considered by Liu, Xu et al. (2013) who proposed rule-based and CSR-based methods to identify comparative sentences. Comparative words, adverbs, and syntactic patterns were used by the rule-based approach. For the CSR-based approach, they combined sequential rules mining and machine-learning techniques. The identified comparative sentences were then classified into three classes according to their sentiment orientation. CSR was also considered by Liu, Huang, et al. (2013) who used those rules to train the classifier to detect comparative sentences written in Chinese. The results were significantly better when CSR features were considered during the training.

Rule mining was also used by Liu, Hu, and Cheng (2005) for developing a framework with the aim to analyze and compare consumer opinions. They proposed a system called Opinion Observer that could be used for comparing opinions on different products. Supervised rule mining was applied to extract product features that were part of the comparison. The generation of rules was based on POS tags and n-grams. Experiments on a data set extracted from Epinions showed that the proposed system was effective.

In addition to CSR, Jindal and Liu (2006b) also applied label sequential rules (LSR) to identify comparative sentences. LSR are defined as $X \Rightarrow Y$, where Y and X are sequences and X is produced by Y by using wildcards to replace some of the items of Y. Jindal and Liu (2006b) employed POS and comparative keywords as features. Experiments on consumer reviews, forum postings, and news articles showed that LSR managed to effectively identify the comparative sentences.

Instead of mining rules from text, a number of researchers manually defined a number of rules that they considered being indicative of comparison. This approach was applied by Gu and Yoo (2010) who proposed using sentence pattern rules and comparison words to detect comparative sentences that were written in Korean. To this end, they defined sen-
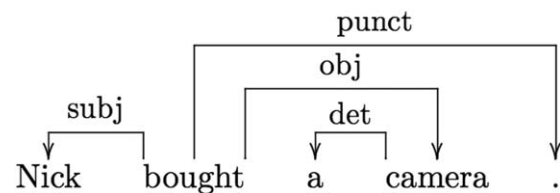
tence rules indicative of comparative sentences. In their approach, they used comparison words and POS tagging. They evaluated their method on comments about restaurants.

A pattern-based approach was used by He, Yuan, and Wang (2012) on reviews about mobile phones written in Chinese. He et al. defined six patterns to extract the comparative relations from reviews. A comparative and a superlative feature lexicon were used to detect the polarity of the features. The results showed that the pattern-based method managed to significantly improve the baseline method that was on the label sequence rules.

A similar approach was proposed by Kurashima, Bessho, Toda, Uchiyama, and Kataoka (2008) who used language patterns to detect entities and extract comparative relations. In their study, they manually prepared language patterns to extract entities and opinion. Experimental results on reviews about movies showed that the proposed method can effectively extract comparative relations.

NLP Approaches

NLP is a field of computer science that focuses on developing methods that can process and analyze written or spoken languages. NLP methods can be applied to analyze language at two different levels: syntactic analysis and semantic analysis. Syntactic analysis parses and analyzes the syntax of sentences. Semantic analysis aims to identify and analyze the meaning of words, phrases, and sentences. Here we present the studies that applied a syntactic or a semantic analysis to the comparative opinion mining task.

Syntactic analysis. Dependency parsing is a type of syntactic analysis and aims to identify the syntactic relations between words. The output of dependency parsing is the syntax tree that reflects the syntactic structure of a sentence. Figure 1 shows an example of a syntax tree of the sentence "Good diagrams are an important tool."

Dependency parsing has been used in many domains including computational linguistics, information retrieval, opinion mining, and so on. Different information can be derived by the dependency parsing including dependency grammar graph. Dependency grammar graph identifies and assigns grammatical roles to the words of a sentence. For example, it is used to identify the subject, the direct object, the indirect object, and so on. To make this clearer, we provide the dependency tree of the sentence "Nick bought a

camera" in Figure 2, which shows the syntactical relations within the sentence.

Identifying the grammatical roles is useful in comparative opinion mining because it can recognize the directions of comparative relations or mine feature opinion pairs. To this end, Sun, Long, Zhu, and Huang (2009) proposed an automated system based on dependency grammar graph and evolution tree. Their system could compare and recommend products to the customers from both subjective and objective perspectives. Dependency grammar graphs were used to mine feature opinion pairs. Then, the authors proposed an evolution tree to visualize how a specific product evolved. The evolution tree was also proposed to recommend customers the products about which customers had a good opinion.

In addition to the dependency grammar graph, Xu et al. (2011) extracted syntactic paths to extract comparative relations. Syntactic paths were useful in detecting comparative relations because they were usually expressed with similar syntactic patterns. Their experiments on Amazon customer reviews showed that syntactic features are useful for extracting comparative relations. Dependency trees were also employed by Liu, Huang, et al. (2013) who analyzed 2,000 comparative sentences with the aim to identify similar syntactic patters. After this analysis, they concluded that a similar syntactic structure is followed in the two parts of the comparative sentences.

Semantic analysis. Semantic analysis tries to address the problem of language ambiguity by extracting the meaning of the sentence. One popular method for semantic analysis is semantic role labeling (SRL). SRL detects the semantic roles and relationships associated with the verbs or predicates of a sentence and the semantic arguments. In SRL, the predicate represents an event and the arguments represent the participants of this event. Typical examples of semantic roles are Agent, Recipient, Patient, and so on. For example, in the sentence "The teacher talked to the student in the front desk," "the teacher" is the agent, "talked" represents the predicate, and "the student in the front desk" is the recipient. Assigning semantic roles is of high importance for several domains, such as Question Answering and Summarization.

Kessler and Kuhn (2013) proposed a model based on SRL to detect the entities and the predicate in comparative sentences. Using a standard pipeline from SRL, they identified the predicate and the arguments. The arguments were then classified in terms of being positive or negative arguments. A regularized linear logistic regression classifiers was used for the classification. The proposed approach managed to outperform the baselines on a data set containing blog posts about cameras and cars (Kessler, Eckert, Clark, & Nicolov, 2010).

Hou and Li (Hou, 2008) used the SRL approach to extract comparative relations in Chinese text. They used SRL to detect six elements in comparative sentences: holder, entity1, comparative predicates, entity2, attributes, and sentiments. POS, phrase type, position to predicate, and the comparative predicate were used as features in the SRL. As a
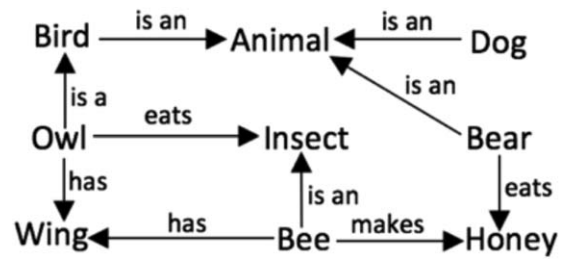

FIG. 3.   Example of a semantic network.

last step, they trained CRF classifiers using those features to label the comparative sentences. Their experiments on two different data sets of 200 comparative sentences each showed that using SRL can effectively mine Chinese comparative sentences.

Semantic network analysis is a semantic approach that is based on a network to reflect the semantic relations between concepts. Figure 3 shows an example of a semantic network on which a set of entities and their attributes are depicted. The most well-known example of semantic network is WordNet (Fellbaum, 1998), which is a large lexical resource of English. Words are grouped together into sets of synonyms that are linked to each other by semantic and lexical relations. WordNet has been also used in comparative opinion mining to detect synonyms and extend the list of comparative words (Jindal & Liu, 2006a), find relations between features and opinion words (Kim & Zhai, 2009), and identify synonyms of features (Liu et al., 2005).

Another example of a semantic network is SenticNet,[2] a knowledge base that includes a set of semantics, sentics, and polarity associated with 30,000 natural language concepts. SenticNet has been used to address concept-level sentiment analysis by leveraging semantics and linguistics (Cambria & Hussain, 2015). Poria, Cambria, Ku, Gui, and Gelbukh (2014) used SenticNet for aspect extraction from product reviews. In their work, they used common-sense knowledge and dependency trees to extract implicit and explicit aspects, a task that is very useful in comparative opinion mining.

Semantic network analysis was employed by Kurashima et al. (2008) who proposed a graph-based method in which the nodes represent the entities and the edges reflect the relations between two entities. According to the authors, the graph structure was appropriate because it could model the behavior of a potential customer who goes around until he finds the best entity. The authors also proposed the graph centrality score to measure the importance of an entity compared to the remaining entities.

Kim and Zhai (2009) also proposed a semantic approach to address the contrastive opinion summarization problem. This task was focused on summarizing opinions by detecting sentences with contrastive perspectives on the same product. For example, the reviews "The screen quality of MacBook Air is very good" and "MacBook Air comes with a poor

---

[2]http://sentic.net/

screen quality" are contrastive. Kim and Zhai (2009) addressed the contrastive opinion summarization problem with an optimization framework. The proposed method was based on several similarity measures. They also proposed two additional measures for this task, representativeness and contrastiveness. Representativeness measures how well the summary represented the original text whereas contrastiveness reflected the similarity between positive and negative sentences. The authors used reviews from Amazon to evaluate their method.

