

OntoLSA – An Integrated Text Mining System for Ontology Learning and Sentiment Analysis

Ahmad Kamal, Muhammad Abulaish, and Jahiruddin

Abstract Since the inception of the Web 2.0, World Wide Web is widely being used as a platform by customers and manufactures to share experiences and opinions regarding products, services, marketing campaigns, social events, etc. As a result, there is enormous growth in user-generated contents (e.g. customer reviews), providing an opportunity for data analysts to computationally evaluate users' sentiments and emotions for developing real-life applications for business intelligence, product recommendation, enhanced customer services, and target marketing. Since users' feedbacks (aka reviews) are very useful for products development and marketing, large business houses and corporates are taking interest in opinion mining and sentiment analysis systems. In this chapter, we propose the design of an Ontology Learning and Sentiment Analysis (*OntoLSA*) system for ontology learning and sentiment analysis using rule-based and machine learning approaches. The rule-based approach aims to identify candidate concepts, which are analyzed using a customized HITS algorithm to compile a list of feasible concepts. Feasible concepts and their relationships (both structural and non-structural) are used to generate a domain ontology, which is later on used for opinion mining and sentiment analysis. The proposed system is also integrated with a visualization module to facilitate users to navigate through review documents at different levels of granularity using a graphical user interface.

Ahmad Kamal
Department of Mathematics, Jamia Millia Islamia (A Central University), New Delhi-25, India,
e-mail: ah.kamal786@gmail.com

Muhammad Abulaish (*corresponding author*)
Department of Computer Science, Jamia Millia Islamia (A Central University), New Delhi-25,
India, e-mail: abulaish@ieee.org

Jahiruddin
Department of Computer Science, Jamia Millia Islamia (A Central University), New Delhi-25,
India, e-mail: jahir.jmi@gmail.com

1 Introduction

The emergence of Web 2.0 has caused rapid proliferation of e-commerce and social media contents. Web is widely used as a platform by users and manufactures to share experiences and opinions regarding social events, political movements, products, services, and marketing campaigns. Enormous availability of user-generated contents (e.g., customer reviews) on the Web have attracted researchers to design and develop linguistically motivated computational paradigms to retrieve most cogent and desired information for developing real-life applications, including business intelligence, recommendation system, target marketing, and Web surveillance. However, exponential growth of unstructured and semi-structured user-generated contents on the Web and their uncontrolled generation consisting of various natural language nuances poses a big challenge for research community to fully automate the process of information component extraction. Thus, distillation of knowledge from such sources is a technically challenging task and requires research at the intersection of various disciplines, including natural language processing, ontology engineering, information retrieval and extraction, and computational intelligence.

1.1 *Opinion Mining and Sentiment Analysis*

Over the past decade, a new research discipline of opinion mining and sentiment analysis as a special case of web content mining has emerged, which mainly focuses to analyze review documents. Sentiment analysis is the computational evaluation of users' opinions, subjectivity, appraisals, emotions, and feedbacks expressed in review documents [1]. Since opinions are very informative in developing marketing and product development plans, large business houses and corporates are taking interest in opinion mining systems for developing business intelligence applications and recommendation systems. Such systems can process users' opinions and sentiments to predict better recommendations for target marketing of products and services. Usage of opinion mining and sentiment analysis has encouraged governments and security agencies in enhancing their abilities to analyze web usage patterns of users on the Web. Intelligence system can be able to track sentiments of Web users on the basis of their usage patterns, friends, and communities they form and participate. Such analysis is very useful in observing and discarding bashing or flames used in social communication and helpful in developing spam detection system [2].

Various research efforts attempted to mine opinions from customer reviews at different levels of granularity, including word-, sentence-, and document-level sentiment analyses. However, development of a fully automated opinion mining and sentiment analysis system is still elusive. A study conducted in [3] reveals that a product feature can appear as an explicit or implicit mention in review documents. Explicit features precisely and clearly appear in one or more review sentences, whereas an implicit feature does not explicitly appear. For example, consider the first and second review sentences of Table 1, in which product features are italicized and opinion

words are marked as bold. In first sentence, the features *video quality* and *software* are explicitly mentioned along with their respective opinion bearing words *great*, *amazing* and *friendly*. In contrast, second sentence does not mention any product feature explicitly, rather an opinion word *affordable* is appearing which can be used to infer the implicit feature *price*. Another challenge remains in the fact that a large number of feature words are referenced by anaphoric pronouns present in succeeding sentences of a review document. For example, anaphora word *it* in the fourth sentence of Table 1 is actually referring to the product feature *battery* mentioned in the third sentence.

Table 1: Sample opinion bearing sentences

S. No.	Sentence
1	The <i>video quality</i> is great . The <i>software</i> is amazing and friendly .
2	So much packed in a small case and very affordable !
3	<i>Battery</i> goes off so quickly.
4	Many times it becomes very hot while talking.
5	The <i>battery life</i> is very good .
6	The <i>battery life</i> is almost good .

The challenges of opinion mining and sentiment analysis have been viewed from different perspectives, including feature and opinion extraction to sentiment classification. A study conducted in [4] on movie reviews reveals that sentiment classification problem is different from traditional topic-based classification, as topics are frequently inferable by keywords, whereas sentiment can be expressed in a more subtle or delicate manner. For example, the opinion sentence “how could anyone sit through this movie?,” does not contain any single word that reflects negative opinion. This study also reveals that the task of identifying the seed set of keywords representing positive or negative sentiment is less trivial. Moreover, opinions appear with different strength and levels, including some very strong and weak opinions [5]. Therefore, degree of expressiveness of opinion is required for describing the expression level of an opinion [6]. For example, consider the fifth and sixth sentences of Table 1 that contain an opinion bearing word *good*, but the associated modifiers *very* and *almost* are fuzzy qualifiers expressing different levels of customer satisfaction regarding the *battery life* feature of the product.

In general, sentiment analysis task is context-sensitive and domain-dependent despite of the fact that general views of positive and negative opinions remain consistent across different domains [7]. For example, same opinionated word *low* expresses different sentiments (positive or negative) in the feature-opinion pairs ⟨price, low⟩ and ⟨quality, low⟩, respectively. Moreover, presence of negations in review sentences makes the task of automatic sentiment classification a more challenging task. For example, in the review sentence “the picture quality of this camera is not really good”, the opinionated word *good* requires inference in negative sense for proper polarity determination with respect to the *picture quality* feature of the product.

1.2 Role of Ontology in Sentiment Analysis

Ontology provides a knowledge management structure to represent concepts (e.g. product features) and their relationships in an inter-related and comprehensible manner. Such knowledge management structure is useful for both users and software agents. Ontology is helpful in providing structural framework for information expression and enhancing the quality of information extraction. However, ontology development and maintenance is an expensive task and require significant participation of domain experts. Overall task related to relevant knowledge extraction and structuring is both non-trivial for a particular domain. Besides, an effective solution to this problem might be to develop a system that automatically extracts knowledge from source documents for ontology creation and enrichment [8]. Since product features are generally nouns [3] and opinions are adjectives [9], rule-based approaches based on Parts-Of-Speech (POS) information along with some statistical techniques for feasibility analysis can be used to identify concepts (representing features and opinions) and their relationships for ontology learning from text documents.

