

Feature Based Sentiment Analysis On Customer Feedback: A Survey

Avani Jadeja

Department of computer Engineering
Hasmukh Goswami College of Engineering
Ahmadabad – 382330
India

Prof. Indr jeet Rajput

Department of computer Engineering
Hasmukh Goswami College Of Engineering
Ahmadabad - 382330
India

ABSTRACT: *With the sheer volume of customer feedback on E –commerce site such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis. The number of the customer reviews that a product receives grows rapidly. After using product, Consumers usually express their experience of feedback via e-commerce sites. People are intended to develop a system that can identify and classify opinion or sentiment as represented in a customer feedback. An accurate method for prediction sentiments could enable us to extract opinions from the internet and predict online customer’s preferences, which could prove valuable for economic or marketing research. There are different problems in this research like sentiment classification, feature based classification and handling negations. This survey paper covering the techniques and methods in feature based sentiment analysis and challenges appear in the field.*

Key words: sentiment, opinion, semantic, machine learning, Sentiment classification.

1. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

There are several challenges in Sentiment analysis. The first is a opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very

different from "the picture was not great". People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs, the more likely people are to combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as its last movie" is entirely dependent on what the person expressing the opinion thought of the previous model.

The users hunger is on for and dependence upon online advice and recommendations the data reveals is merely one reason behind the emerge of interest in new systems that deal directly with opinions as a first-class object. Sentiment analysis concentrates on attitudes, whereas traditional text mining focuses on the analysis of facts. There are few main fields of research predominate in Sentiment analysis:

sentiment classification, feature based Sentiment classification and opinion summarization. Sentiment classification deals with classifying entire documents according to the opinions towards certain objects. Feature-based Sentiment classification on the other hand considers the opinions on features of certain objects. Opinion summarization task is different from traditional text summarization because only the features of the product are mined on which the customers have expressed their opinions. Opinion summarization does not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization.

For the sake of convenience the remainder of this paper is organized as follows: Section 2 presents the data sources used for customer feedbacks. Section 3 introduces different approaches for feature based sentiment classification. Section 4 presents some applications of sentiment classification. Last section concludes our study and discusses some future directions for research.

2. DATA SOURCE

User's opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, review sites, data and micro blogs provide a good understanding of the reception level of the products and services.

2.1. Blogs

With an increasing usage of the internet, blogging and blog pages are growing rapidly. Blog pages have become the most popular means to express one's personal opinions. Bloggers record the daily events in their lives and express their opinions, feelings, and emotions in a blog (Chau & Xu, 2007). Many of these blogs contain reviews on many products, issues, etc. Blogs are used as a source of opinion in many of the studies related to sentiment analysis (Martin, 2005; Murphy, 2006; Tang et al., 2009).

2.2. Review sites

For any user in making a purchasing decision, the opinions of others can be an important factor. A large and growing body of user-generated reviews is available on the Internet. The reviews for products or services are usually based on opinions expressed in much unstructured format. The reviewer's data used in most of the sentiment classification studies are collected from the e-commerce websites like

www.amazon.com(product reviews), www.yelp.com (restaurant reviews), www.CNET download.com (product reviews) and www.reviewcentre.com, which hosts millions of product reviews by consumers. Other than these the available are professional review sites such as www.dpreview.com, www.zdnet.com and consumer opinion sites on broad topics and products such as www.consumerreview.com, www.epinions.com, www.bizrate.com (Popescu & Etzioni, 2005; Hu, B. Liu, 2006; Qinliang Mia, 2009; Gamgaran Somprasertsi, 2010).

2.3. Dataset

Most of the work in the field uses movie reviews data for classification. Movie review data's are available as dataset <http://www.cs.cornell.edu/People/pabo/movie-review-dat> a). Other dataset which is available online is multi-domain sentiment (MDS) dataset. (<http://www.cs.jhu.edu/mdredze/datasets/sentiment>). The MDS dataset contains four different types of product reviews extracted from Amazon.com including Books, DVDs, Electronics and Kitchen appliances, with 1000 positive and 1000 negative reviews for each domain. Another review dataset available is <http://www.cs.uic.edu/liub/FBS/CustomerReviewData.zip>. This dataset consists of reviews of five electronics products downloaded from Amazon and Cnet (Hu and Liu, 2006; Konig & Brill, 2006; Long Sheng, 2011; Zhu Jian, 2010; Pang and Lee, 2004; Bai et al., 2005; Kennedy and Inkpen, 2006; Zhou and Chaovalit, 2008; Yulan He 2010; Rudy Prabowo, 2009; Rui Xia, 2011).