Li et al. (2011) employed an unified graphical model to represent each aspect of a relation. Each graphical model is based on all the comparative reviews obtained for a particular aspect. Their model was evaluated on reviews about phones and audio players. Finally, Fujimoto (2012) investigated mathematically potency-magnitude relations to model the degree of difference between two objects. To this end, they proposed a Q-magnitude Relation Map (Q-Map) and a Priority Message-Class Map (P-Map). Those models used evaluation target size and evaluation scale size to reflect how the relations change. One interesting conclusion of this study was that nonequal gradable messages that included intensified comparisons were most likely to change.

Conclusions

After presenting the techniques that have been used in the different tasks of comparative opinion retrieval, we observe that most of the research employed a rule based or machine learning approach. A pattern-based approach is suitable for comparative opinion mining because those sentences follow a specific pattern. However, they have to be combined with other methods to obtain a better performance. Unsupervised or NLP approaches that are still underexplored in this field could be suitable. For example, including synonyms to improve the performance of the approaches would be an interesting direction.

Mining Comparative Opinion Elements

This section explains comparative opinion mining from a different point of view compared to the previous section. In this section, comparative opinion mining is reviewed from the perspective of elements. In comparative opinion mining, research was seen in comparative opinion sentence detection, entity detection, relation detection, and feature detection. The advantage of this categorization is that it identifies the attributes that we have focused and helps in narrowing down the scope of the research works. It also helps the future researchers to identify the research gaps easily and more clearly. Finally, we highlight that the works done on each of these elements attains different objectives and therefore their comparison is not easy. Thus, this section is focused on exploring each one of these elements together with the techniques used. In the past, linguists spent some significant number of years in studying comparatives constructs in English language. Some had defined the problem as a universal quantifier over degrees of gradables (Lerner

& Pinkal, 1992), whereas others defined it as ordering of objects with some degrees of gradables (Kennedy, 2004). There is research that focused mainly on defining the syntax and semantics of comparative constructs from a linguistic perspective (Moltmann, 1992). Just by analyzing comparative opinion sentences with comparative words is ineffective. For example, "In order to be the best, we have to work hard" is not a comparative opinion statement although it contains the comparative adverb "best." There are also comparative sentences that will not contain any of the comparable words in lexicons, but belong to comparative opinion statements. There is also research on comparative sentiment identification through the use of logics (Ballard, 1988; Rayner & Banks, 1990; Von Stechow, 1984). Work reported in Friedman (1989) used comparatives in question answering systems. Subsequently, Staab and Hahn (1997) used a multi-layered interpretation model that combines syntactic and semantic approaches. All of these studies cover the logics of comparative statements and not detecting comparative opinions. The following subsections explore the elements of comparative opinion mining, which are comparative opinion sentence detection, entity detection, relation detection, and feature detection.

Comparative Opinion Sentence Detection

In this section, we look at past research on mining comparative opinion at sentence level. The first work on computational methods in comparative sentence extraction in opinionated text was seen in Jindal and Liu (2006b). However, research by Zhai, Velivelli, and Yu (2004) with a cross-collection mixture model for comparative text mining existed even before Jindal and Liu (2006b), but their research focused on how to retrieve different common themes across all of the available text collections and not on sentence-level identification. On the other hand, Liu et al. (2005) studied analyzing and comparing competing products based on opinions rather than on comparative sentences as a whole. Meanwhile, Kim and Zhai (2009) worked on summarizing contradictory opinions from opinionated text and not on comparative opinion sentences. For the past 9 years (2006–2015), the focus of the research on comparative opinion sentence detection covers machine learning, associative rule, and lexicon approaches. In total, there are six research articles on these approaches. The research focuses on the supervised machine learning technique (Wang et al., 2015) and rule mining (Liu et al., 2013). The associative rule approach in comparative sentence identification was initiated by Jindal and Liu and also used by Liu et al. (Jindal & Liu, 2006a, 2006b; Liu, Xu, et al., 2013). Meanwhile, Wei, Zhang, Yuan, and Chi (2014) explored a lexicon-based approach. The details on each one of these approaches are covered in depth in the following subsections.

Associative rule mining techniques. All research that falls under the category of associative rule mining uses CSR. CSR was used by Jindal and Liu and Liu et al. in

distinguishing the comparative and the noncomparative text. Jindal and Liu (2006a) have grouped comparatives into four types (nonequal gradable, equative, superlative, and nongradable). "Canon is better than Nikon" is an example of a nonequal gradable type that contains expressions that shows greater or lesser than. Meanwhile, an equative type shows the equality between the entity or entities. An example would be the following: "Canon is as good as Nikon." For superlative comparative type, the relation of one entity is compared with all the others, and the following is an example: "Canon is the cheapest of all." On the other hand, nongradable types do not grade the objects with which they are compared. A comparative sentence such as "Some of Canon's features are different than Nikon's" is an example of a nongradable type. Jindal and Liu (2006a) managed to distinguish between comparative and noncomparative sentences by using all of the different types of comparative sentences based on a set of 83 keywords. They carried out several experiments based on keywords, naïve Bayes, and SVM. The performance of naïve Bayes on both CSR and manual rules shows a significant improvement in precision but as expected, recall dropped.

Jindal and Liu (2006b) focused on gradable comparative sentences that are composed of nonequal, equative, and superlative, and neglecting the nongradable sentences. Nongradable sentences are sentences that do not have explicit grading. For example, "Hotel A has air conditioning, but Hotel B does not." This kind of comparison is implicit and cannot be graded. They also extended their own studies, which worked on CSR and took advantage of the label sequential rule (LSR). They filtered sentences that are not comparative by using the same set of keywords that they used previously using supervised learning approach. Both techniques, CSR and LSR, performed reasonably well.

Liu, Xu et al. (2013) also used CSR and a rule-based method on Chinese comparative sentences. The rules are compiled based on syntactic patterns and words comprising comparative words, comparative content words, and relative degree adverbs. On the other hand, the CSR method integrates class sequential rule mining and machine learning techniques. A list of comparative words is created and these words are used to generate the sequential data set from the training corpus. A multiminimum support was used to extract all the patterns that match the minimum threshold. This leads to an increase in their precision, but dropped recall. The reasons for low recall is because of the incomplete list of comparative keywords used, the complexity of the text, wrong spelling, and punctuation.

All of the work in the area of associative rule mining used sets of keywords and syntactic patterns in identifying comparative sentences. This only tackles the problem at the basic level because lexicons do not portray the meaning of these sentences. A sentence may have keywords and syntactic pattern that belongs to comparative category, but may not necessarily be a comparative opinion sentence. Thus, a semantic method is needed to handle this problem and improve the performance of the classifications obtained so far.

Machine learning techniques. Using machine learning approaches to detect comparative sentences has emerged in recent years. This can be seen with the study conducted by Wang et al. (2015) on supervised approaches that uses SVM and Liu et al. (2005) on rule mining techniques. Wang et al. (2015) worked on comparative sentence identification in Chinese text. They divided comparative sentences into two types, which are comparative sentences comprising comparative keywords and noncomparative keywords. The keyword search technique is used to identify sentences with comparative indicators and syntactic technique was used to identify comparative sentences that do not have comparative keywords on the Chinese Opinion Analysis Evaluation (COAE2012). A keyword lexicon was created that consist of 102 words and 30 syntactic sequences. SVM was used for filtering noncomparative sentences that contain keywords. To handle superlative comparative sentences, 45 patterns were also constructed. The precision performance obtained is the best precision in comparative sentence detection so far.

A supervised rule mining technique was used by Liu, Huang, et al. (2013) in identifying comparative sentences in the Chinese Opinion Analysis Evaluation (COAE2013) and online product reviews sites. They constructed three kinds of sentence structure based on these data sets and the characteristics of comparative sentences. There is a slight similarity between this research and the previous research that used SVM (Wang et al., 2015), in which both used syntactic structure for differentiating the sentences. The difference between these two studies is the use of the sentence structures. Liu, Hang, et al. (2013) used sentence structure for identifying comparative and noncomparative sentences. On the other hand, Wang et al. identified comparative sentences that do not have comparative keywords. The study also reported that high recall was achieved in identifying explicit comparative sentences compared to implicit ones. To handle the implicit comparative sentences, researchers mined more information by using dependency relations, and as a result they found two parts of the comparative sentences that have similar syntactic structures. A statistical technique was also used to compute the similarity dependency relation and to exclude all the sentences that were below a determined threshold. Their approach managed to give a satisfactory recall and not on satisfactory precision. This is because most of the sentences managed to fit into the broad template that was created. As a consequence, they proposed a SVM classifier that included a CSR, comparative words, and statistical feature words. They also extended their experiment on SVM by incorporating sentence structure and similarity of dependency relation. The result of this experiment showed an increase of all the performance measures compared to their own previous experiments.