Though the development of feature-based opinion mining and sentiment analysis systems are gaining momentum, fewer research efforts have been applied in exploiting the benefits of structuring product features for sentiment analysis. It is observed that some of the features exhibit relationship with each others in product domains. According to Cadilhac et al. [10] and Wei et al. [11], the knowledge of hierarchical relationships among product features is not utilized properly in opinion mining and sentiment analysis research. Cadilhac et al. [10] emphasized that knowledge management structure to represent concepts, relationships, and lexical information can be exploited to extract implicit features. Moreover, hierarchical analysis with ontological information can be exploited in propagating feature specific sentiment information to infer sentiments at higher levels of abstraction, specially in the cases where explicit mentions of sentiment information with associated features are missing. In addition, product ontology expedites comprehensive review summarization and facilities in visualizing customer reviews at different levels of granularity.

In this chapter, we propose the design of an Ontology Learning and Sentiment Analysis (*OntoLSA*) system for ontology learning and sentiment analysis using rule-based and machine learning approaches. The proposed system is equipped with a crawler to retrieve relevant reviews from merchant sites and applies various pre-processing techniques for their cleaning, chunking, and parsing using a statistical parser. A rule-based system using POS information and dependency relationships between the words is developed to identify candidate concepts, which are analyzed using a customized HITS algorithm for feasibility analysis [12]. Since ontology is not merely a data store rather a knowledge management tool, domain expert supervision on the list of feasible concepts is performed for validation. Feasible concepts and their relationships (both structural and non-structural) are used to generate an ontology, which later on used to extract opinionated words associated with the feasible concepts. We propose word-level sentiment classification method with the aid of statistical and machine learning techniques to determine users' sentiments. One of the crucial requirements when developing an opinion mining and sentiment analysis

system is the ability to browse through the review documents and to visualize various concepts and their relationships along with sentiment information present within the collection in a summarized form. Thus, the proposed system is also integrated with a visualization module to facilitate users to navigate documents collection at different levels of granularity using a graphical user interface.

The rest of the chapter is structured as follows. Section 2 presents a review of the related works on ontology-based opinion mining and sentiment analysis. Section 3 presents the architectural and functional details of the proposed Ontology Learning and Sentiment Analysis (*OntoLSA*) system. Section 4 presents the design of a graphical user interface for review summarization and visualization. The experimental setup and evaluation results are presented in section 5. Finally, section 6 concludes the chapter with future directions of work.

2 Related Works

Opinion mining and sentiment analysis is the computational evaluation and representation of users' sentiments and emotions expressed in customer reviews for developing practical real-life applications. Over the past decade, various research attempts have been made to mine product- or service-specific users' sentiments at document level to classify reviews as positive, negative, or neutral [4, 13]. Document-level sentiment analysis approaches treat each review document as a basic information unit [14]. Such approaches neither acquire insight knowledge regarding product features and their relationships nor exhibit users' feature-specific sentiments. As observed by many researchers, a positive review might contain features that are liked as well as disliked by an opinion holder, but the overall sentiment on the review remains positive. Similarly, a negative review does not mean that the opinion holder dislikes everything regarding the product. Therefore, research work based on feature-based opinion mining is proposed in [3], wherein authors applied a three-step process for feature and opinion extraction. In the first step, product features commented by the end users are identified. In the second step, all opinion sentences containing extracted features are identified and marked as either positive or negative. Finally, review documents are compiled on the basis of individual feature to classify positive and negative opinion sentences. In [15], the authors proposed the design of OPINE system based on an unsupervised pattern mining approach which extracts product features using feature assessor and web PMI statistics. Various extraction rules are applied to associate product features with potential opinion phrases. Relaxation labeling approach is used to classify the sentiment information of opinion phrases. In [16, 17, 18], the authors applied a finer-grained analysis on customer reviews to associate product features and sentiment bearing words on the basis of their explicit co-occurrence at sentence level. In [16], a supervised multi-knowledge based approach is proposed for movie reviews analysis, which applies grammatical rules to identify feature and opinion pairs. In contrast, Hu and Liu in [3] associated each feature word with its nearest opinion bearing word to form feature and

opinion word pairs. As a result many ambiguous and noisy pairs were extracted due to complexity of the sentences in movie reviews. A study in [17] explored the link between opinion target (product feature) and opinion expression in review sentences. Using data sets containing car and camera reviews in which features and opinions are manually annotated, a Support Vector Machine (SVM) based classifier is trained for identifying related features and opinions. Considering the algorithm proposed in [19] as a baseline, authors claimed to achieve F-score of 0.70, outperforming the baseline which yields F-score of 0.45 only. A study in [18] emphasized that a complete opinion is always expressed in one sentence along with its relevant feature. Therefore, to avoid false association between features and opinions, it is better to identify them at sentence level. Further, it is highlighted that feature and opinion word pairs retain intra-sentence contextual information for reflecting relevant opinions. Similarly, inter-sentence contextual information is also considered to imitate the relationship among opinions on the same topic or feature. Thus, both intra-sentence and inter-sentence contextual information are exploited for development of effective opinion retrieval method. In [20], a semi-supervised technique based on double propagation approach is proposed to extract opinion words and product features using a small seed of opinion lexicon, and thereafter the newly extracted opinion words and product features are exploited further for feature and opinion words extraction. Syntactic relationships are exploited using dependency parser to associate opinion words with appropriate features.

Applications and usage of ontology-based text information processing for knowledge management structure in a machine-interpretable format have attracted a number of researchers in the field of opinion mining and sentiment analysis. In [21], the authors presented the development of a system (OMINE) for identifying product features using premodeled knowledge. Manually created ontology for car domain (car functions, properties, and components) is used as a set of candidate features. The ontology is dynamically enriched by searching for phrases which match any of the manually defined lexical patterns. In [22], an ontology supported polarity mining technique is proposed for movie reviews in which a hybrid approach for ontology development is adopted. Authors stated that their ontology based approach is not only helpful in improving the efficiency of sentiment analysis task over a standard baseline, but also helps in understanding and management of movie reviews. In [23], a fuzzy ontology based context-sensitive opinion mining system is presented in which a variant of the Kullback-Leibler divergence statistical learning technique is used to predict sentiment polarities. In [24], the authors proposed an ontology-based combined approach for sentiment classification on movie reviews. Ontology development is performed using Web Ontology Language (OWL), and sentiment classification is performed using Support Vector Machine (SVM). In [11], a supervised Hierarchical Learning (HL) process is adopted to propose a manually constructed Sentiment Ontology Tree (SOT) to structure product attributes and associated sentiments from digital camera reviews. Knowledge derived from HL-SOT is utilized to handle sentiment analysis in a hierarchical classification process. Mukherjee and Joshi exploited ontological information in aggregating feature-specific sentiments to derive the overall polarity of review documents [25]. Ontology construction is

accomplished using ConceptNet 5¹, a very large machine usable semantic network of common sense knowledge. Kontopoulos et al. proposed an ontology based approach for fine-grained sentiment analysis of twitter posts with respect to the subjects discussed in tweets [26]. A sentiment grade for each aspect is obtained, which facilitates in detailed analysis of post opinions with respect to a specific topic. A semi-automatic, data-driven ontology editor OntoGen is used for ontology learning [27] and OPenDover² web service is employed for sentiment analysis. However, authors emphasized that use of third party sentiment analysis tool may be considered as one of the limitations of their proposed work as working approach of such tool cannot be verified; the source code and methodology are not publically available.

3 Proposed *OntoLSA* System

In this section, we present the architectural and functional details of our proposed *OntoLSA* system. The major functionalities of the proposed system are – *review document crawling, document preprocessing, concept mining and ontology learning, sentiment analysis, and review summarization and visualization*. The functional details of these modules are presented in the following sub-sections.

