Sentiment Analysis of Reviews

Subhabrata Mukherjee
Max Planck Institute for Informatics
Germany
smukherjee@mpi-inf.mpg.de

ACM Reference Format:
Subhabrata Mukherjee. 2017. Sentiment Analysis of Reviews. In *Encyclopedia of Social Network Analysis and Mining, Springer New York, 2017 (ESNAM).*
https://doi.org/10.1007/978-1-4614-7163-9_110169-1

SYNONYMS

Opinion mining; Social media analysis; Aspect or facet mining; Knowledge discovery; Data mining; Sentiment classification;

GLOSSARY

SA Sentiment Analysis
SO: Semantic Orientation
PMI: Point-wise Mutual Information
NB: Naive Bayes classifier
SVM: Support Vector Machines
MaxEnt: Maximum Entropy Classifier
POS: Part of Speech
KB: Knowledge-bases
JST: Joint Sentiment Topic Model
JAST: Joint Author Sentiment Topic Model

DEFINITION

Sentiment Analysis (SA) of reviews refers to the task of analyzing natural language text in forums like Amazon, TripAdvisor, Yelp, IMDB etc. to obtain the writer's feelings, attitudes, and emotions expressed therein towards a particular topic, product, or entity. It involves overlapping approaches in several domains like Natural Language Processing (NLP), Computational Linguistics (CL), Information Extraction (IE), Text Mining, and Machine Learning (ML).

INTRODUCTION

In recent years, the explosion of social networking sites (e.g., Facebook, Twitter), blogs (e.g., Mashable, Techcrunch), and online review portals (e.g., Amazon, TripAdvisor, IMDB) provide overwhelming amount of information about products and services. Millions of people express uninhibited opinions about various product (and, service) features, and their nuances. As consumers cannot test the functionality of a product (or, service) prior to consumption, these reviews help them make an informed decision to buy the product (or, service) or not. Sentiment Analysis aims to tap this goldmine of information by analyzing the vast repository of largely unstructured text — in the form of reviews, comments, questions, and requests etc. — to retrieve users' opinions about featured products and services. It can be performed at different levels of *granularity* as follows. Consider the following movie review:

Example 1.1. "This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story. However, the film does not quite make the mark due to sloppy acting."

- *Document-level* SA [Pang 2002, Dave 2003] aims to find the *overall* polarity of the review as *positive* or *negative*, depending on whether its author liked it or not (for instance, it is *negative* in Example 1.1).
- *Sentence-level* SA [Yu 2003, Mukherjee 2012a] analyzes individual sentences in the review to find its polarity. For example, the second sentence in Example 1.1 depicts a *positive* sentiment about the movie.
- *Phrase-level* SA [Wu 2009, Mukherjee 2012a] analyzes individual phrases to determine its polarity. For example, "does not quite make the mark" depicts a *negative* sentiment. This is particularly interesting, in case of *negation* operators in the scope, that flips the polarity of the phrase: for instance, contrast "not only good..." (positive) with "not good" (negative).
- *Facet-level* SA [Lin 2009, Mukherjee 2014a, Lakkaraju 2011] finds the polarity of a review with respect to the underlying *facets* (attributes, features, or aspects) of the item in the review under consideration. For instance, the sentiment in Example 1.1 is *positive* w.r.t. the movie facets "plot" and "director"; however, it is *negative* w.r.t. "acting".
- *Author-level* SA [Mukherjee 2014a, Mukherjee 2013a] aims to find the polarity of a review w.r.t. the *person* who wrote the review. Since reviews are subjective, the same review may attain different polarities depending on the preferences of the person who authored it.

Considering all of the different aspects outlined above, sentiment (or, opinion) can be defined as:

Definition 1.2. "Sentiment (or, opinion) is a quintuple: $\langle o_j, f_{jk}, so_{ijkl}, h_i, t_l \rangle$, where o_j is a target object (entity or item), f_{jk} is a feature (facet/aspect) of the object o_j, so_{ijkl} is the sentiment on feature f_{jk} of object o_j, h_i is the opinion holder, and t_l is the time when the opinion is expressed by h_i. The sentiment so_{ijkl} is +ve, -ve, or neutral, or expressed with different strength/intensity levels, e.g., 1 to 5 stars as used by most review sites on the Web." [Pang 2008]

ESNAM, 2017, Springer New York
2017. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1007/978-1-4614-7163-9_110169-1

KEY POINTS

This work gives a broad overview of the various approaches, and state-of-the-art systems for sentiment analysis of reviews. It focuses on three key areas: (i) commonly used lexical resources for SA, (ii) a broad overview of the various features, and aspects of SA, and (ii) some prominent Machine Learning approaches to SA with a brief overview of some of the commonly used classifiers and techniques.