It is clear from the previously mentioned work that supervised learning is used for identifying comparative sentences in Chinese text only. To the best of our knowledge, there is no research in comparative sentence detection that uses supervised techniques on English text or that covers other

languages. On the other hand, strangely there is no research on unsupervised learning of comparative sentence identification. Future research is welcomed to explore machine learning approaches in comparative opinion sentence identification.

Lexicon/NLP techniques. There are two articles that use syntactic techniques in comparative sentence detections in Chinese text (Gu & Yoo, 2010; Wei et al., 2014). Gu and Yoo (2010) used sentence structures and statistics in identifying comparative sentences. The sentences that contains graded comparative words such as "than," "more," "far," and "more" were extracted and classified as comparative sentences

Meanwhile, Wei et al. (2014) proposed a method that uses some usual conceptual arrangements of words in a sentence and the structural characteristics of comparative sentences. However, there are a few drawbacks in both of these studies: they did not use word segmentation, and this leads to identifying non comparative sentences as comparative. In addition, they failed to handle negation as well as comparative sentences that had omitted "than" but used "is."

Research on lexicon based approach should evolve. Semantic based technique would be a good area that needs to be explored as no research was found on this technique so far.

Conclusions. The research on mining comparative sentences is still at its infancy. Most of the techniques used are based on some kind of pattern based approach or machine learning approach with regard to the languages covered (English, Chinese). Surprisingly there are only two research articles on English text (Jindal & Liu, 2006a, 2006b), and the rest are on Chinese text (Gu & Yoo, 2010; Liu et al., 2013; Liu, Xu, et al., 2013; Wang et al., 2015; Wei et al., 2014). Future research on comparative sentences should venture into unsupervised approaches and lexicon based semantic approaches. Besides that, the collections of comparative keywords need to be increased to achieve better recall and precision because past research used sets of keywords for identifying comparative sentences. One easy approach could be to include synonyms for keywords as well as incorporating domain knowledge. But again, using just words in identifying comparative sentences is a conventional approach. Research should flourish beyond this aspect.

There is a large research gap on comparative opinion mining. Past research only focuses on mining comparative sentence and not on mining comparative opinion sentences. For example, "Everest is the highest mountain in the world" is a comparative sentence which contains fact rather than opinion. On the other hand, "my car is faster than yours" is an opinionated sentence. Unfortunately, none of the research managed to differentiate between comparative facts and comparative opinions. All of the articles covered so far are giving a general methodology that could be used for differentiating comparative with the noncomparative text. When researchers detected comparative sentences, they basically

managed to identify comparative sentences that exist in their data set (it could be reviews, blogs, or forums that belong to opinionated text and nonopinionated text, such as the news). Jindal and Liu (2006a, 2006b) and Liu, Hang, et al. (2013) uses reviews as well as news data sets for evaluating their work. In comparative opinion retrieval, this could be biased because they have considered news data sets that are prone to have factual data rather than opinionated text. In addition, assuming that all reviews are opinionated is also wrong. A review may contain facts as well as opinions. For example, "KLCC is taller than the parliament building of Malaysia" (fact), "Samsung is more trendy than iPhone" (opinion). Much research on comparative sentence mining manually annotated their sentences into comparative and noncomparative, but none has explicitly mentioned that they annotated it based on only comparative opinion sentences (Jindal & Liu, 2006a, 2006b; Liu, Xu, et al., 2013; Wang et al., 2015). Wang et al. (2015) eliminated direct opinion sentences and only concentrated in comparative opinions. However, they did not omit comparative sentences that do not contain opinions. On the other hand, Ganapathibhotla and Jindal (2008) worked on opinion in comparative sentences in which they made use of opinionated comparative words, context, and application domain when detecting opinion in comparative sentences. However, they focused more on identifying preferred entities rather than comparative opinion sentences. As a consequence, they did not perform any evaluation to support their findings on mining opinion in comparative sentences. In conclusion, research on mining comparative opinion sentences is still at its infancy.

Entity Detection

Entity detection is one of the main components of comparative opinion mining because it provides valuable information to consumers on other options that are available for the products in which they are looking. It might also help business organizations to know who are their current rivals. Besides that, the overall structure of the market can be identified and viewed. The entity can be represented as product names, service names, person names, and so on. Entity detection basically covers identifying brand names or model names.

In this section, we focus on how entities are identified in comparative opinion sentences. A closer look is needed because we are not handling general text with standard names or properly mentioned abbreviations or proper language text, with correct spellings and grammar. It is not right to expect reviews given by consumers to not have all of these. Dealing with informal language requires more specific techniques compared to formal text.

Machine learning techniques. Research on CRF is implemented in detecting entities in comparative sentences. Liu et al. (Liu, Xu, et al., 2013) worked on retrieving two entities in a single comparative statement. For a comparative statement that contains a single entity, a "NULL" value is

replaced. An example of such statement would be "the picture quality is good as that of Canon." They classified comparative sentence to contain four elements, which are subject, subject attribute, object, and object attribute. An example of such sentence is "Mercedes's the speed is faster than BMW's speed." In this case, the subject is Mercedes, the subject attribute is speed, the object is BMW, and the object attribute is speed. They proposed CRF to identify product names in Chinese opinion text. The product names are often mentioned at the end of a comparative sentence. Thus, CRF is used to identify entities by using only the end word in each of the element phrase as the head word that contains the entity name that exists in the domain knowledge. The performance of classification that uses a head word shows better precision and recall when compared with a baseline approach that tries to classify entities by using each words in a sentence. Very low recall is achieved, and this is caused mainly by the limited number of words in the list. In addition, this research also showed that higher domain knowledge delivers better results. In terms of machine learning, only CRF is used in retrieving entities in Chinese text. The CRF technique could also be used by researchers who deal with other languages to see whether this technique can be applied to non-Chinese text. Other techniques from machine learning still remain unexplored in entity detection from comparative opinion text.

Lexicon techniques. This section discusses all of the relevant literature on lexicon-based approaches in detecting entities in comparative sentences. Lexicon-based approaches are more popular when compared with other approaches. This could be seen from the number of articles published on this approach. Xu et al. (2011) worked on mobile phone data sets and had also used their own lexicon dictionary that contains mobile phone names and attributes to detect entities in a sentence. Because mobile phones may have different abbreviations for the same product, for example, "iPhone 6 Plus" is also known as "6+" or "iPhone6+" in reviews, they also included all of the possible abbreviations in building their lexicon. Anaphora that addresses the issues of word referring back to a word used earlier such as it, they, she, he, and so on were also studied by adopting a closest-first method. However, they realized the importance of performing postprocess steps that caters to products that are mentioned differently but refer to the same entity. For example, "Samsung S2" could be referred to as "S2." These kinds of different entity names needs to be normalized to avoid the duplication of entities. Because the focus of their research is not on entity normalization, they have neglected this phase.

Tkachenko and Lauw (2014) used a data-driven approach in detecting entities in comparative sentences. Their study revealed that to identify comparable entities, these entities must have been compared at least once in the opinionated text. Besides that, there exist entities that are compared indirectly with one another. To accommodate this, they found that if entity1 is compared with entity2 and entity2 is compared with entity3 then entity1 can be compared with entity3. They have focused their research to deal with comparative sentences that contains two entities only. Because there is no Name Entity Relation (NER) for the entities that they were detecting, they created their own dictionary based on product titles and employed a dictionary matching approach.

Feldman et al. (2007), identified an entity by emphasizing on the entities that are mentioned together in a sentence. Named entity recognition methods were used to accomplish this task. A graph was generated to identify which entity coexists with another. Overall visualization on how the market is structured based on entities from the perspective of consumers can be obtained. They also used a dictionary-based approach that collects all of the terms used to refer to the same entity. However, this approach is not sufficient to accommodate comparative opinion text because it contains many variations and it is not possible to make sure that all of these variations are covered by the dictionary. They realized that brand names are easier to identify, when compared with model names and model numbers. This is because a model number could be written in many formats or variations, for example, only numbers are written in a review without stating a model name. Another issue that needs to be taken into consideration is that names could be misspelled. To eliminate all of these problems, researchers devised a process of entity detection into two stages. The first stage was matching the sentences with a list of search strings from the web and using regular expressions when identifying the entities. Once these entities were identified, the second stage filtered out and modified the wrongly extracted entities from the list. In other words, it removed falsely recognized entities. All comparative sentences' sequences of tokens, POS tags, and NP chunks that satisfied minimum support values were retrieved. From these, frequent sequence sets were identified. These patterns have been filtered to accommodate only patterns that contain specific words or parts of speech. Identifying patterns with parsing failed to give fruitful results because of grammatical or spelling errors in the opinionated text. Their findings also show that simple pattern matching gives better results when compared with parsing.