3.1 *Review Document Crawling*

For a target review site, the crawler retrieves review documents and stores them locally for further processing. The data samples used in our experimental work consist of review documents on different electronic products crawled from *www.amazon.com* using crawler4j API³.

3.2 *Document Preprocessing*

Crawled review documents are filtered to remove markup language tags and unwanted texts, such as meta-data information containing details regarding source of a review, author name, description, posting date, and star ratings. Filtered review documents are tokenized into record-size chunks (sentences), boundaries of which are decided heuristically on the basis of the presence of special characters. Depending on the application, a record-size chunk may contain a sentence, a paragraph, or

¹ <http://conceptnet5.media.mit.edu>

² <http://opendovr.nl>

³ <http://code.google.com/p/crawler4j>

a complete document. Thereafter, the sentences are parsed using Stanford parser⁴, which assigns POS tags to every word in a sentence. The POS tag reflects the syntactic category of the words and plays vital role in identification of candidate constituent for ontology learning from texts. The POS tags are also useful to identify the grammatical structure of the sentences, like *noun*, *verb*, *adverb*, and *adjective* phrases and their inter-relationships.

Stanford parser is also used to convert each sentence into a set of dependency relations between the pair of words. The dependency relations between a pair of words w_1 and w_2 is represented as $relationType(w_1, w_2)$, in which w_1 is called head or governor and w_2 is called dependent or modifier. The relationship $relationType$ between w_1 and w_2 can be of two types - (i) direct relationship, or (ii) indirect relationship [28]. In a direct relationship, one word depends on the other or both of them depend on a third word directly, whereas in an indirect relationship, one word depends on the other through other words or both of them depend on a third word indirectly. A list of sample review sentences along with the POS tags and dependency relationships generated by the Stanford parser are shown in Tables 2, 3, and 4, respectively.

Table 2: Sample review sentences

S. No.	Sentence
S1	During calls the speaker volume is very good.
S2	The sound sometimes comes out very clear.
S3	Nokia N95 has a pretty screen.

Table 3: Sample sentences with POS tags

S. No.	Sentence with POS tags
S1	During/IN calls/NNS the/DT speaker/NN volume/NN is/VBZ very/RB good/JJ ./.
S2	The/DT sound/NN sometimes/RB comes/VBZ out/IN very/RB clear/JJ ./.
S3	Nokia/NNP N95/NNP has/VBZ a/DT pretty/JJ screen/NN ./.

3.3 Concept Mining and Ontology Learning

The functionality of this module is to facilitate the linguistic and semantic analyses of texts to identify candidate constituents for knowledgebase. The candidate concepts are analyzed using a customized HITS algorithm to compile a list of feasible concepts. Feasible concepts and their relationships (both structural and non-

⁴ <http://nlp.stanford.edu/software/lex-parser.shtml>

Table 4: Sample sentences with dependency relationships

S. No.	Dependency relationships
S1	during(good-8, calls-2)det(volume-5, the-3)nn(volume-5, speaker-4)nsubj(good-8, volume-5)aux(good-8, is-6)advmod(good-8, very-7).
S2	det(sound-2, The-1)nsubj(comes-5, sound-2)det(times-4,some-3)nsubj(comes-5, times-4)dep(clear-8, out-6)advmod(clear-8, very-7)acomp(comes-5, clear-8)
S3	nn(N95-2, Nokia-1)nsubj(has-3, N95-2)det(screen-6, a-4)amod(screen-6, pretty-5) dobj(has-3, screen-6).

structural) are used to generate an ontology, which later on used for opinion mining and sentiment analysis purpose.

3.3.1 Concept Extraction

The concepts or entities normally emerge as a noun phrase in customer reviews. Thus, every n-gram (1- and 2-grams) of a review sentence is accessed on the basis of POS information and dependency relationships between the words. For example, in the sample sentence S1, the bigram *speaker volume* is a concept in cell phone domain and can be identified using *nm* tag, which is a noun compound modifier used by the Stanford parser. During candidate concept extraction, various noises are noticed. For elimination of noise during extraction, cleaning steps are performed by removing the stop-words and apostrophes, numerals, special characters, and symbols associated with a word. Cleaning operation proceed further by discarding all noun phrases representing person, organization, or location, as the probability for a noun phrase representing a named entity to be considered as a valid concept in product domain is very low. For named entity annotation, we have used the NER module of Gthis, we have

Apart from candidate concept extraction, opinion mining and sentiment analysis research requires extraction of associated opinion information for sentiment analysis. The dependency relationship between words is very helpful for this purpose. Considering the sample sentence S1, the word *volume* (part of a candidate concept) is related to an adjective *good* with *nsubj* relation (a dependency relationship type used by the Stanford parser). Thus, *good* can be identified as opinion. Further, using *advmod* relation, the adverb *very* can be identified as a modifier to represent the degree of expressiveness of the opinion word *good*. In sentence S2, the noun word *sound* is a nominal subject of the verb *comes*, and the adjective word *clear* is adjectival complement of it. Therefore, *clear* can be extracted as opinion for the concept *sound*. Further discussion on such dependency relationships between words and their usage in sentiment analysis can be found in our previous works [29, 30]. We have defined various rules to tackle different types of sentence structures to identify information components constituting candidate concepts, modifiers, and opinions.

3.3.2 Feasibility Analysis

It is observed during concept extraction phase that various noun, verb, and adjective words extracted from review documents are not relevant features or opinions. Though the ontology concepts normally emerge as a noun phrase in customer reviews, but sometimes verbs are considered as nouns due to parsing error. The basic reason for occurrence of the noises is the presence of ordinary nouns, verbs, and adjectives that are not actual features and opinions, but extracted due to parsing errors and their association with each other. Another issue is that very often several customers comment on same product feature, and in many cases their opinions contradict with each other. To handle these issues, a reliability score, $0 \leq r \leq 1$, is calculated for each concept-opinion pair with respect to the review documents collection. A higher reliability score reflects a tight integrity between the elements of a pair. We follow the opinion retrieval model based on the Hyperlink-Induced Topic Search (HITS) algorithm [12] used by Li et al. in [31]. The extracted concept-opinion pairs are represented as an undirected bipartite graph and treated by the HITS algorithm to generate reliability scores for concept-opinion pairs. For applying iterative HITS algorithm, extracted concept-opinion pairs and review documents are modeled as hubs and authorities, respectively. A hub object (a concept-opinion pair) has links to many authorities (review documents) because the same concept-opinion pair may occur in many review documents. Similarly, an authority object (a review document) contains many concept-opinion pairs, and as a result many hubs (i.e., concept-opinion pairs) are linked to it. Formally, a bipartite graph is represented as a triplet of the form $G = \langle V_p, V_d, E_{dp} \rangle$, where $V_p = p_{ij}$ is the set of concept-opinion pairs that have co-occurrence at sentence level. $V_d = d_k$ is the set of review documents containing concept-opinion pairs, and $E_{dp} = \{e_{i,j}^k | p_{ij} \in V_p, d_k \in V_d\}$ refers to the correlation between documents and feature-opinion pairs. Each edge $e_{i,j}^k$ is associated with a weight $W_{i,j}^k \in [0, 1]$ denoting the strength or integrity of the relationship between the pair p_{ij} and document d_k . The weight of a pair p_{ij} across all sentences of the document d_k is calculated using equations 1, 2, 3, and 4, where $|d_k|$ is the number of sentences in document d_k and $0 \leq \alpha \leq 1$ is used as a trade-off parameter.