HISTORICAL BACKGROUND

Early works in sentiment analysis starting with the seminal work of [Pang 2002, Turney 2002, Dave 2003, Yu 2003, Pang 2004b] considered reviews as bag-of-words, and focused on classifying them as positive, negative, or neutral using classifiers. Later works developed more sophisticated features based on phrasal and dependency relations, narratives, perspectives, lexical resources etc. The flurry of activity in this domain, in recent times, can be attributed to the availability of large-scale datasets with the explosion of social networking sites and online review portals, and significant advances in Machine Learning algorithms for Natural Language Processing, Text Mining, and Information Extraction tasks.

LEXICAL RESOURCES

The simplest approach to Sentiment Analysis is to consider *word-level* features. Given a review document with a sequence of words, and access to a lexical resource containing the *annotated* polarity (positive, negative, or objective) of opinion words and phrases — a count-based approach assigns the majority polarity of the opinion words in the document as the polarity of the review. The following are some commonly used lexical resources for SA:

- SentiWordNet [Esuli 2006] is a lexical resource, where each wordnet synset (i.e., sense) s is associated to three numerical scores (each ranging from $[0-1]$ on a simplex) — Obj(s), Pos(s) and Neg(s) — denoting the corresponding polarity scores of the synset. This corresponds to a graded evaluation of opinion, as opposed to a hard one. For instance, the synset "estimable", with the sense "may be computed or estimated", has an Obj score of 1.0, and corresponding Pos and Neg scores as 0.0; whereas its Obj, Pos and Neg scores corresponding to the sense "deserving of high respect or high regard" are 0.25, 0.75, and 0 respectively.
- Subjectivity lexicon [Wiebe 2004] is a resource that annotates words with tags like parts-of-speech, prior polarity, magnitude of prior polarity (weak / strong) etc. The prior polarity can be positive, negative or neutral.
- Inquirer [Stone 1966] is a list of words marked as positive, negative and neutral.
- Taboada [Taboada 2004] is a word-list that gives a count of collocations with positive and negative seed words. A word closer to a positive seed word is predicted to be positive, and vice versa.
- Bing Liu Sentiment Lexicon [Hu 2004] contains a list of manually annotated positive and negative opinion words.

FEATURES AND ASPECTS OF SENTIMENT ANALYSIS

Feature engineering is a basic and essential task for most Machine Learning based approaches to Sentiment Analysis. Converting a piece of review text to a feature vector is the basic step in any data driven approach to SA. In the following section, we will see different aspects and features for Sentiment Analysis.

Term Presence vs. Term Frequency. Term (or, word) frequency has always been considered essential in traditional Information Retrieval (IR) and Text Classification tasks. [Pang 2002] empirically found term presence to be more important to Sentiment Analysis than term frequency. This involves binary-valued feature vectors, with entries indicating the presence or absence of words. Unlike in traditional text classification tasks, the presence of a strong sentiment bearing word can change the overall polarity of a sentence in SA. It has also been seen that the occurrence of rare words contain more information than frequently occurring words — a phenomenon called Hapax Legomenon.

Term Position. Words appearing in certain positions in the text carry more sentiment or weight than words appearing elsewhere. This is similar to IR where words appearing in topic titles, subtitles or abstracts etc. are given more weight than those appearing in the body. For instance, the review text in Example 1.1 contains more positive words than negative; however, the presence of negative sentiment in the *concluding* sentence makes the overall sentiment of the review *negative*. This is a typical example of *thwarting* which has been investigated in some recent works [Mukherjee 2014b, Ramteke 2013].

N-grams. N-grams are capable of capturing context in texts, and are widely used in Natural Language Processing tasks. Whether higher order N-grams are increasingly useful is a matter of debate. [Pang 2002] reported unigrams to outperform bigrams when classifying movie reviews by sentiment polarity; however, [Dave 2003] found that in some settings bigrams and trigrams perform better.