Kessler and Kuhn (2013) proposed semantic role labelling (SRL) for detecting entities. They had used predicates with three arguments compared to the normal two. These three are entities that are evaluated positively, entities that are evaluated negatively, and the aspects in which these entities are evaluated in a comparative sentence. All of these three arguments were filled as semantic roles. The data set was annotated by humans on positive entity, negative entity, and aspect. It was then used as a training set for the SRL system. They were aware that identifying entity does not focus on comparing against a set of comparative keywords that lies between the entities alone. Thus, their research used POS in entity identification. This approach filled the gap that exists in achieving better recall because it managed to identify comparative words that are not stated in the set of comparative keywords as what had been used by other researchers such as Jindal and Liu (2006b). As a

consequence, any comparative opinion text could adopt this method without much hassle. The difference between this research and the other research is that entities are associated with roles and not with predicates. Because they worked with comparative sentences that contain two entities, they successfully detected both entities. Their results were compared with Jindal and Liu (2006b) that used a rule-based approach, which is discussed in depth in the next section, and their findings reveal that they outperformed all performance measurements (recall, precision, and *F*-score).

Research on entity detection based on lexicon approaches managed to cover a number of techniques when compared with other approaches. Techniques on dictionary-based, data-driven, regular expressions and semantic role labeling were used in detecting entity in comparative opinion.

Rule-based techniques. The rule-based technique is also used in entity detection. Jindal and Liu (2006b) worked with two entities and used LSR, which uses nouns and pronouns in extracting entities in comparative sentences. These entities were extracted with a set of 83 comparative keywords composed of four POS tags (comparative adjectives such as "taller," "bigger"; comparative adverb such as "faster," "more frequently"; superlative adjective such as "best," "most desirable"; superlative adverb such as "earliest," "hardest") and 79 keywords referring to indicative words such as outperform, beat, and so on. The rules were used to extract the components of the relation that consists of entities and features. LSR reports good results with high precision. However, the data set that they used for testing contained only 564 gradable comparatives, which included three types of comparatives (nonequal gradable, equative, and superlative). Although the data set contained mostly reviews and forum discussions, it also contained random news articles that may include factual information rather than opinionated information. Thus, the performance results obtained are questionable, because of the biases that exist in the data set.

Text mining techniques. He et al. (2012) worked with a mobile data set similar to Xu et al. (2011) using a lexicon-based approach. They used a text mining approach in identifying entities in comparative sentences. In this research, they categorized entities in comparative sentences in a different way compared to the previous approaches. Each of the comparative sentences contains a subject and an object, and this subject and object can be from mobile brands or mobile models. The Jaccard coefficient was used as a mining technique in identifying the entities in comparative sentences. The similarity value obtained from this technique was used in differentiating the subject as well as the object in the comparative sentence. The entity with high similarity value in the comparative sentence is categorized as a subject and the entity with a lower value is categorized as an object.

Conclusions. Not much research was found that solely catered for entity detection in comparative opinion mining. Research dealt with just using the existing approach to entity retrieval because the focus of the present research lied in other areas of comparative opinion mining. The accuracy of entity retrieval is important because it is one of the main backbones of comparative opinion mining. It is pointless to find the relation between entities or finding the features of entities when the entities are wrongly identified. Extracting an entity in comparative opinion text requires more precise techniques when compared with general text. It is tedious to cater for some fixed rules or a fixed number of lexicons. It is also tedious to involve human efforts to accomplish good results in entity detection because the data we are dealing with is large and requires automated mining. Data variations become a more complex problem to handle because opinionated text contains many variations in writing styles and is ambiguous. The use of a dictionary, POS tag analysis, or domain knowledge alone is not be sufficient to solve the problem of identifying entities in a comparative opinion. Up to now, only CRF techniques of machine learning are used for identifying entities in comparative text. There are many other techniques that could be explored, such as SVM, hidden Markov model, and so on. Other text mining approaches could also be used for identifying entities. Research so far has been narrowed to a maximum of two entities in a comparative opinion sentence. For example, "Rolex is expensive compared to Swatch." This sentence contains only two entities, which are Rolex and Swatch. Future research should study comparative sentences that contain more than two entities. In addition, the problem of pronoun resolution needs to looked into more carefully as it contributes significantly to entity detection.

Relation Detection

Relation detection is an important aspect because it helps in knowing the relation that exists between entities and how to rank these entities. This kind of information is valuable for business organizations in knowing the strengths and weaknesses of their products when compared with their competitors. In addition, individuals will also be able to benefit from relation detection because they will know what options are available and what suits them the most based on their requirements. In this section, we look at past research that handled relation detection and to what extent this relation detection had been explored in comparative opinion mining.

Machine learning techniques. This section explores research that used machine learning techniques in relation detection. Tkachenko and Lauw (2014) had proposed a generative model called CompareGem tackles both sentence-level as well as entity-level comparisons using both supervised and unsupervised approaches. It is a model that managed to fill up the gap that exists in the pipeline approach. For pipeline approach, it managed to show the entity that is superior than the other. However, there are some limitations, for instance, in Figure 4 it shows a pipeline example.

A is better than B
B is better than C
C is better than E
C is better than D

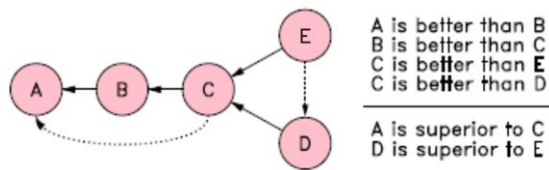A is superior to C
D is superior to E

FIG. 4. Pipeline example. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In this example, we cannot conclude whether E is superior to D or the other way around. In addition, they realized that using bags of words alone is insufficient to achieve this goal. Thus, this research successfully achieved it through a sigmoid ranking function. A sentence-level comparison gave an output of which entity is superior to the other. Significant results were obtained for CompareGem at the sentence level. Meanwhile, entity-level comparisons determine the overall ranking of these two entities that were compared in the comparative sentence. An entity that is preferred in a comparative statement does not mean it would be preferred when compared to the overall comparative statements. This kind of comparison is similar to the one reported in Kurashima et al. (2008), which used a graphical model in detecting relations but that takes into consideration the overall comparative relation in finding the superior entity by using queries. For further details on Kurashima et al. (2008), see the graphical model subsection. Although CompareGem managed to give a different perspective on the overall relation between comparative sentences, it failed to rank the top most entities. In their research, they did not cater to the equality that exists between entities because they claimed that it would not produce any effect. For example, A is better than B, B is better than C and D is equal to A. The probability that D is better than B is higher. However, to measure which entity is superior, equality plays an important role. Meanwhile, although CompareGem managed to produce highest accuracy for entity ranking, the results obtained are quite similar to the baselines. However, their supervised configuration managed to produce better result compared to their unsupervised approach.

Lexicon techniques. Lexicon-based techniques are used in detecting relation in comparative opinion mining. There are four research works that used this technique in their relation detection. Jindal and Liu (2006b) first researched proposed mining comparative relation in opinionated text. They defined comparative relation to be represented as follows:

relation word, features, entity 1, entity 2, type

The *relation word* is identified with a set of comparative keywords, *entity 1* and *entity 2* are the respective entities that exist in the comparative statement, and *type* represents one of the three different types that had been identified: nonequal gradable, equative, and superlative. Nonequal gradable expresses the preferences, for instance, "worse than" or

"better than." Equative comparative relation treats both entities as equal, and superlative sentences identify the relation between entities in which one of the entities is preferred to all others. Although these types of comparative relation classifications cover most of the comparative sentences, it failed to handle all types of comparative sentences that belong to nongradable. Nongradable comparative relations focus on comparing more than two entities without any explicit grading. In this study, they only handle comparative statements with only one relation. This caused drawbacks because user opinions may contain more than one relation that could exist between entities or features. Besides that, Jindal and Liu (2006b) only focus on categorizing comparative sentences into the three different categories that were mentioned earlier. They also failed to make use of the direction of the relations that identify the preferred entities. On top of that, because they used LSR, their rules are insufficient to be adopted by other domains.

Hou and Feng (Hou, 2008) represents the structure of the comparative relation as attributes, entity 1, operator and entity 2. An example of attributes would be price, speed, distance, capacity, and so on. Entity 1 and entity 2 are composed of products, services, persons, or locations. Meanwhile, operator is one of $\{>,<,=,\square,\square\}$ in which ">" represents that entity 1 is better than entity 2 and "<" represents vice versa. The "=" operator shows that both entity 1 and entity 2 are equal. Entity 1 is the best is shown using "$\square$" operator and "$\square$" means that entity 1 is the worst. In this study, they used a large number of comparative sentences: 1,800 sentences and 277 different comparative predicates. These sentences are used in training the CRF model. However, complex sentences that did not manage to go through the parser are excluded in the evaluation process. This would have led to some important sentences missing in the evaluation process and thus producing higher precision because the complex sentences are discarded automatically. Meanwhile, they only catered for sentence-level relation detection and not on overall relation. This means an entity can be considered superior, inferior, equal, worst, or best in a comparative sentence and not by considering the overall comparative sentences.