The feature score is calculated using term frequency (tf) and inverse sentence frequency (isf) in each sentence of the document, $tf(f_i, s_l)$ is the number of times f_i occurs in sentence s_l . N is the total number of sentences in the document, and $sf(f_i)$ is the number of sentences where the feature f_i appears [31, 32, 33]. Similarly, $tf(o_j, s_l)$ is the number of times opinion o_j appears in a sentence s_l and asl is the average number of sentences in the document d_k [34].

$$W_{i,j}^k = \frac{1}{|d_k|} \sum_{p_{ij} \in s_l \in d_k} [\alpha \times fScore(f_i, s_l) + (1 - \alpha) \times oScore(o_j, s_l)] \quad (1)$$

$$fScore(f_i, s_l) = tf(f_i, s_l) \times isf(f_i) \quad (2)$$

$$isf(f_i) = \log\left(\frac{N+1}{0.5 \times sf(f_i)}\right) \quad (3)$$

$$oScore(o_j, s_l) = \frac{tf(o_j, s_l)}{tf(o_j, s_l) + 0.5 + \left\{1.5 \times \frac{len(s_l)}{asl}\right\}} \quad (4)$$

The authority score $AS^{(t+1)}(d_k)$ of document d_k and hub score $HS^{(t+1)}(p_{ij})$ of p_{ij} in $(t+1)^{th}$ iteration are computed as a function of the hub scores and authority scores obtained in t^{th} iteration, using equations 5 and 6, respectively.

$$AS_{(d_k)}^{(t+1)} = \sum_{P_{ij} \in V_p} w_{ij}^k \times HS_{(p_{ij})}^{(t)} \quad (5)$$

$$HS_{(p_{ij})}^{(t+1)} = \sum_{d_k \in V_d} w_{ij}^k \times AS_{(d_k)}^{(t)} \quad (6)$$

In order to apply HITS algorithm, initial weight for each feature-opinion pair and review document is set to 1, and the process is repeated till the convergence point is achieved. In line with [32], the convergence is reached when score computed for two successive iterations for any review document or feature-opinion pair falls below a given threshold. In our experiment, the threshold value is set to 0.0001. After convergence, the generated hub scores of (f_i, o_j) pairs represent the soundness of the integration of the respective features and opinions. Thereafter, the reliability score r_{ij} for a pair (f_i, o_j) is calculated by normalizing the hub score using *min-max* normalization given in equation 7, where $HS(p_{ij})$ denotes the hub score of p_{ij} , and *NewMin* and *NewMax* are set to 0 and 1, respectively. The reliability score of a feature-opinion pair represents the reliability of the relatedness of the feature and opinion.

$$r_{ij} = \frac{HS(p_{ij}) - \min_{xy}\{HS(p_{xy})\}}{\max_{xy}\{HS(p_{xy})\} - \min_{xy}\{HS(p_{xy})\}} \times (NewMax - NewMin) + NewMin \quad (7)$$

3.3.3 Relationship Extraction

After identification of feasible concepts, their validation is done by the domain expert to make sure that only valid concepts are added in the ontology – as ontology is not a general data store, rather a knowledge management tool. A relationship is defined as a specific association between two ontology concepts and assumed to be binary in nature. It can be structural or generic. The structural relations (IS-A, HAS-PART, etc.) also known as conceptual semantic and lexical relations are obtained using WordNet [35], which is a large lexical database containing English words. In wordNet, nouns, verbs, adjectives and adverbs are grouped into synsets, each repre-

senting a distinct concept. Synsets are interconnected through conceptual-semantic and lexical relations.

In order to identify generic relationships between ontological concepts, a two-steps process is devised. In the first step, parts-of-speech information and dependency relationship between terms are exploited to identify candidate relations that associate ontology concepts; whereas in second step, a feasibility analysis is applied to retain only valid relations between the concepts. For example, consider the sample sentence S3 of Table 4, wherein “N95” (a constituent of the bigram “Nokia N95”) is the nominal subject of the verb “has” and “screen” is its direct object. Thus “screen” and “Nokia N95” can be identified as concepts related through “has” relationship.

3.4 Sentiment Analysis

Sentiment analysis can be treated as a ternary classification problem in which the opinion words are mapped into one of the *positive*, *negative*, or *neutral* class, with respect to a particular concept or feature. In this section, we present a two-step process for word-level sentiment classification. In first step, popular statistical text classification methods are used to generate score for target opinion words, whereas in second step, generated scores are used to engineer an enriched set of features to develop a word-level sentiment classification system using supervised machine learning techniques. Thereafter, sentiments at higher levels of abstraction are determined using domain ontology.

3.4.1 Statistical Feature Identification

This section presents the proposed statistical approach for feature extraction. As proposed in [36], popular association measures, including Pointwise Mutual Information (PMI) [37], Mutual Information (MI) [38], Chi-square (commonly known as Karl Pearson’s Chi-square), and Log Likelihood Ratio (LLR) [39] are used to compute score for each feasible opinion words. For association measures, a set of positive seed-words ($N^{(+)}$) and a set of negative seed-words ($N^{(-)}$) are compiled. Thereafter, for each opinion word, positive opinion score ($Score^{(+)}$), negative opinion score ($Score^{(-)}$), and final opinion score ($OpnScore$) are calculated using equations 8, 9, and 10, respectively; wherein “AssociationFunction” represents PMI, MI, Chi-square, or LLR. Similarly, $SeedPos_j$ and $SeedNeg_j$ represent the j^{th} positive and negative seed words, respectively.

$$Score^{(+)}(w_i) = \sum_{j=1}^{|N^{(+)}|} AssociationFunction(w_i, SeedPos_j) \quad (8)$$

$$Score^{(-)}(w_i) = \sum_{j=1}^{|N^{(-)}|} AssociationFunction(w_i, SeedNeg_j) \quad (9)$$

$$OpnScore(w_i) = Score^{(+)}(w_i) - Score^{(-)}(w_i) \quad (10)$$

The value of $OpnScore(w_i)$ is used in training data set to determine the sentiment polarity of opinion words. If $OpnScore(w_i)$ value is greater than 0, it is an indication that the word w_i has higher association with positive seeds set and its sentiment orientation is considered as positive. Similarly, a negative score refers higher association of the word w_i with the negative seeds set and reflects negative sentiment. An opinion score value as zero represents that w_i is equally associated with both positive and negative seeds set, and its polarity is considered as neutral.

3.4.2 Sentiment Determination using Machine Learning Techniques

In addition to a rich set of statistical features discussed in the previous section, some linguistic features, including negation, tf-idf, and modifiers are also considered for classification purpose. The proposed sentiment analysis system is implemented as a two-phase process – model learning (aka training phase) and classification (aka testing phase). The training phase uses the feature vectors generated from training data set to learn classification models, whereas the classification phase is used to determine the sentiments of words from test data set. Four different classifiers, including Naive Bayes [40], Decision Tree (J48) [41], Multilayer Perceptron (MLP) [42], and Bagging [43] are considered initially, but finally settled with Bagging and Decision Tree (J48) algorithms implemented as a part of WEKA [44] due to their best performance. Once the sentiment class of individual opinionated words is determined, the semantic orientation at higher levels of abstraction is determined with the help of domain ontology.