Part-of-Speech. Part-of-Speech information is most commonly exploited in all NLP tasks. One of the important reasons is that they provide a crude form of Word Sense Disambiguation (WSD). Adjectives have been used most frequently as features amongst different parts-of-speech. There is a strong correlation between *Adjectives* and subjectivity in SA. Although many part-of-speech tags are important, most of the common sentiment bearing words are Adjectives. [Pang 2002] achieved an accuracy of around 82.8% in movie review domains using only Adjectives as features.

Apart from using only Adjectives, the *Adjective-Adverb* combination is also informative. Most of the Adverbs have no prior polarity. But when they occur with sentiment bearing Adjectives, they can play a major role in determining the sentiment of a sentence. [Benamara 2007] demonstrated how Adverbs can alter the sentiment value of the Adjectives that they are used with. For example, the sentiment orientation of 'immensely good' is more positive than that of using only the Adjective 'good'. Similarly 'barely good' is more negative than the positive sentiment-bearing Adjective 'good'. This fine-grained analysis can be used to assign graded polarity scores to sentiment-bearing words and phrases, rather than assigning a hard polarity label as positive or negative.

Semantic Orientation with PMI. Semantic Orientation (SO) [Turney 2002] refers to a real number measure of the positive or negative sentiment expressed by a word or phrase. The SO of a phrase is determined based on the phrase's Pointwise Mutual Information (PMI) with the words "excellent" and "poor". PMI is defined as follows: $PMI(w_1, w_2) = log_2(p(w_1 \& w_2)/p(w_1)p(w_2))$.

The SO for a phrase is the difference between its PMI with the word "excellent" and its PMI with the word "poor". This yields values above zero for positive sentiment-bearing phrases, and below zero for negative ones. A SO value of zero would indicate a neutral semantic orientation.

Discourse and Modalities. "An important component of language comprehension in most natural language contexts involves connecting clauses and phrases together in order to establish a coherent discourse" [Wolf 2004]. The presence of linguistic constructs like connectives, modals, conditionals and negation can alter sentiment at the sentence level as well as the clausal or phrasal level.

Example 1.3. "I'm quite excited about Tintin, despite not really liking original comics. Probably because Joe Cornish had a hand in."

The overall sentiment of Example 1.3 is *positive*, although there are an equal number of positive and negative sentiment-bearing words. This is due to the connective "despite" which gives more *emphasis* on the previous discourse segment. Any bag-of-words model would be unable to classify this sentence without considering the discourse marker "despite". [Mukherjee 2012b] probes the influence of these discourse markers and semantic operators (like, modals and negations) for fine-grained sentiment analysis of sentences at the *discourse-level*. They outline a lightweight approach to incorporate these insights as rules and features in a simple classifier that is effective for SA, even in noisy mediums (e.g., Twitter with short and noisy texts), where heavy-weight approaches based on *Parsing* typically fail..

Parsing. In dependency grammar, structure is determined by the relation between a head and its dependents. The dependent is a modifier or complement, and the head plays a more important role in determining the behaviors of the pair. Dependency parsing captures short range and long range dependencies between words in a sentence.

Example 1.4. "I have an iPod and it is a great buy but I'm probably the only person that dislikes the iTunes software."

The review in Example 1.4 is *positive* w.r.t. "iPod", but *negative* w.r.t. "iTunes". Feature-specific SA finds the polarity of a review with respect to the target facets or features in the review. [Mukherjee 2012a] uses dependency parsing (e.g., Stanford Dependency Parser) for *facet-level* SA to capture the relations between the facets (aspects, or features of items in a review) and their associated opinions. The idea is to capture the association between any specific facet, and the expressions of opinion that come together to describe that feature. However, not all dependency relations are important for SA. The authors show that dependency relations like "nsubj, dobj, advmod, amod, neg, prep_of, acomp, xcomp, conj_and_ccomp, iobj" etc. are more important for SA than other relations.

[Wu 2009] uses *phrase dependency parsing* for opinion mining. This approach trades-off the information loss of the word-level dependency in Dependency Parsing, as it does not explicitly provide local structures and syntactic categories of phrases, and the information gain in extracting long distance relations. Hence they extend the Dependency Tree node with phrases.

[Chen 2010] uses dependency parsing and shallow semantic analysis for Chinese opinion related expression extraction. They categorize relations as: topic and sentiment located in the same sub-sentence and quite close to each other (like the rule "an adjective plus a noun" is mostly a potential opinion-element relation), topic and sentiment located in adjacent sub-sentences and the two sub-sentences are parallel in structure (that is to say, the two adjacent sub-sentences are connected by some coherent word, like although/but, and, etc.), topic and sentiment located in different sub-sentences, either being adjacent or not, etc.