Semantic Role Labelling (SRL) was used by Gu and Yoo (2010) in Korean text. It is the first research that studies relation extraction in Korean text. For the first time, their research managed to identify the comparative directions in a relation and also identified the superior entity that exist in comparative statements. The structure of comparative Korean text consists of:

subject, target, feature, and superiority

In this comparative sentence, "Fabric A is more expensive than fabric C", the subject is "fabric A," "fabric C" is the target, feature is the price, superiority is "fabric A." Predicate information that lies next to the superiority or inferiority comparison word is used to indicate the relation that exists

in comparative sentences. Thus, they performed a rule-based approach to identify which is the superior entity. The rules are applied to all predicates which are located on the right hand side of a comparative word. The data set used for this study is unique as they used only comments on restaurants. Besides that, their evaluation centered only on retrieving comparative sentence based on four different comparative keywords, which are "than," "compared to," "far," and "more." The usage of this very limited amount of keywords shows that the research needs to be enhanced further in terms of keywords as well the rules used. They have basically tackled only the general part of comparative relation in opinionated text, although they achieved surprisingly good result in terms of precision.

He et al. (2012) built a comparative relation lexicon that contained comparative words. They identified relation through comparative words used in a sentence. This study suggested six matching patterns to extract comparative words from opinionated text. The matching pattern identified was syntactic, very simple, and straightforward. Figure 5 shows the matching patterns.

These patterns are too general to be adopted in detecting relations in comparative opinion sentences. They are also insufficient to cater to the complexity and the variations of Chinese language. It also clearly shows that no attention was given to comparative opinions, but only on the part of speech of comparative sentences.

Graphical model techniques. Graphical models in comparative sentences based on opinion are first used by Kurashima et al. (2008). They created a model that represents the relations that exist only between entities without including their features. The entities were also ranked against their competitors. Their approach is based on query in which the users can make a query on single entity or multiple entities, and the results will be based on overall comparative relations. This approach is somewhat similar to the one proposed in Tkachenko and Lauw (2014). In this research, they managed to find all of the competitors (entities) and aspects that were similar to the searched query. The searched query basically provided an idea on which entity cluster the user was interested in. A graph was generated based on the comparative relation in which nodes represented entities and edges represented the relations. All of the nodes relating to the queried entity or entities were retrieved. Their result shows the ranking of all the competitors based on the users' query. Thus, the customers could have an idea of the available options and of the best option that they could pick.

Xu et al. (2009) proposed a SVM-based map for comparative relation. They were the first to introduce a comparative relation map. This map consisted of entities, attributes, and relations between each one of these attributes with other entities. Their model managed to identify if a comparative relation exists and differentiates it into three sentiments categories, which are "better," "worse," and "same." They broke the norm by dealing with three entities in which handling these amounts of entities is a multiclass classification prob-

Pattern Number	Patterns
1	adv.+adj.
2	adv.+v
3	adv.+n.
4	adj.+m.
5	v.+m.
6	v.+n.

FIG. 5.   Matching patterns.

lem. However, the relation extractions that they handled were straightforward. In addition, the data set that they obtained from Amazon contained only 217 comparative relations, which were quite low to train their model that was based on a machine-learning approach. They should have used a larger data set for training, which could have given a better performance in their research work.

Xu et al. (2011) also used a graphical model to interpret the relation that could exist between entities. This graphical representation managed to show all the entities that could exist in the data set. This kind of relation visualization helps in supporting decisions made by business organizations in tracing the information about their competitors' strengths and the strengths of their own products or services. They identified how to differentiate comparative relation from other existing relation extraction models. Their findings revealed that comparative relation belonged to higher order relation in which it contained four entities or arguments. On the other hand, the existing relation contained only two entities. Besides that, comparative relation contained direction in which such directions do not exist in normal opinion statements. For example, "I like iPhone" is a normal opinion without any directions. On the other hand, "I like iPhone compared to Samsung," is a comparative statement that contains direction. To accommodate these needs, these researchers proposed a two-level CRF model with unfixed interdependencies that is a more powerful modelling tool compared to the existing CRF model. Although they managed to retrieve relations that were useful for business intelligence, their results are not very promising.

Graphical models were also used by Li et al. (2011) in relation detection. The difference between their model and others is in the comparisons made. They compared relation based on aspects such as design, feature, performance, and ease of use. Each aspect is represented with a unified graphical model that takes into account all the comparative reviews. This unified approach helped the customers to have a clear picture on which product seemed to be the best based on superiority scores of the aspects. A K-mean of clustering was used to group the comparative sentences in reviews and community based question answering (cQA) pairs. For polarity classification of these sentences, SVM is used. The results obtained higher average score in using both reviews

and cQA compared to just reviews. Although their model gave a different perspective on how comparative relations are performed, not all of the products could use the same aspects to make comparisons. For example, using performances aspects to compare fabrics, clothes, handbags, book, and so on are not feasible. Besides that, these aspects cannot be expanded to be used in nonproduct domains, such as services, because aspects will be different.

One-side association techniques. One-side association techniques are used in relation detection of comparative opinion mining. Ganapathibhotla and Jindal (2008) worked on identifying which is the preferred entity in a comparative opinion mining sentence. This research expected to benefit the consumers in deciding their best options. They emphasized on sentimental aspects of a sentence to determine if either Entity1 or Entity 2 is preferred. This sentiment is derived from context dependent opinion, which indicates whether the comparative word gives positive or negative opinions. For example, "more" does not mean to be positive all the time, it depends on which context this word is used. For instance, "Device A has more power consumption than Device B." In this case, the comparative word "more" and "power consumption" give negative value to an opinion. Thus, they derived the following rules:

Increasing comparative word + Positive ⇒ Entity 1 preferred
Decreasing comparative word + Positive ⇒ Entity 2 preferred
Increasing comparative word + Negative ⇒ Entity 2 preferred
Decreasing comparative word + Negative ⇒ Entity 1 preferred

Because context-dependent information were used in this research, they proposed the one-side association method to identify the number of times comparative words and features co-occur. The base forms of these words are also included. Besides that, synonyms and antonyms, which were retrieved from WordNet, were also used. The count increased every time comparative word and feature (including the synonyms) co-occured in positive comparative statements. The count was also increased if antonyms of comparative word and feature together with its synonym co-occur in negative statements. Besides features, adjective or adverbs were also used to detect the preferred entity. On the other hand, these researchers tackled negation statements as well. They had compiled a list of 26 negation words and negated the rules mentioned earlier in detecting the preferred entity. There are times in which negation in comparative opinion statements will not show any preference to any of the entities. An example of this kind of statement would be "Printer A's performance is not better than that of Printer B." The result obtained showed significant improvement compared to the baselines.

Conclusions. There are nine articles that worked on relation detection in comparative opinion mining in the past 9 years (2006–2015). From these nine studies, two studies were on machine learning approaches, two are from rule-based approaches, and the others used lexicon-based approaches. None of the research was focused on an unsupervised approach. Exploring unsupervised approaches for an instance clustering technique could group entities that belong to the same product or services category. From this, the user will be able to identify the relation that could exist between the products or services. There are also many other techniques that have yet to be ventured on, such as other text mining approaches that could be explored to identify which entities are always being compared with. Business organizations could benefit from this kind of association because it could identify the top most rivals that exist or allows consumers to know what are the best options available for a product or services that they are looking for. Besides these approaches, the languages that were covered in comparative opinion mining so far are only on English, Chinese, and Korean text. Future work may also make use of some past approaches to focus on other languages.

Feature Detection

Feature information on comparative opinion mining is very important. In fact, it is often useful to look at products' or services' features. Detailed information on certain entities could only be found in these features. For example, "I like Obama more than George Bush," failed to indicate based on which attribute (feature) the "liking" is. This kind of statement does not provide much information to the reader, especially in decision-making purposes. On the other hand, for instance, "Obama's administration capability is better than George Bush's" clearly shows that the preference is based on Obama's administration capability. The value of an entity is appreciated based on the feature in which the entity is compared rather than just the entity itself. In this section, we will focus on how features are detected and explored in comparative opinion mining.