2.4. Micro-blogging

Twitter is a popular micro blogging service where users create status messages called "tweets". These tweets sometimes express opinions about different topics. Twitter messages are also used as data source for classifying sentiment.

3. SENTIMENT CLASSIFICATION

Much research exists on sentiment analysis of user opinion data, which mainly judges the polarities of user reviews. In these studies, sentiment analysis is often conducted at one of the three levels: the document level, sentence level, or attribute level. In addition to that, the nature language processing techniques (NLP) is used in this area, especially in the document sentiment detection. Current-day sentiment detection is thus a discipline at the crossroads of NLP and Information retrieval, and as such it shares a number of characteristics with other tasks such as information extraction and text-mining,

computational linguistics, psychology and predicative analysis. In sentiment analysis, methods are machine Learning, .Semantic Orientation; Role of negation, Feature based sentiment classification. In relation to sentiment analysis, the literature survey done indicates one technique feature based sentiment analysis with different research papers.

3.1 Mining and Summarizing Customer Reviews [1]

In this paper, authors presented novel idea for extracting sentiment analysis and summary generation of customer reviews. They had crawled customer reviews and created review database. This paper first extracts features and then identifies opinions expressed for it. To extract features, they applied POS tagging [13] to every sentences of review. POS tagging identifies noun, noun group, noun phrase, verb, adverb, adjectives and other sentence parts. For POS tagging, they used NLProcess linguistic parser [7]. After POS, they applied deletion of stop words, stemming and fuzzy matching on result. Stop words are articles (a, an, the), supporting verbs (is, are, was, were, am) and other words which provides grammatical support. Stemming means converting plural into singular, past participle word into normal word and etc. Fuzzy matching [8] converts word variants and misspellings into correct word. They stored POS tagged sentences in review database and created transaction file for identifying frequent features.

This paper identifies product features on which many people have expressed their opinions. These features called frequent features and this step is called frequent feature generation. Hu and Liu use association rule mining [9] to find all frequent item sets, which is a set of words or phrases that occurs together. The association rule miner CBA [10] based on Apriori algorithm [9], finds all frequent item sets in the transaction file which appears in more than 1% of the review sentences.

In next step, authors applied Feature Pruning which aims to remove those incorrect features. Two types of pruning are presented: (a) Compactness pruning [8] checks features that contain at least two words, which are named feature phrases, and removes those that are likely to be meaningless because their words do not appear together in a specific order. (b) Redundancy pruning [8] removes redundant features that contain single words. For instance, *life* by itself is not a useful feature while *battery life* is a meaningful feature phrase. So Redundancy pruning removes word like *life*.

Paper performs opinion words extraction with all the remaining frequent features after pruning. Depending on previous work on subjectivity by Bruce and Wiebe (2000)[11], this paper uses adjectives as opinion words. Sentences containing one or more frequent features and one or more opinion words (i.e. adjectives) than they are called Opinion sentences. Paper extracts opinion words with following manner:

For each sentence in the review database
 If (it contains a frequent feature, extract all the adjective words as opinion words)
 For each feature in the sentence .The nearby adjective is recorded as its effective opinion
 /* A nearby adjective refers to the adjacent adjective that modifies the noun / noun phrase that is a frequent feature. */

Next step is infrequent feature identification. Hu and Liu suppose people like to use the same opinion word to describe different features. So they can use the opinion words to look for features that cannot be found in frequent feature generation step. If one sentence contains no frequent feature but one or more opinion words, find the nearest noun or noun phrase of the opinion word as an infrequent feature.

After identifying frequent, infrequent features and their adjacent opinion words in previous steps, paper next identifies orientation of opinion words and eventually determines orientation of opinion sentence. Authors had created database of opinion words and their orientation as follows:

Authors took 30 opinion words (i.e. adjectives) and determined their orientation manually. These adjectives were used as seed to create database. For every adjective, they identify synonyms and antonyms from WordNet [12]. To every synonyms of adjective, they assigned same orientation as adjective has and to every antonyms of adjective, they assigned opposite orientation. E.g. Nice is adjective with positive orientation, so its synonyms like beautiful, wonderful etc. will have positive orientation, while its antonyms like ugly, disgraceful etc. will have negative orientation. They had done above step recursively for every synonym and antonym of adjective. This way, database had been created with almost every adjective and their orientation.

After creating adjective database, they determined orientation of opinion sentence as follows. For every opinion word of opinion sentence, they got its orientation value from above database. Summation of orientation of opinion words had been counted for

every opinion sentence. If summation is positive than sentence orientation is positive otherwise negative. For every discovered feature, related opinion sentences were put into positive and negative categories according to previous step. A count was computed to show how many reviews positive / negative opinions to the feature. Features were ranked according to their frequency of their appearances in the reviews i.e. most frequent feature would be on top of summary.