World Knowledge: Encyclopedic, Semantic, Ontological Resources. Sentiment Analysis often requires access to external or background knowledge (e.g., Knowledge-Bases (KB) like the Knowledge Graph (KG) [Dong 2014]) to understand entity-specific concepts in the text as in Example 1.5. This *objective* synopsis of the movie will be misclassified as *negative* using any lexical resource for word-level SA due to the presence of the word "dangerous" — which is a part of the movie plot.

Example 1.5. "L.I.E. stands for Long Island Expressway, which slices through the strip malls and middle-class homes of suburbia. Filmmaker Michael Cuesta uses it as a metaphor of dangerous escape for his 15-year old protagonist, Howie (Paul Franklin Dano)."

[Mukherjee 2012c] uses Wikipedia to understand concepts specific to movies like the movie plot, fictional characters, objective facts about the crew and cast, their past performance, and characteristics of the movie genre. This is used for extractive summarization to derive relevant subjective (opinionated) sentences significant for sentiment analysis of the movie review, and not objective facts about the movie.

[Yu 2003] proposes to find subjective sentences using lexical resources where the authors hypothesize that subjective sentences will be more similar to opinionated sentences than to factual sentences. As a measure of similarity between two sentences they used different measures including shared words, phrases and the WordNet. [Potthast 2010] focuses on extracting top sentiment keywords based on Pointwise Mutual Information (PMI) measure [Turney 2002].

The pioneering work for subjectivity detection is done in [Pang 2004a], where the authors use *min-cut* to leverage coherency between the sentences. The fundamental assumption is that local proximity preserves the objectivity or subjectivity relation in the review. [Agarwal 2005] integrates graph-cut with linguistic knowledge in the form of WordNet to exploit similarity in the set of documents to be classified.

Example 1.6. "I bought a Canon EOS 7D (DSLR). It's very small, sturdy, and constructed well. The handling is quite nice with a powder-coated metal frame. It powers on quickly and the menus are fairly easy to navigate. The video modes are nice, too. It works great with my 8GB Eye-Fi SD card. A new camera isn't worth it if it doesn't exceed the picture quality of my old 5Mpixel SD400 and this one doesn't. The auto white balance is poor. I'd need to properly balance every picture taken so far with the ELPH 300. With 12 Mpixels, you'd expect pretty good images, but the problem is that the ELPH 300 compression is turned up so high that the sensor's acuity gets lost (softened) in compression."

Review Example 1.6 depicts the complexity involved in analyzing product reviews. The review has a mix of good and bad comments about various features of the product. A flat classification model which considers all features to be equally important will fail to capture the proper polarity of the review. The reviewer seems happy with the "camera size, structure, easy use, video modes, SDHC support" etc. However, the "auto-white balance" and "high compression" leading to "sensor acuity" seem to disappoint him. Now, the primary function of a camera is to take good pictures and videos. Thus "picture, video quality, resolution, color balance" etc. are of primary importance, whereas "size, video mode, easy use" etc., are secondary in nature. The overall review polarity, in this example, should be *negative* as the reviewer shows concerns about the most important features of the camera.

In order to analyze this review, we not only need to understand the different underlying *facets*, but also their *relations*. This can be captured using *ontological* information — where, the concepts are nodes, and edges between them capture relations. Typically, the concepts higher up in the ontology are more important than the concepts further down the hierarchy.

[Wei 2010] propose a hierarchical learning method to label a product's attributes and their associated sentiments in product reviews using a Sentiment Ontology Tree (HLSOT). The HLSOT approach is supervised, requiring the reviews to be annotated with product attribute relations, as well as feature-specific opinion expressions. [Mukherjee 2013b] use ConceptNet [Liu 2004] as a knowledge resource to automatically construct a domain-specific ontology tree for product reviews, without requiring any labeled training data. They present a novel sentiment aggregation approach to combine the feature-specific polarities with ontological information to find the overall polarity of the review. ConceptNet relations have an inherent structure which helps in the construction of an ontology tree from the resource.