3.4.3 Ontology-Based Sentiment Propagation

As discussed earlier, knowledge management structure provided by ontology in a machine-interpretable format is very helpful in propagating feature-specific sentiment information to infer sentiments at higher levels of abstraction, especially in the case where explicit mentions of sentiment information with associated features are missing. Motivated by the work presented in [11, 25], we propose an unordered sentiment propagation ontology tree T with a finite set of one or more nodes, in which there exists a specially designated node called *root*, and remaining nodes are partitioned into $n \geq 1$ disjoint subsets T_1, T_2, \dots, T_n , where each T_i ($i = 1, 2, \dots, n$) is a tree and T_1, T_2, \dots, T_n are called subtrees of the root. Each node of the proposed ontology tree is a quadruple of the form $\langle c, m, o, s \rangle$, where c is a *concept*, m is an optional modifier representing degree of expressiveness of opinion word o , and s is the sentiment information referring *positive*(+), *negative*(-), or *neutral*(=) orien-

tation. Figure 1 shows an exemplar sentiment propagation ontology tree generated using the documents from iPod domain, wherein opinion words and subsequently sentiment information are missing for the concept node *software* and its child node *interface*. Propagating sentiment in a bottom-up manner, the sentiment orientation of the node *interface* can be derived as *positive*, since both of its left and right child nodes at depth one (i.e., *Apps* and *Menu*) are referring positive sentiment. Similarly, sentiment orientation of the concept node *software* can be derived as *positive*.

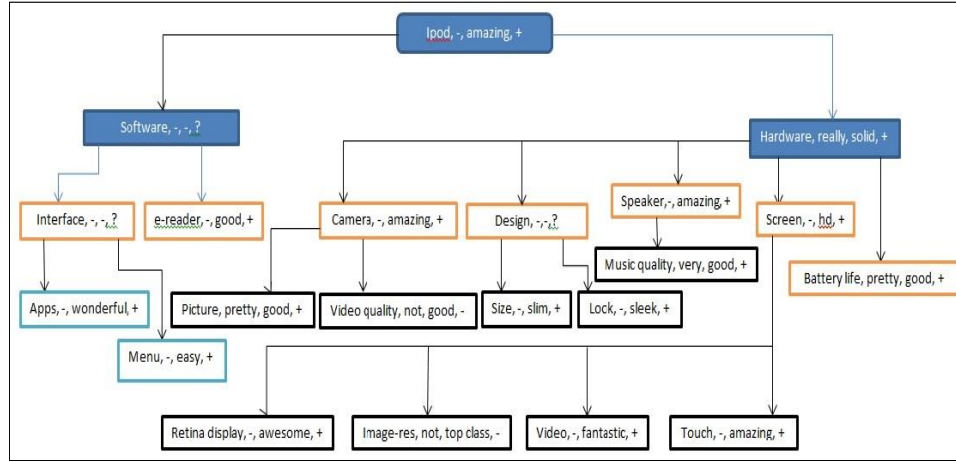


Fig. 1: Sentiment propagation using ontology tree

4 Review Summarization and Visualization

This section presents the design of Opinion Summarization and Visualization (OSV) module to present extracted information components in a graphical form which facilitates users to have a quick glance over the mined concepts, opinions, and sentiments polarity without reading the pile of review documents. As mentioned in one of our previously published papers [45], OSV module is capable to visualize mining results both from single as well as multiple review documents. It also provides graphical tools for end users to explore and visualize summarized information components using bar charts and pie charts generated using Google chart API⁵.

Extracted information components along with opinion summary statistics are presented using Java Script Object Notation (JSON)⁶ object, which is a language independent, lightweight text-data interchangeable format. Figure 2 shows the JSON

⁵ http://code.google.com/apis/ajax/playground/#chart_wrapper

⁶ <http://www.json.org/>

representation of an object describing information component and opinion summary statistics. The object uses string fields for feature, modifier, opinion, and sentiment polarity; a number field for reliability score, and contains an array of objects for opinion score. During execution, OSV retrieves all required information from database to form JSON object and uses the same as an input for visualization purpose. Figure 3 (taken from one of our previous research paper [45]) shows the main screen of OSV module, consisting of two rows viz. the upper-row and the lower-row. The upper-row is divided into three panels – upper-left, upper-middle, and upper-right. The upper-left panel contains list of reviews crawled from merchant sites. When a user selects a particular review from the upper-left panel, its description and metadata appear in the upper-middle and upper-right panels, respectively. Metadata of a review consists of information such as source from where the review was crawled, domain, author name, description, date of posting, and star rating. The lower-row of the main screen is also divided into two panels viz. lower-left and lower-right. The lower-left panel uses pie chart for opinion summarization of a particular review selected by the user from upper-left panel. The pie chart makes use of different colour combination mainly blue, red, and green to visualize the number of positive, negative, and neutral opinions, respectively. For a selected review document, the lower-right panel presents the list of extracted results that includes feature, modifier (if any), opinion, sentiment polarity (positive, negative, or neutral) and opinion indicator of the feature-opinion pair. For visibility purpose, colour scheme is used to highlight the information components extracted from review documents. On clicking a feature word appearing in the lower-right panel, the constituents of the corresponding feature-opining pair is highlighted using orange and yellow colors, respectively, and the relevant snippets (containing feature and opinion words) of the review document also accentuates. When a user clicks to a highlighted snippet representing product feature in upper-middle panel, a pop-up window appears visualizing the percentage of positive, negative, and neutral opinions using pie-chart.

Figure 4 (taken from one of our previous research paper [45]) shows the percentage of opinions expressed on a product feature from a corpus of customer reviews. As discuss earlier, OSV module facilitates users to navigate through the pile of customer reviews in an efficient way to produce feature-based opinion summary. Thus, pop-up window appearing in the above mentioned step contains a *view more* option, clicking which causes the window to expand in size, visualizing opinion score summary for the respective product feature. Figure 5 shows an expanded pop-up window, where size of each slice in the 3D pie-chart represents the degree of expressiveness of opinion. Opinion scores are calculated using Chi-square value due to its best performance. Higher the score for an opinion, larger the size of the slice in 3D pie-chart.

```

{
  "concept": "Speaker quality",
  "modifier": "very",
  "opinion": "bad",
  "scoreReliabilityPair": 0.0108,
  "scoreOpinion": [
    {
      "type": "pmi",
      "number": -0.7344
    },
    {
      "type": "mi",
      "number": -109.3725
    },
    {
      "type": "chi",
      "number": -850.0066
    }
  ],
  "orientation": "negative"
}

```

Fig. 2: JSON representation of information components and opinion scores

5 Experimental Setup and Results

In this section, we present the experimental results of the proposed opinion mining and sentiment analysis system. The data set used in our experiment consists of 1200 review documents related to different electronic products crawled from www.amazon.com. Information component extraction mechanism is implemented using a rule-based system. Reliability score is calculated for each concept and associated opinion using a reliability score generator implemented in Java. A partial list of feasible concepts along with their opinions and modifiers are presented in Table 5. After analysis, we found that occurrence frequency of genuine concepts is very high in review documents, which is due to the tendency that various reviewers refer to same product feature with different opinion words to express distinct sentiments.