Statistical techniques. This section focuses on statistical approaches used in feature detection and how they are implemented. Sun et al. (2009) handled features in comparative opinion text by building a sentiment-based product features database. The comparison between products was made based on the structured data that were extracted and saved in the database. Visualization based on tree models was used to identify options as well as criteria such as lowest price or bestselling items. Moreover, customers would prefer to have some detailed comparison that includes the feature aspect of the products that they wish to compare rather than some general information. To accommodate this, they proposed a recommender system that took into account all of the previous reviews and suggested products based on feature comparisons. For instance, a customer could post a query such as "Suggest a carto me that is better than BMW." Their system would list all the cars that are better than BMW based on each of the common features between BMW and other cars. Sentiment values on features based on past reviews were also included in suggesting products. Rather than just giving
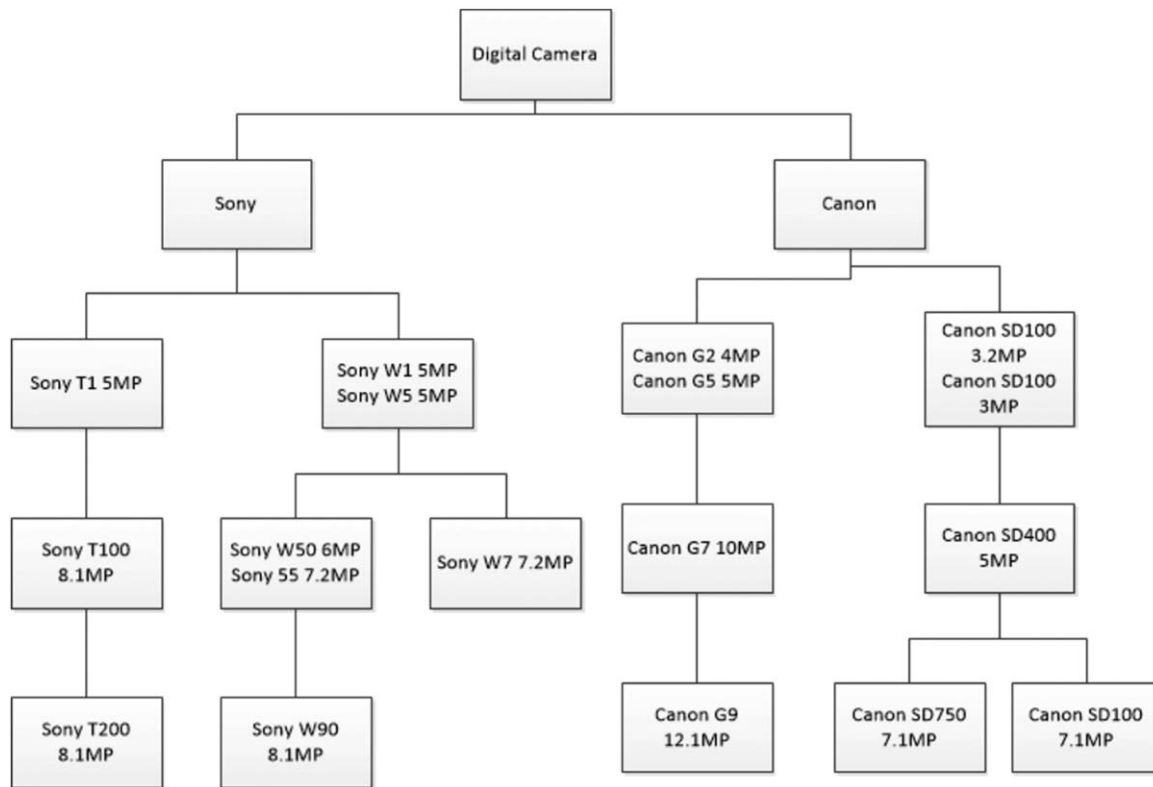
FIG. 6.   Evolutionary tree.

binary values sentiments, they used SentiWordnet to weigh each of the sentiment strength accordingly. This definitely gives a positive impact to the comparative opinion text they are handling because stronger sentiment opinion has a higher value than the weaker one.

Their system was tested using reviews obtained from Amazon with 18 different products' features. Furthermore, the evolutionary tree that was proposed managed to give a clear picture of the product evolution process and served as a guide to customers for getting the latest or better products in the market. These researchers looked into the utilization of features in comparative opinion mining from totally a different angle. A sample of their evolutionary tree is shown as the Figure 6.

On the other hand, they failed to handle synonyms of features that were mentioned differently in the review. For example, "photo," "image," and "picture" were all representing the same feature. Although this looked simple, their performance would have been much better with this integration.

Pattern matching techniques.   Pattern matching techniques were explored in feature detection. Tkachenko and Lauw (2014) proposed a generative model that includes information that can be derived on sequence of the features. These features can be categorized to syntactic feature that analyzes the placement of a word, whether it is before, between, or after the mentioned entity and negation features that exist to a targeted word. Because they are focusing on a data set of digital cameras, four different aspects that were frequently used to compare cameras such as functionality, form factor, image quality, and price were included as features that needed to be analyzed. Their analysis revealed some interesting points as indicated in Figure 7.

The sequence of existence of words is deemed very important. For example, column 1, row 1 of the element of the table (#1 from #2) shows that the sequence (#1 from #2) is the top discriminative feature in terms of functionality that contains in a comparative sentence that favored entity 1. Another example would be column 8, row 1(#1 more #2) reveals that if entity 1 is more than entity 2, then entity 2 is favored. The opposite of this can be seen in column 7, row 1(#1 less #2) which favors entity 1. However, because the features are manually identified in the camera domain, there is a possibility for this research to exclude some important features such as size.

Conclusions.   So far only pattern matching and statistical approaches were used for feature detection in comparative opinion sentences. Statistical approaches managed to give the detailed aspects of features and pattern matching enable to identify the preferred entity based on preidentified features. Research works were seen in using the features as relations and not much research was founded on solely feature detection studies. There are many aspects that can be looked into in feature detection for comparative opinion mining. Explicit and implicit feature detection in comparative opinion sentences could be a new research area. Besides that, topic modeling could also be used for identifying features from comparative opinion reviews.

Functionality		Form Factor		Image Quality		Price	
#1 is favoured(c=0)	#2 is favoured (c=1)	#1 is favoured(c=0)	#2 is favoured (c=1)	#1 is favoured(c=0)	#2 is favoured (c=1)	#1 is favoured(c=0)	#2 is favoured (c=1)
#1 from #2	from #1 #2	#1 lighter #2	# heavier #2	#1 than #2	#1 #2 better	#1 less #2	#1 more #2
#1 than #2	#1 #2 upgrad	#1 smaller #2	#1 #2 hand	#1 better #2	#1 us #2	purchas #1 #2	#1 cost #2
#1 over #2	had #1 #2	size #1 #2	#1 larger #2	qualiti #1 #2	#1 #2 detail	#1 better #2	about #1 #2
#1 better #2	#1 decid #2	#1 had #2	#1 #2 feel	pictur #1 #2	#1 #2 which	#1 instead #2	would #1 #2
#1 ha #2	#1 year #2	#1 over #2	#1 #2 lightest	imag #1 #2	#1 out #2	x #1 #2	camera #1 #2

FIG. 7.   Top five most discriminative features.

	Predicted *Positive*	Predicted *Negative*
Truly *Positive*	TP	FN
Truly *Negative*	FP	TN

FIG. 8.   Confusion matrix.

Conclusions

Research on categorizing comparative and noncomparative opinion sentences is not being explored yet. It is an important area that needs to be explored. This is because comparative opinion sentences are unique by themselves. A comparative sentence which contains opinion words need not necessarily be a comparative opinion sentence. Besides that, research on handling non gradable comparative sentences need to be explored as no research is found. Moreover, entity normalization is an important area that needs to be looked into. Past research used manual insertion by listing the possible abbreviations. This approach tends to miss some abbreviations which may affect the overall entity identification results. We found that most research on comparative opinion mining works only on two entities. Future research should cater for more than two entities. Many research studies were conducted on relation mining as compared to other elements mining. Although relation seems to be popular compared to the others, there is not enough research done on evaluation metrics in ranking the entities in comparative reviews. In addition, feature detection needs to be researched further as it contains the least amount of research studies.

Performance Measures

Comparative opinion mining can be considered as a classification problem because it aims at classifying if a sentence, entity, relation, or feature is retrieved or identified correctly. For this reason, the most frequent measures used in the literature are *accuracy*, *precision*, *recall*, and *F-score* that are adopted from traditional classification field.

To better explain these measures we provide one example. Consider that we want to measure the performance of a classifier on its ability to detect comparative sentences. Given a sentence, the classifier annotates as positive the sentences that are considered as comparative, otherwise the sentences are annotated as negative. Figure 8 summarizes the results of the classifier on the test data by providing the number of *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), and *False Negatives* (FN) instances. TP represents the number of sentences predicted as positive and

which were indeed positive, whereas FP is the number of sentences incorrectly predicted as positive. TN and FN have a corresponding meaning for the negative class. Based on this matrix, we now explain the most frequent comparative opinion mining measures.