MACHINE LEARNING APPROACHES

In the previous section, we reviewed various aspects of Sentiment Analysis and different classes of features that are commonly used, like n-grams, part-of-speech, parsing (e.g., dependency relations like "nsubj, dobj, advmod" and "amod"), sentiment lexicons for prior polarity of words, discourse and modality features etc. Now given access to all of these features, we can create a feature vector $f = \{f_1, f_2, \cdots, f_m\}$ (considering m number of features) for each review text, and use a Machine Learning Classifier to classify it as *positive* or *negative*. We now review some commonly used classifiers for Sentiment Analysis. Starting with the seminal work of [Pang 2002, Pang 2004a], the most commonly used Machine Learning classifiers for SA have been Naive Bayes, Maximum Entropy and Support Vector Machines.

Naive Bayes (NB). The simplest approach is to assign a sentiment label l to a document d that maximizes the conditional probability: $l^* = argmax_l P(l|d)$. Assuming *conditional independence* between the features: $P(l|d) = \frac{P(l)\prod_{i=1}^{m} P(f_i|l)^{n_i(d)}}{P(d)}$, where $n_i(d)$ denotes the number of times feature f_i appears in document d, and $P(f_i|l)$ is computed based on the number of times f_i is observed in documents with the sentiment label l. Despite the conditional independence assumptions, this approach works fairly well with a good *smoothing* technique (e.g., Laplace smoothing).

Maximum Entropy (MaxEnt). Unlike Naive Bayes, Maximum Entropy does not make any assumptions regarding the relationships between the features, and therefore may perform better when conditional independence assumptions are not met. Its estimate of $P(l|d)$ has the following exponential form: $P(l|d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,l} \dot{f}_{i,l}(d,l))$, where $Z(d)$ is a normalization function. $F_{i,l}$ is a feature/class function for feature f_i and class l, defined as a binary indicator function that assumes the value 1 or 0 corresponding to whether the feature f_i is present for class l, or not. The weight λ_i for f_i learned from data depicts the importance of that feature for the classification task. The weights are learned to maximize the entropy of the distribution subject to the constraint that the expected values of the feature/class functions with respect to the model are equal to their expected values with respect to the training data. The underlying philosophy is to choose the model that makes the fewest assumptions about the data, while remaining consistent with it.

A more generalized model of this flavor that have been recently shown to outperform other classifiers is the *Conditional Random Field (CRF)*. It is typically used for sequence classification tasks, where the objective is to predict a *sequence* of labels instead of a single one — where, the model leverages interaction between the features of different label classes for more accurate predictions. It also does not suffer from the label bias problem due to global normalization, unlike the (locally normalized) Maximum Entropy Markov Model (MEMM) classifiers.

Support Vector Machines (SVM). Support Vector Machines (SVM) have been shown to be extremely effective for text classification tasks. SVM maps the features (using Kernels) to a high dimensional space, and constructs a hyperplane to separate the two categories. Although there can be an infinite number of such hyperplanes possible, SVM constructs the one with the largest functional margin given by the distance of the nearest point to the hyperplane on each side of it.

The solution can be written as $\overline{w} = \sum_j \alpha_j \cdot c_j \cdot \overline{d_j}$, where the α_j's are obtained by solving a dual optimization problem. The $\overline{d_j}$ corresponding to $\alpha_j > 0$ are called support vectors, since they are the only document vectors contributing to $\overline{w}$. New points are mapped to the same space and classified to a category based on which side of the hyperplane they lie.

Facet or Aspect Specific Sentiment Analysis. A review may have multiple facets or topics, with a different opinion about each facet. Consider the review in Example 1.1. The review is positive with respect to the topics "direction" and "story", but negative with respect to "acting". In order to analyze the sentiment of this review, it is necessary to identify the different facets or aspects in the review; and then analyze the sentiment about those facets. The overall sentiment of the review can then be computed as a weighted aggregation of the facet or aspect-specific sentiments.

Latent Aspect Rating Analysis Model (LARAM) [Wang 2010, Wang et al. 2011] jointly identifies latent aspects, aspect ratings, and weights placed on the aspects in a review. A shallow dependency parser is used to learn product aspects and aspect-specific opinions in [Yu et al. 2011] by jointly considering the aspect frequency and the consumers' opinions about each aspect. A rated aspect summary of short comments is done in [Lu 2009]. A topic model is used in [Titov 2008] to assign words to a set of induced topics. The model is extended through a set of maximum entropy classifiers, one per each rated aspect, that are used to predict aspect specific ratings.