An opinion score generator implemented in Java is applied to compute the scores for the feasible opinion words. Table 6 presents statistical feature values for a partial list of opinion words. Thereafter, each opinion word is characterized using the features described in section 3.4.2. It can be observed from Table 6 that majority of the opinion scores obtained using Log Likelihood Ratio (LLR) is found negative. Thus, it is discarded from further analysis. Subsequently, a feature vector generator

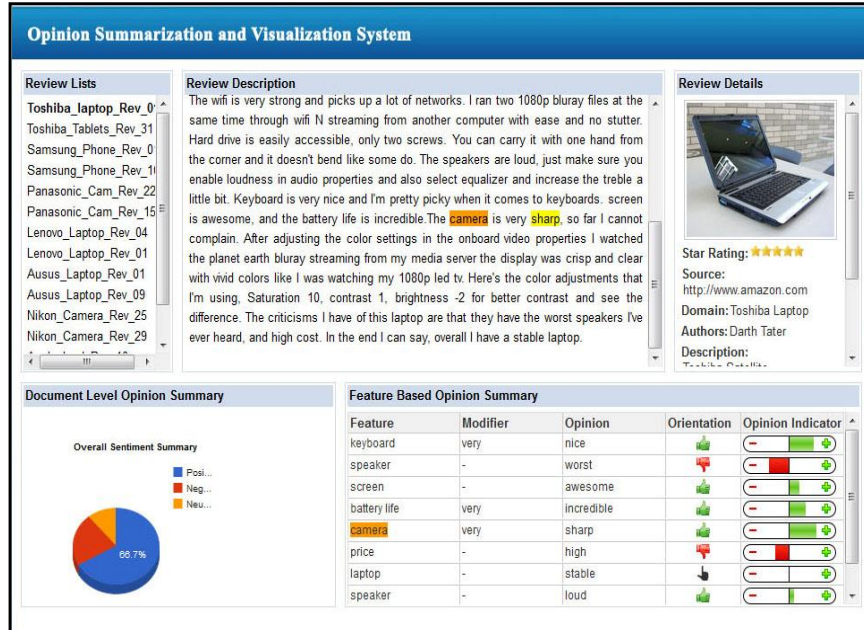


Fig. 3: Opinion summarization and visualization using OSV module

is implemented in Java to generate feature values for each opinion bearing word, and a ternary classification model is learned to classify polarity of a word as positive, negative, or neutral. Evaluation of the experimental results is performed using standard Information Retrieval (IR) metrics *Precision*, *Recall*, and *F-score* which are defined in equations 11, 12, and 13, respectively. In these equations, TP represents *true positives*, FP represents *false positives*, and FN represents *false negative*.

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F-Score = \frac{2 \times precision \times recall}{precision + recall} \quad (13)$$

5.1 Evaluating Concept and Opinion Extraction Process

To the best of our knowledge, no benchmark data is available in which concepts and related opinion information are marked simultaneously for electronic product

Table 5: A partial list of candidate information components including concepts, opinions, and modifiers

Product	Concept	Modifier	Opinion
Digital Camera	lens	too	heavy
	battery	-	standard
	picture	really	great
	price	very	reasonable
	access	-	quicker
IPod	ipod touch	not	great
	retina display	-	beautiful
	sound	-	great
	resolution	-	fine
	battery life	very	good
Laptop	screen	too	glossy
	keyboard	very	nice
	software	-	expensive
	battery life	-	incredible
	picture quality	just	great
Cell Phone	OS	-	beautiful
	touch screen	-	awesome
	feel	-	great
	picture	quite	good
	battery life	not	bad
Tablet	scrolling	very	smooth
	processor	-	faster
	screen	-	beautiful
	file transfer	-	painless
	speaker	really	good

Table 6: A partial list of opinionated words and their opinion scores obtained using different statistical measures

Opinionated word	Opinion score			
	PMI	MI	Chi-square	LLR
Bad	-0.7344	-109.3725	-850.0066	-10419.1723
Expensive	-0.2984	-51.8493	-556.9257	-11182.3347
Poor	-0.5560	-54.0277	-378.8968	-2483.0364
Slow	-0.6935	-66.1516	-369.4841	-6.5339
Horrible	-0.9389	-34.7363	-240.8866	-8247.2498
Bittersweet	0.0000	0.0000	0.0000	0.0000
Unbelievable	0.0000	0.0000	0.0000	0.0000
Amazing	2.3403	172.7484	1177.2141	-47112.7543
Bright	0.3932	47.6693	245.7392	-21448.2128
Beautiful	1.0260	76.5412	485.5965	-15660.7247
Fantastic	1.4603	98.6912	607.5986	-24541.5964
Wonderful	2.0459	75.8369	419.2278	-10558.8426

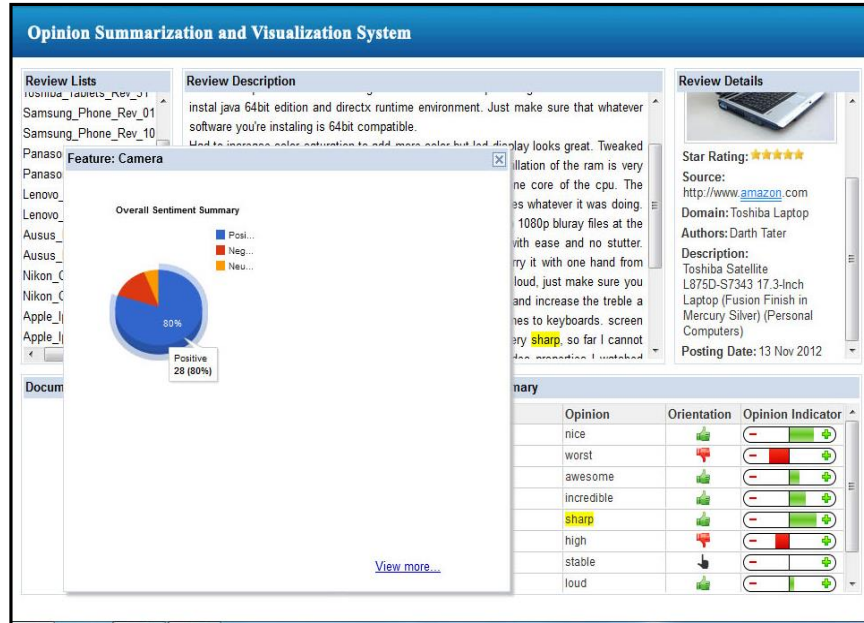


Fig. 4: Feature-based opinion summarization using OSV module

domain. Therefore, manual evaluation is performed to analyze the overall performance of the proposed system. From a corpus of 1200 review documents, a total of 120 documents (Cell Phone: 30, Laptop: 30, iPod: 30, Digital Camera: 15, and Tablet: 15) are randomly selected for testing purpose. Our rule-based method is applied to extract concepts and related opinion words. Initially, the total count obtained for TP, FP, and FN are 841, 866, and 450, respectively. We have observed that direct and strong relationship between words causes extraction of nouns, verbs, and adjectives, representing irrelevant concepts and opinion words. As a result, the number of false positives (FP) increases, which has an adverse effect on the precision value. To overcome this problem, post-processing step is applied to remove noisy extraction by performing feasibility analysis discussed in section 3.3.2. After removal of the noisy concepts and associated opinions, the total count of FP reduces to 273. Macro-averaged values are calculated to present a synthetic performance measure by simply averaging the results of different categories of the products. Table 7 summarizes the performance measure values in the form of a misclassification matrix.

Since most of the reviewers use informal approach while commenting, reviews are generally lack in grammatical correctness and pose a number of challenges for natural language parser. The recall value is lower than the precision is an indication of system inability to extract certain concepts and associated opinions correctly.