- *Accuracy*: Accuracy measures how often the method being evaluated made the correct prediction. It is calculated as the sum of the true predictions divided by the total number of predictions, that is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Precision*: Precision is calculated as the ratio of sentences that were correctly predicted as positive divided by the total number of sentences that were predicted as positive, that is:

$$Precision = \frac{TP}{TP + FP}$$

- *Recall*: Recall represents the ratio of sentences that were correctly predicted as positive divided by the total number of the positive sentences and is calculated as:

$$Recall = \frac{TP}{TP + FN}$$

- *F-score*: Calculating only recall and precision is not enough. A combination of the two measures is more suitable to measure the performance of the methods. The *F*-score combines recall and precision and is calculated as:

$$F\text{-}score = 2 \cdot \frac{precision\,recall}{precision + recall}$$

These measures are mostly used for performance evaluation in comparative opinion mining research. These kind of metrics are important for future researchers in conducting their evaluations correctly. Besides that, using the right performance metrics will enable research to be benchmarked appropriately against the past research works.

Resources for Comparative Opinion Mining

This section reports on different resources, such as popular collections and preprocessing tools that have been used in research on comparative opinion mining.

Test Collections

A number of different test collections have been used in the literature of opinion mining. In most of the cases,

researchers built their own collections by crawling reviews from Amazon, Epinions, forums, and so on.

However, in most cases, a detailed description of the data set is missing, and researchers do not make collections publicly available. The lack of standard test collections creates considerable difficulties in comparing the effectiveness of different methods and systems.

In an effort to cover the lack of standard collections, some researchers built collections for comparative opinion mining and made them publicly available. Here, we present these data sets and indicate where they could be accessed:

- J&L data set[3]: the J&L data set that was created by Jindal and Liu (2006b) and contains about 650 comparative sentences extracted from reviews, blogs, and forum discussions. This test collection was built specifically for the task of detecting comparisons.
- JDPA data set[4]: the JDPA data set was built by Kessler et al. (2010) and contains 506 sentences about cameras and 1,100 about cars. The sentences were extracted from blog posts.
- Kessler14 data set[5]: the Kessler and Kuhn (2014) data set contains annotations for 1,707 sentences extracted from Epinions. The sentences are about cameras and are annotated with comparisons.

Preprocessing Tools

The task of comparative opinion mining requires a preprocessing phase that includes tokenization, stemming, POS-tagging, and named entity detection. A number of preprocessing tools have been used for such a phase. Apart from these tools, a number of tools that contain implementations of popular methods have been used to facilitate the research. Those include NLP toolkits and tools that contain implementations of machine learning methods. Although those tools are very important, in many cases they are not explicitly mentioned in articles. Here, we present the tools that have been used in the literature of mining comparative opinions:

- Gate[6]: Gate is a freely available toolkit for language processing that is provided by the Natural Language Processing Research Group of the University of Sheffield since 1995. Gate is a widely used toolkit for text mining. It provides a number of features including sentence tokenization, entity recognition, past of speech tagging, semantic tagging etc. Gate was used by Xu et al. (2011) to detect linguistic features from the reviews which were then used in mining comparative opinions.
- Stanford CoreNLP[7]: Stanford CoreNLP is a framework that provides a set of natural language analysis tools including tokenization, part-ofspeech tagging, named entity recognition, parsing, and coreference. Stanford CoreNLP was used

by Kessler and Kuhn (2013) for sentence segmentation and tokenization.
- Stanford Parser[8]: Stanford Parser is a natural language parser used to detect the grammatical structure of the sentences. For example, it can detect which term is the subject or object of a verb. It provides a number of features including producing dependency trees. Xu et al. (2011) employed Stanford Parser to detect the grammatical roles of the various entities in the sentences.
- SVM multi-class[9]: SVM multiclass contains an implementation of the SVM method. It includes a number of features including solving classification and regression problems. Xu et al. (2011) used this tool to build a model that can mine comparative opinions.
- CRF toolbox[10]: CRF toolbox is a Matlab toolbox that implements the CRF method. Xu et al. (2011) used this tool to build a two-level CRF model for extracting comparative relations.
- WEKA[11]: WEKA is a toolbox that contains the implementation of machine learning methods. Tkachenko and Lauw (2014) used WEKA to train their baselines which were based on SVM and naïve Bayes.

Conclusions and Future Work

To the best of our knowledge, this is the first work that reviews comparative opinion mining. There are some significant findings that were obtained from this review. Research on comparative opinion mining published in English venues so far were explored only in English, Chinese, and Korean languages. This finding opens the door to future researchers to embark upon and explore other languages as well. Our review also demonstrates that the approaches used in comparative opinion are skewed to NLP. Many studies were found using NLP based approaches in achieving their objectives. There are many other techniques that remain unexplored, such as unsupervised approaches or computational approaches that uses some mathematical theories, and so on.

Research in the area of comparative opinion mining should focus more on comparative opinion and not on comparative statements to avoid factual information. In terms of application areas that were covered so far are mainly focused on comparative opinions of products, although very few studies were seen using restaurant and bus service review data sets. The focus on other domains are important as well. Research that handles comparative opinion mining on public figures or politicians is not attempted at all.

Besides that, there was no research found in spam detection on comparative opinion mining. This is a useful research path that needs serious considerations among future researcher because spams pollute reviews and give the wrong perception to the decision makers who are relying on comparative reviews in making their future decisions.

Past research also shows that there is no proper data set that is available for the purpose of experimentation

[3]http://www.cs.uic.edu/liub/FBS/data.tar.gz
[4]http://verbs.colorado.edu/jdpacorpus/
[5]http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/reviewcomparisons/
[6]http://gate.ac.uk/
[7]http://nlp.stanford.edu/software/corenlp.shtml

[8]http://nlp.stanford.edu/software/lex-parser.shtml
[9]http://svmlight.joachims.org/
[10]http://www.cs.ubc.ca/~murphyk/Software/CRF/crf.html
[11]http://www.cs.waikato.ac.nz/ml/weka/

and evaluations in comparative opinion mining. Although Kessler and Kuhn's (2014) data set provided annotated comparative sentences, they still failed to check whether it is based on opinions or just mere comparative sentences. Many researchers created their own data set from publicly available reviews. However, the use of this kind of private data set made the evaluation biased and not transparent. It is also hard for the researchers to benchmark their results with others because the data set used for the evaluation purpose differs from one another. Public data set creation will help other researchers use and evaluate their research based on standardized data. Information scientists will be able to explore some of the research gaps that exist, which have been highlighted in this review. They may contribute in many ways, from building a proper data set to be used for comparative opinion mining to defining an evaluation measures that managed to rank the entities, sentences, and features that are part of the reviews.

Moreover, this survey points to another possible direction that the future researchers should embark, which is on the use of the time feature pertaining to comparative opinion mining. So far there is no research found on the use of time and this feature is indeed very useful for trend analysis, brand sustainability, preference analysis, and change analysis. This survey also revealed that more than 60% of researchers who worked on comparative opinion mining are Chinese researchers. This opens the door for sociological studies to find out the reason on why this group of people are more interested in comparative opinion mining when compared with other scholars.

Acknowledgment

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia (Eds.), Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (pp. 207–216). New York: ACM. doi: 10.1145/170035.170072.

Arunachalam, R., & Sarkar, S. (2013). The new eye of government: Citizen sentiment analysis in social media. In IJCNLP 2013 Workshop on Natural Language Processing for Social Media. (pp. 23–28). Asian Federation of Natural Language Processing.

Asur, S., & Huberman, B.A. (2010). Predicting the future with social media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 492–499). Washington, DC: IEEE Computer Society. doi: 10.1109/WI-IAT.2010.63

Ballard, B.W. (1988). A general computational treatment of comparatives for natural language question answering. In Proceedings of the 26th Annual Meeting on Association for Computational Linguistics (pp. 41–48). Association for Computational Linguistics.

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., & Goldstein, G. (2011). Identifying and following expert investors in stock micro-blogs. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1310–1319). Association for Computational Linguistics.

Bickart, B., & Schindler, R.M. (2001). Internet forums as influential sources of consumer information. Journal of Interactive Marketing, 15(3), 31–40.

Bos, J., & Nissim, M. (2006). An empirical approach to the interpretation of superlatives. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 9–17). Association for Computational Linguistics.

Buzzetto-More, N.A. (2012). Social networking in undergraduate education. Interdisciplinary Journal of Information, Knowledge, and Management, 7(1), 63–90.

Cambria, E., & Hussain, A. (2015). Sentic computing: A commonsense-based framework for concept-level sentiment analysis. Cham: Springer.

Casella, G., & George, E.I. (1992). Explaining the Gibbs sampler. The American Statistician, 46(3), pp. 167–174. Retrieved from http://www.jstor.org/stable/2685208

Feldman, R., Fresko, M., Goldenberg, J., Netzer, O., & Ungar, L. (2007). Extracting product comparisons from discussion boards. In Proceedings of the 2007 Seventh IEEE International Conference on Data Mining (pp. 469–474). Washington, DC: IEEE Computer Society. doi:10.1109/ICDM.2007.27

Fellbaum, C.D. (1998). Wordnet: An electronic lexical database. Cambridge, MA: MIT Press.

Fiszman, M., Demner-Fushman, D., Lang, F.M., Goetz, P., & Rindflesch, T.C. (2007). Interpreting comparative constructions in biomedical text. In Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing (pp. 137–144). Association for Computational Linguistics.