A Joint Sentiment Topic model (JST) is described in [Lin 2009] which detects sentiment and topic simultaneously from text. In JST, each document has a sentiment label distribution. Topics are associated to sentiment labels, and words are associated to both topics and sentiment labels. In contrast to [Titov 2008] and some other similar works [Wang 2010, Wang et al. 2011, Yu et al. 2011, Snyder 2007, Lu 2009] which require some kind of a supervised setting like aspect ratings or overall review rating [Mukherjee 2013a], JST is fully unsupervised. The CFACTS model [Lakkaraju 2011] extends the JST model to capture facet coherences in a review using Hidden Markov Model. All of these generative models have their root in Latent Dirichlet Allocation Model [Blei 2003]. LDA assumes a document to have a probability distribution over a mixture of topics, and topics to have a probability distribution over words.

Author-Specific Sentiment Analysis. However, the models described so far do not consider any authorship information to incorporate author preferences for the facets, or author style information for maintaining coherence in reviews. The overall sentiment for review Example 1.1 will differ for different authors depending on their topic preferences; if a reviewer watches a movie for a good story and narration, then his (overall) sentiment for the movie will be different than that if he watches it only for the acting skills of the protagonists in this example.

An approach to capture author-specific topic preferences is described in [Mukherjee 2013a]. The work considers *predefined* seed facets for restaurants like "food, ambience, service" etc. and uses dependency parsing with a sentiment lexicon to find the sentiment about each facet. A WordNet similarity metric is used to assign each facet to a seed facet. Thereafter, they use linear regression to learn author preference for the seed facets from review ratings.

Joint Author Sentiment Topic Model (JAST) [Mukherjee 2014a] describes a generative process of writing a review by an author, without any supervision, building an author-layer over the Joint Sentiment Topic Model [Lin 2009]. Authors have different topic preferences, 'emotional' attachment to topics, writing style based on the distribution of semantic (topic) and syntactic (background) words and their tendency to switch topics. JAST uses Latent Dirichlet Allocation to learn the distribution of author-specific topic preferences and emotional attachment to topics. It uses a Hidden Markov Model to capture short range syntactic, and long range semantic dependencies in reviews to capture coherence in author writing style. JAST jointly discovers the topics in a review, author preferences for the topics, topic sentiment as well as the overall review sentiment from the point of view of the author of the review.

Temporal Evolution. Opinions are *dynamic* in nature, evolving with user experience, maturity, and social interactions. For instance, a review on a given item (e.g., movie) can have different polarities at different points in time — depending on when, or, at what level of maturity the user authored it in his lifecycle in the community [Mukherjee 2015, Mukherjee 2016]. For example consider the following camera reviews written by the same user at different points in time and maturity.

Example 1.7. "(a) My first DSLR. Excellent camera, takes great pictures with high definition, without a doubt it makes honor to its name. [Aug, 1997]

(b) The EF 75-300 mm lens is only good to be used outside. The 2.2X HD lens can only be used for specific items; filters are useless if ISO, AP,... . The short 18-55mm lens is cheap and should have a hood to keep light off lens. [Oct, 2012]"

Clearly, the user is much more experienced in appreciating fine-grained facets of the camera in the second review. The authors propose approaches based on Hidden Markov Model and Geometric Brownian Motion to capture the evolution of users' expertise based on their style of writing — where users at higher levels of expertise exhibit distinctive facet preferences and use sophisticated vocabulary compared to amateur users. Therefore, such approaches can be used for personalized sentiment analysis and item recommendation tailored to individual user's tastes, maturity, and expertise.

The dynamics on how users update their opinions on topics over time is affected by their social interactions as well. These can be used for forecasting users' sentiment, emotion, and attitudes over time [De 2015].

KEY APPLICATIONS

Word of mouth (WOM) is the process of conveying information from person to person and plays a major role in customer buying decisions. In commercial situations, WOM involves consumers sharing attitudes, opinions, or reactions about businesses, products, or services with other people. WOM communication functions based on social networking and trust. People rely on families, friends, and others in their social network. Research also indicates that people appear to trust seemingly disinterested opinions from people outside their immediate social network, such as online reviews. This is where Sentiment Analysis comes into play. Growing availability of opinion rich resources like online review sites, blogs, social networking sites have made this "decision-making process" easier for us. With explosion of Web 2.0 platforms consumers have a soapbox of unprecedented reach and power by which they can share opinions. Major companies have realized that these consumer voices shape voices of other consumers. Sentiment Analysis thus finds its use in:

- Consumer Market for product reviews for knowing consumer attitudes and trends for Marketing-related activities
- Social Media for finding general opinion about recent hot topics in town
- Product domains (like, movies and electronics) to find whether a recently launched item is gaining popularity

[Pang 2008] broadly classifies the applications into the following categories:

- Applications to review related websites like movie and product reviews
- Applications as a sub-component technology detecting antagonistic, heated language in mails, spam detection, context sensitive information detection etc.
- Applications in Business and Government Intelligence to know consumer attitudes and trends
- Applications across different domains to know public opinions for political leaders, or their notions about rules and regulations in place etc.