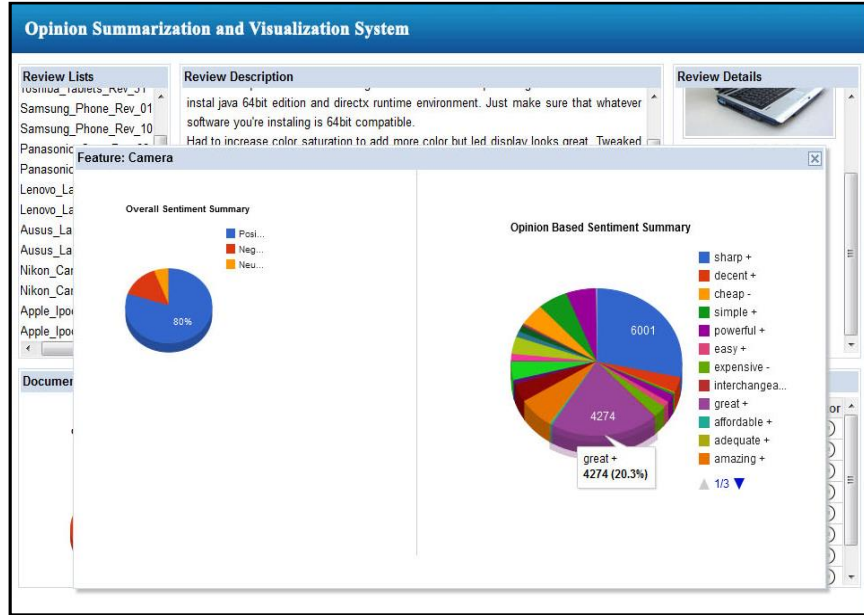


Fig. 5: Opinion-based sentiment summary using OSV module

Table 7: Performance evaluation of information component extraction process

Product Category	TP	FP	FN	Precision	Recall	F-Score
Cell Phone	181	51	120	78.02	60.14	67.92
Laptop	319	37	99	89.61	76.32	82.43
Camera	104	23	68	81.89	60.47	69.57
IPod	129	124	96	50.99	57.34	53.98
Tablet	108	38	67	73.98	61.72	67.29
Macro-Average	841	273	450	75.50	65.15	69.94

5.2 Evaluating Sentiment Classification Process

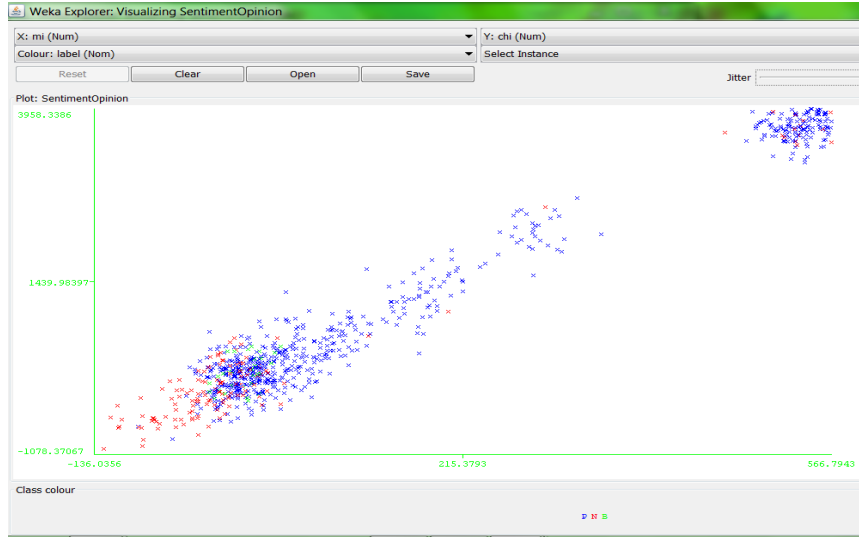
In our experiment, we have considered 1000 and 200 opinion bearing words for training and testing purpose, respectively. A feature vector generator is implemented in Java to generate numeric values of the features for each opinion bearing word. Table 8 shows the information gain ranking of attributes on the basis of WEKA's attribute evaluator, in which the Chi-square feature seems to be most discriminative followed by Mutual Information (MI).

Figure 6 visualizes the effect of classifying opinion bearing words on the basis of Chi-square (χ^2) and MI scores placed on X-axis and Y-axis, respectively.

We have experimented with some prominent classifiers that are best suited for the classification task, but settled with Bagging and Decision Tree (J48) algorithm

Table 8: A ranked list of sentiment classification features based on information gain values

Attribute Name	Information gain
Chi-square	0.3609
MI	0.3542
PMI	0.1680
Negation	0.0657
TF-IDF	0.0236
Modifier	0.0103

Fig. 6: Visualization of sentiment classification results based on χ^2 and MI values

due to their best performance during training and testing phases. Figure 7 presents Receiver Operating Characteristics (ROC) curves of all four classifiers, visualizing their comparative accuracy in terms of true positive rate and false positive rate during training. The ROC curve is generated using WEKA by plotting false positive rate and true positive rate on X-axis and Y-axis, respectively. Best classification performance of Bagging (curve consisting of + symbol) can be observed easily due to its appearance at the extreme left and higher in the ROC space.

A ternary classification model is learned to determine the class of an opinion word as *positive*, *negative*, or *neutral*. To judge the overall performance of the classifiers used in our experiment, Weighted Average Precision ($P_{\omega a}$), Weighted Average Recall ($R_{\omega a}$), and Weighted Average F-score ($F_{\omega a}$) values are calculated using equations 14, 15, and 16, respectively. In these equations, P_{pos} , R_{pos} , and F_{pos} represent the Precision, Recall, and F-score, respectively of the positive instances; P_{neg} , R_{neg} , and F_{neg} represent the Precision, Recall, and F-score, respectively of the nega-

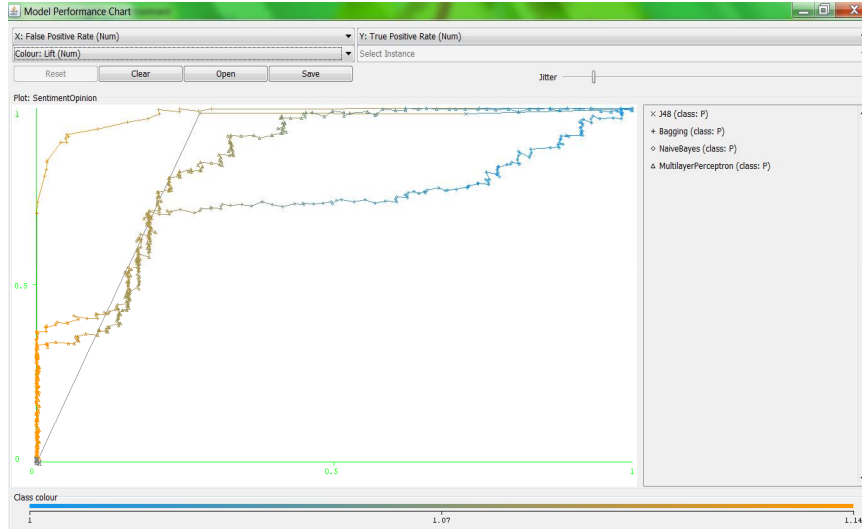


Fig. 7: ROC curves of classifiers for sentiment analysis on training data set

tive instances; and P_{neu} , R_{neu} , and F_{neu} represent the Precision, Recall, and F-score, respectively of the neutral instances. Similarly, ω_{pos} , ω_{neg} , and ω_{neu} represent the weights (ratio of the number of instances of a particular category to the total number of instances) of the positive, negative, and neutral instances, respectively. Table 9 presents the comparison results of the classifiers in terms of these metrics calculated over the training and test data sets.