Friedman, C. (1989). A general computational treatment of the comparative. In Proceedings of the 27th Annual Meeting on Association for Computational Linguistics (pp. 161–168). Association for Computational Linguistics.

Fujimoto, K. (2012). Investigation of potency-magnitude relations of eWOM messages with a focus on intensified comparative expressions. In Proceedings of the 11th International Conference on Cognitive Informatics Cognitive Computing (pp. 163–173). IEEE.

Ganapathibhotla, M., & Liu, B. (2008). Mining opinions in comparative sentences. In Proceedings of the 22nd International Conference on Computational Linguistics (Vol. 1, pp. 241–248). Association for Computational Linguistics.

Gu, Y.H., & Yoo, S.J. (2010). Searching a best product based on mining comparison sentences. In SCIS & ISIS (Vol. 2010, pp. 929–933). Japan Society for Fuzzy Theory and Intelligent Informatics.

He, S., Yuan, F., & Wang, Y. (2012). Extracting the comparative relations for mobile reviews. In Proceedings of the Second International Conference on Consumer Electronics, Communications and Networks (CECNET) (pp. 3247–3250). IEEE.

Hou, F. (2008). Mining Chinese comparative sentences by semantic role labeling. In Proceedings of the International Conference on Machine Learning and Cybernetics, 2008 (Vol. 5, pp. 2563–2568). IEEE.

Huang, X., Wan, X., & Xiao, J. (2011). Comparative news summarization using linear programming. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers (Vol. 2, pp. 648–653). Association for Computational Linguistics.

Huang, X., Wan, X., Yang, J., & Xiao, J. (2008). Learning to identify comparative sentences in chinese text. In PRICAI 2008: Trends in Artificial Intelligence (pp. 187–198). Berlin Heidelberg: Springer.

Inui, K., Abe, S., Hara, K., Morita, H., Sao, C., Eguchi, M., . . . Matsuyoshi, S. (2008). Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In Proceedings of the WI-IAT '08. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008(Vol. 1, pp. 314–321). doi: 10.1109/WIIAT.2008.373

Jindal, N., & Liu, B. (2006a). Identifying comparative sentences in text documents. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 244–251). ACM.

Jindal, N., & Liu, B. (2006b). Mining comparative sentences and relations. In A. Cohn (Ed.), Proceedings of the 21st National Conference on Artificial Intelligence (Vol. 2, pp. 1331–1336). AAAI Press.

Kennedy, C. (2004). Comparatives, semantics of. In K. Allen (Ed.), Encyclopedia of language and linguistics, second edition, lexical and logical semantics section. Oxford: Elsevier.

Kessler, J.S., Eckert, M., Clark, L., & Nicolov, N. (2010). The ICWSM 2010 JDPA sentiment corpus for the automotive domain. In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC). Washington, D.C.

Kessler, W., & Kuhn, J. (2013). Detection of product comparisons-how far does an out-of-the-box semantic role labeling system take you? In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (pp. 1892–1897). Association for Computational Linguistics.

Kessler, W., & Kuhn, J. (2014). A corpus of comparisons in product reviews. In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14) (Vol. 14, pp. 2242–2248). European Language Resources Association (ELRA).

Kim, H.D., & Zhai, C. (2009). Generating comparative summaries of contradictory opinions in text. In Proceedings of the 18th ACM Conference on Information and Knowledge Management (pp. 385–394). ACM.

Kurashima, T., Bessho, K., Toda, H., Uchiyama, T., & Kataoka, R. (2008).Ranking entities using comparative relations. In Database and expert systems applications (pp. 124–133). Berlin Heidelberg: Springer.

Lafferty, J.D., McCallum, A., & Pereira, F.C.N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning (pp. 282–289). San Francisco, CA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=645530.655813

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. The American Political Science Review, 97(2), 311–331. Retrieved from http://www.jstor.org/stable/3118211

Lerner, J.-Y., & Pinkal, M. (1992). Comparatives and nested quantification. In Proceedings of the Eighth Amsterdam Colloqium (pp. 329–347). ILLC, University of Amsterdam.

Li, S., Zha, Z.-J., Ming, Z., Wang, M., Chua, T.-S., Guo, J., & Xu, W. (2011). Product comparison using comparative relations. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1151–1152). ACM.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and com-paring opinions on the web. In Proceedings of the 14th International Conference on World Wide Web (pp. 342–351). ACM.

Liu, C., Xu, R., Liu, J., Qu, P., Wang, H., & Zou, C. (2013). Comparative opinion sentences identification and elements extraction. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 4, pp. 1886–1891). IEEE.

Liu, Q., Huang, H., Zhang, C., Chen, Z., & Chen, J. (2013). Chinese comparative sentence identification based on the combination of rules and statistics. In Advanced Data Mining and Applications (pp. 300–310). Berlin Heidelberg: Springer.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. Cambridge, MA: The MIT Press.

Moltmann, F. (1992). Coordination and comparatives (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). Mining product reputations on the web. In Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 341–349). New York, NY: ACM. Retrieved from http://doi.acm.org/10.1145/775047.775098. doi: 10.1145/775047.775098

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1–135.

Retrieved from http://www.nowpublishers.com/product.aspx?product=INR&doi=1500000001http://dl.acm.org/citation.cfm?id=1454712 doi: 10.1561/1500000011

Park, D.H., & Blake, C. (2012). Identifying comparative claim sentences in full-text scientific articles. In Proceedings of the Workshop on Detecting Structure in Scholarly Discourse (pp. 1–9). Association for Computational Linguistics.

Poria, S., Cambria, E., Ku, L.-W., Gui, C., & Gelbukh, A. (2014). A rule-based approach to aspect extraction from product reviews. In Proceedings of the Second Workshop on Natural Language Processing for Social Media (SOCIALNLP) (pp. 28–37).

Qualman, E. (2010). Socialnomics: How social media transforms the way we live and do business. New York: John Wiley and Sons.

Rayner, M., & Banks, A. (1990). An implementable semantics for comparative constructions. Computational Linguistics, 16(2), 86–112.

Romero, D.M., Galuba, W., Asur, S., & Huberman, B.A. (2011). Influence and passivity in social media. In Machine learning and knowledge discovery in databases (pp. 18–33). Berlin Heidelberg: Springer.

Staab, S., & Hahn, U. (1997). In Proceedings of the 14th National Conference on Artificial Intelligence and 9th Innovative Applications of Artificial Intelligence Conference (pp. 616–621). Menlo Park, CA: The AAAI Press.

Sun, J., Long, C., Zhu, X., & Huang, M. (2009). Mining reviews for product comparison and recommendation. Polibits, 39, 33–40.

Sun, J.-T., Wang, X., Shen, D., Zeng, H.-J., & Chen, Z. (2006). CWS: A comparative web search system. In Proceedings of the 15th International Conference on World Wide Web (pp. 467–476). ACM.

Tjong Kim Sang, E.F., & Buchholz, S. (2000). Introduction to the conll-2000 shared task: Chunking. In Proceedings of the Second Workshop on Learning Language in Logic and the Fourth Conference on Computational Natural Language Learning (Vol. 7, pp. 127–132). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from http://dx.doi.org/10.3115/1117601.1117631. doi: 10.3115/1117601.1117631

Tkachenko, M., & Lauw, H.W. (2014). Generative modeling of entity comparisons in text. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (pp. 859–868). ACM.

Von Stechow, A. (1984). Comparing semantic theories of comparison. Journal of Semantics, 3(1), 1–77.

Wang, W., Zhao, T., Xin, G., & Xu, Y. (2015). Exploiting machine learning for comparative sentences extraction. International Journal of Hybrid Information Technology, 8(3), 347–354.

Wei, X., Zhang, Q., Yuan, Y., & Chi, Z. (2014). Conceptual collocations of words in comparative sentences. In Proceedings of the 15th Chinese Lexical Semantics Workshop.

Xu, K., Liao, S.S., Lau, R.Y., Tang, H., & Wang, S. (2009). Building comparative product relation maps by mining consumer opinions on the web. AMCIS 2009 Proceedings.

Xu, K., Liao, S.S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. Decision Support Systems, 50(4), 743–754.

Yang, S., & Ko, Y. (2009). Extracting comparative sentences from Korean text documents using comparative lexical patterns and machine learning techniques. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers (pp. 153–156). Association for Computational Linguistics.

Yang, S., & Ko, Y. (2011). Extracting comparative entities and predicates from texts using comparative type classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, pp. 1636–1644). Association for Computational Linguistics.

Zhai, C., Velivelli, A., & Yu, B. (2004). A cross-collection mixture model for comparative text mining. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 743–748). ACM.

Zhang, R., & Jin, Y. (2012). Identification and transformation of comparative sentences in patent Chinese-English machine translation. In Proceedings of the International Conference on Asian Language Processing (IALP), 2012 (pp. 217–220). IEEE.