FUTURE DIRECTIONS

Sentiment Analysis has largely been restricted to textual analysis. An interesting extension can be to perform a *multi-modal* analysis by considering speech, images, and video to capture human behavior, interactions and sentiment in real-time which can have applications in the gaming industry, and improving security and intelligence.

With the recent advance of deep learning and representation learning in Natural Language Processing tasks, efforts have been underway to understand the compositionality of forming sentiment expressions in text at fine-grained granularity, and develop end-to-end systems for more robust sentiment predictions using *Neural Networks* [Socher 2013]. This also provides an easy framework to incorporate multi-modal data as separate embeddings in the model.

REFERENCES

[Agarwal 2005] Alekh Agarwal. *Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified.* In In Proceedings of the International Conference on Natural Language Processing (ICON, 2005.

[Benamara 2007] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato and V.S. Subrahmanian. *Sentiment analysis: Adjectives and adverbs are better than adjectives alone.* In Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 2007.

[Blei 2003] David M. Blei, Andrew Y. Ng and Michael I. Jordan. *Latent dirichlet allocation.* J. Mach. Learn. Res., vol. 3, pages 993–1022, March 2003.

[Chen 2010] Mosha Chen and Tianfang Yao. *Combining dependency parsing with shallow semantic analysis for Chinese opinion-element relation identification.* In 4th International Universal Communication Symposium, IUCS 2010, Beijing, China, October 18-19, 2010, pages 299–305, 2010.

[Dave 2003] Kushal Dave, Steve Lawrence and David M. Pennock. *Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews.* In Proceedings of the 12th International Conference on World Wide Web, WWW '03, pages 519–528, New York, NY, USA, 2003. ACM.

[De 2015] Abir De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya and Manuel Gomez-Rodriguez. *Modeling Opinion Dynamics in Diffusion Networks.* CoRR, vol. abs/1506.05474, 2015.

[Dong 2014] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun and Wei Zhang. *Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion.* In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, pages 601–610, New York, NY, USA, 2014. ACM.

[Esuli 2006] Andrea Esuli and Fabrizio Sebastiani. *SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining.* In In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06, pages 417–422, 2006.

[Hu 2004] Minqing Hu and Bing Liu. *Mining and Summarizing Customer Reviews.* In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04, pages 168–177, New York, NY, USA, 2004. ACM.

[Lakkaraju 2011] Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharya and Srujana Merugu. *Exploiting coherence in reviews for discovering latent facets and associated sentiments.* In SDM '11, pages 28–30, 2011.

[Lin 2009] Chenghua Lin and Yulan He. *Joint sentiment/topic model for sentiment analysis.* CIKM '09, pages 375–384, 2009.

[Liu 2004] H. Liu and P. Singh. *ConceptNet: A Practical Commonsense Reasoning Tool-Kit.* BT Technology Journal, vol. 22, no. 4, pages 211–226, October 2004.

[Lu 2009] Yue Lu, ChengXiang Zhai and Neel Sundaresan. *Rated aspect summarization of short comments.* In WWW, pages 131–140, 2009.

[Mukherjee 2012a] Subhabrata Mukherjee and Pushpak Bhattacharyya. *Feature Specific Sentiment Analysis for Product Reviews.* volume 7181 of *Lecture Notes in Computer Science*, pages 475–487. Springer, 2012.

[Mukherjee 2012b] Subhabrata Mukherjee and Pushpak Bhattacharyya. *Sentiment Analysis in Twitter with Lightweight Discourse Analysis.* In COLING, pages 1847–1864, 2012.

[Mukherjee 2012c] Subhabrata Mukherjee and Pushpak Bhattacharyya. *WikiSent: weakly supervised sentiment analysis through extractive summarization with wikipedia.* ECML PKDD'12, pages 774–793, 2012.

[Mukherjee 2013a] Subhabrata Mukherjee, Gaurab Basu and Sachindra Joshi. *Incorporating author preference in sentiment rating prediction of reviews.* In WWW, 2013.