Table 9: Comparison results of classifiers based on weighted average precision, weighted average recall, and weighted average f-score over training and test data sets

Classifier	Weighted average result (On training data set)			Weighted average result (On test data set)		
	Precision	Recall	F-Score	Precision	Recall	F-Score
NB	0.902	0.703	0.779	0.840	0.617	0.683
J48	0.935	0.948	0.939	0.885	0.872	0.855
MLP	0.919	0.940	0.929	0.821	0.857	0.833
Bagging	0.952	0.951	0.942	0.875	0.865	0.848

$$P_{\omega a} = \frac{(P_{pos} \times \omega_{pos}) + (P_{neg} \times \omega_{neg}) + (P_{neu} \times \omega_{neu})}{\omega_{pos} + \omega_{neg} + \omega_{neu}} \quad (14)$$

$$R_{\omega a} = \frac{(R_{pos} \times \omega_{pos}) + (R_{neg} \times \omega_{neg}) + (R_{neu} \times \omega_{neu})}{\omega_{pos} + \omega_{neg} + \omega_{neu}} \quad (15)$$

$$F_{\omega a} = \frac{(F_{pos} \times \omega_{pos}) + (F_{neg} \times \omega_{neg}) + (F_{neu} \times \omega_{neu})}{\omega_{pos} + \omega_{neg} + \omega_{neu}} \quad (16)$$

6 Conclusion and Future Works

Rapid growth in user-generated contents, mainly unstructured or semi-structured textual data, on the Web and their uncontrolled generation containing various natural language nuances provides an opportunity for data analysts to apply computational techniques for determining users' sentiments and emotions with respect to various features of products. Despite various research efforts attempted worldwide, development of a fully automatic opinion mining and sentiment analysis system is still elusive. It has been observed that overall problems associated with the development of opinion mining and sentiment analysis system is non-trivial and requires more research exploration. The main contribution of this work is to propose linguistically motivated computational paradigm for ontology learning and sentiment analysis using rule-based and machine learning approaches. Natural language processing techniques along with statistical analysis are applied to identify feasible product features and opinions, whereas machine learning techniques are applied for sentiment polarity determination. The electronic products ontology developed in this work, helps to structure and manage product features, and improves the performance of sentiment analysis task by propagating sentiment information from lower to higher levels among related concepts. Consideration of other structural information of ontology can be used for sentiment propagation along ontology tree, which is one of our future directions of work. Moreover, handling implicit concepts that are very common with customer reviews is an area in which we wish to direct our future research.

References

1. B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool, 2012.
2. E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.
3. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.
4. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of EMNLP*, 2002, pp. 79–86.

5. T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? finding strong and weak opinion clauses," in *Proceedings of the 19th national conference on Artificial intelligence*, 2004, pp. 761–767.
6. M. Abulaish, Jahiruddin, M. N. Doja, and T. Ahmad, "Feature and opinion mining for customer review summarization," in *Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence (PReMI'09)*, 2009, pp. 219–224.
7. B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.
8. M. Abulaish, "An ontology enhancement framework to accommodate imprecise concepts and relations," *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, pp. 22–36, 2009.
9. V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, 1997, pp. 174–181.
10. A. Cadilhac, F. Benamara, and N. Aussenac-Gilles, "Ontolexical resources for feature-based opinion mining: a case-study," in *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, 2010, pp. 77–86.
11. W. Wei and J. A. Gulla, "Sentiment learning on product reviews via sentiment ontology tree," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 404–413.
12. J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.
13. P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 417–424.
14. B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, Taylor and Francis Group, Boca, 2010, pp. 627–666.
15. A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP05)*, 2005, pp. 339–346.
16. L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, 2006, pp. 43–50.
17. J. S. Kessler and N. Nicolov, "Targeting sentiment expressions through supervised ranking of linguistic configurations," in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, 2009, pp. 90–97.
18. B. Li, L. Zhou, S. Feng, and K.-F. Wong, "A unified graph model for sentence-based opinion retrieval," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 1367–1375.
19. K. Bloom, N. Garg, and S. Argamon, "Extracting appraisal expressions," in *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, 2007, pp. 308–315.
20. G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.
21. X. Cheng and F. Xu, "Fine-grained opinion topic and polarity identification," in *Proceedings of the Poster Session of the Sixth International Conference on Language Resources and Evaluation*, 2008, pp. 2710–2714.
22. L. Zhou and P. Chaovalit, "Ontology-supported polarity mining," *Journal of the American Society for Information Science and technology*, vol. 59, no. 1, pp. 98–110, 2008.
23. R. Y. Lau, C. C. Lai, J. Ma, and Y. Li, "Automatic domain ontology extraction for context-sensitive opinion mining," in *Proceeding of the 13th International Conference on Information Systems*, 2009, pp. 35–53.

24. K. P. P. Shein, "Ontology based combined approach for sentiment classification," in *Proceedings of the 3rd International Conference on Communications and information technology*, 2009, pp. 112–115.
25. S. Mukherjee and S. Joshi, "Sentiment aggregation using conceptnet ontology," in *Proceedings of the 6th International Joint Conference on Natural Language Processing*, 2013, pp. 570–578.
26. E. Kontopoulou, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4065–4074, 2013.
27. B. Fortuna, M. Grobelnik, and D. Mladenic, "Ontogen: Semi-automatic ontology editor," in *Proceeding of the 2007 Conference on Human Interface*, 2007, pp. 309–318.
28. G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation," in *Proceedings of the 21st international joint conference on Artificial intelligence*, 2009, pp. 1199–1204.
29. A. Kamal, M. Abulaish, and T. Anwar, "Mining feature-opinion pairs and their reliability scores from web opinion sources," in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, 2012, pp. 15:1–15:7.
30. A. Kamal and M. Abulaish, "Statistical features identification for sentiment analysis using machine learning techniques," in *Proceeding of the 2013 International Symposium on Computational and Business Intelligence*, 2013, pp. 178–181.
31. B. Li, L. Zhou, S. Feng, and K. Wong, "A unified graph model for sentence-based opinion retrieval," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 1367–1375.
32. F. Li, Y. Tang, M. Huang, and X. Zhu, "Answering opinion questions with random walks on graphs," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 2009, pp. 737–745.
33. J. Otterbacher, G. Erkan, and D. R. Radev, "Using random walks for question-focused sentence retrieval," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 915–922.
34. J. Allan, C. Wade, and A. Bolivar, "Retrieval and novelty detection at the sentence level," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003, pp. 314–321.
35. G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
36. R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.
37. K. W. Church and P. Hank, "Word association norms, mutual information, and lexicography," *Computational Linguistics*, vol. 16, no. 1, pp. 22–29, 1990.
38. R. M. Fano, "Transmission of information: A statistical theory of communications," *American Journal of Physics*, vol. 29, no. 11, pp. 793–794, 1961.
39. T. Dunning, "Accurate methods for the statistics of surprise and coincidence," *Computational Linguistics*, vol. 19, no. 1, pp. 61–74, 1993.
40. N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.
41. J. R. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1993.
42. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1*. MIT Press Cambridge, MA, USA, 1986, ch. Learning internal representations by error propagation, pp. 318–362.
43. L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
44. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

45. A. Kamal, "Review mining for feature based opinion summarization and visualization," *International Journal of Computer Applications*, vol. 119, no. 17, pp. 6–13, 2015.