[Mukherjee 2013b] Subhabrata Mukherjee and Sachindra Joshi. *Sentiment Aggregation using ConceptNet Ontology.* In Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013, pages 570–578, 2013.

[Mukherjee 2014a] Subhabrata Mukherjee, Gaurab Basu and Sachindra Joshi. *Joint Author Sentiment Topic Model.* In Proceedings of the 2014 SIAM International Conference on Data Mining (SDM'14), 2014.

[Mukherjee 2014b] Subhabrata Mukherjee and Sachindra Joshi. *Author-Specific Sentiment Aggregation for Polarity Prediction of Reviews.* In Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014., pages 3092–3099, 2014.

[Mukherjee 2015] Subhabrata Mukherjee, Hemank Lamba and Gerhard Weikum. *Experience-Aware Item Recommendation in Evolving Review Communities.* In 2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015, pages 925–930, 2015.

[Mukherjee 2016] Subhabrata Mukherjee, Stephan Günnemann and Gerhard Weikum. *Continuous Experience-aware Language Model.* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, pages 1075–1084, 2016.

[Pang 2002] Bo Pang and Vaithyanathan Shivakumar Lee Lillian. *Thumbs up?: sentiment classification using machine learning techniques.* EMNLP '02, 2002.

[Pang 2004a] Bo Pang and Lillian Lee. *A Sentimental Education: Sentiment Analysis Using Subjectivity.* In Proceedings of ACL, pages 271–278, 2004.

[Pang 2004b] Bo Pang and Lillian Lee. *A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts.* ACL '04, 2004.

[Pang 2008] Bo Pang and Lillian Lee. *Opinion Mining and Sentiment Analysis.* Found. Trends Inf. Retr., vol. 2, no. 1-2, pages 1–135, January 2008.

[Potthast 2010] Martin Potthast and Steffen Becker. Opinion summarization of web comments, pages 668–669. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[Ramteke 2013] Ankit Ramteke, Akshat Malu, Pushpak Bhattacharyya and J. Saketha Nath. *Detecting Turnarounds in Sentiment Analysis: Thwarting.* In ACL (2), pages 860–865. The Association for Computer Linguistics, 2013.

[Snyder 2007] Benjamin Snyder and Regina Barzilay. *Multiple Aspect Ranking Using the Good Grief Algorithm.* In HLT 2007, pages 300–307. ACL, April 2007.

[Socher 2013] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts. *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.* In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1631–1642, Stroudsburg, PA, October 2013. Association for Computational Linguistics.

[Stone 1966] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith and Daniel M. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. MIT Press, 1966.

[Taboada 2004] Maite Taboada and Jack Grieve. *Analyzing appraisal automatically.* In In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, pages 158–161, 2004.

[Titov 2008] Ivan Titov and Ryan T. McDonald. *A Joint Model of Text and Aspect Ratings for Sentiment Summarization.* pages 308–316. ACL, 2008.

[Turney 2002] Peter 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, ACL '02, pages 417–424, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[Wang et al. 2011] Hongning Wang et al. *Latent aspect rating analysis without aspect keyword supervision.* KDD '11, 2011.

[Wang 2010] Hongning Wang, Yue Lu and Chengxiang Zhai. *Latent aspect rating analysis on review text data: a rating regression approach.* In KDD, pages 783–792, 2010.

[Wei 2010] Wei Wei and Jon Atle Gulla. *Sentiment Learning on Product Reviews via Sentiment Ontology Tree.* In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10, pages 404–413, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[Wiebe 2004] Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell and Melanie Martin. *Learning Subjective Language.* Comput. Linguist., vol. 30, no. 3, pages 277–308, September 2004.

[Wolf 2004] Florian Wolf and Edward Gibson. *Representing discourse coherence: A corpus-based analysis.* In COLING 2004, 20th International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2004, Geneva, Switzerland, 2004.

[Wu 2009] Yuanbin Wu, Qi Zhang, Xuanjing Huang and Lide Wu. *Phrase Dependency Parsing for Opinion Mining.* In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3, EMNLP '09, pages 1533–1541, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[Yu et al. 2011] Jianxing Yu et al. *Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews.* pages 1496–1505. ACL, 2011.

[Yu 2003] Hong Yu and Vasileios Hatzivassiloglou. *Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences.* In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP '03, pages 129–136, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.