The proposal in this paper [1] can produce number of features, but only explicit features could be found and implicit feature does not be extracted. The irrelevant sentences may be thought as opinion sentences, and the nouns in irrelevant sentence would be extracted as features. In addition, the assumption that people use the same opinion word to describe different features, to find infrequent feature is not so reasonable. Different features tend to be described by different opinion words.

3.2 Mining Product Features from Online Reviews [2]

This paper is closely related to Hu and Liu's research paper [1] on mining opinion features in customer reviews. This paper follows same POS tagging [13] process as Hu and Liu's with only change in linguistic parser. It uses *OpenNLP* [14] instead of *NLPProcess* [7].

To identify opinion sentence, this paper propose use of Sentiwordnet [16]. Adjectives, adverbs and verbs are taken as target words to be identified as opinion words. Sentiwordnet stores synsets instead of term, because it assumes that every term can have different opinion-related properties in different senses. Every synsets in Sentiwordnet have 3 scores: positivity, negativity and objectivity. They can range from 0.0 to 1.0 and their sum is 1.0 for each synsets. This paper calculates positive and negative score for each adjective, adverb and verb from Sentiwordnet and calculates average positive or negative score of sentence. If it is greater than certain score, than it marks current sentence as opinion sentence. This paper uses probability based algorithm to generate features as follows:

Probability of any candidate feature being correct feature = Number of occurrence of that candidate feature / Number of sentences if appeared in.

For explicit feature extraction, every noun of sentence is a candidate feature. Hu and Liu use

association rule mining to find all frequent items sets that occur together in one sentence

However its computation is too large and result needs compact pruning. For implicit features, this paper maps adjective used in sentence of implicit feature to one noun i.e. "Heavy" has been mapped to "Weight". Using Word Net, paper gets synonyms and antonyms of every adjective and maps them with noun. This paper uses redundancy pruning to remove redundant features same way Hu and Liu have used. This paper performs its proposal on same data set which Hu and Liu have used and gets better performance in precision and recall.

3.3 Extracting Product Features and Opinions from Reviews [3]

This paper is about one unsupervised information extraction system called "OPINE". Given a particular product and a corresponding set of reviews, OPINE solves the following opinion mining tasks Identify product features, Identify opinions regarding product features, Determine the polarity of opinions, Rank opinions based on their strength. And outputs a set of *product features*, each accompanied by a list of *associated opinions* which are ranked based on strength (e.g. "abominable" is stronger than "bad"). OPINE is built on top of KnowItAll [17], a Web-based, domain-independent information extraction system (Etzioni et al., 2005).

Explicit feature extraction accomplish as follows. OPINE first extracts the noun phrases from reviews and retains those with frequency greater than an experimentally set threshold. Its feature access or, which is instance of KnowItAll system, evaluates each noun phrase by computing Point-Wise-Mutual-Information (PMI) [18] scores, which is estimated from web search hit counts, between the phrase and discriminators associated with the product. It also distinguishes parts from properties of product using WorldNet's IS-A hierarchy. Compared to previous work, OPINE achieves 22% higher precision (with only 3% low recall) on the feature task. Its use of Relaxation-Labeling Technique to determine the semantic orientation of potential opinion words in the context of the extracted product features and specific review sentences results high precision and recall.

3.4 Mining Sentiment from Tweets[4]

In this paper, they identify tweet sentiment using pre-annotated tweets corpus. They use two different datasets which are built using emotions and list of suggestive words respectively as noisy labels. They give a new method of scoring "Popularity Score", which allows determination of the popularity score at the level of individual words of the tweet text.

Calculate positive and negative scores of tweets by applying following methods to calculate score.

Next step is potential opinion phrases extraction. OPINE uses MINIPAR parser to compute syntactic dependencies between explicit features. If an explicit feature is found in a sentence, OPINE applies the extraction rules in order to find the head of potential opinion phrases. Each head word together with its modifier is returned as potential opinion phrase. OPINE examines the potential opinion phrases in order to identify the actual opinion. First, system finds the semantic orientation for the lexical head to each potential opinion phrase. Every phrase whose head word has a positive or negative semantic orientation is opinion phrase.

Relaxation labeling for semantic orientation is unsupervised classification technique which takes set of objects, set of semantic orientation (SO) labels i.e. {positive, negative, neutral}, initial probabilities for each object, possible labels, a set of other objects which influence object's SO label called neighborhood, neighborhood features and support function as input and assigns semantic orientation label to object. It is an iterative procedure whose output is an assignment of semantic orientation labels to objects i.e. features. At each iteration, algorithm used update equation to re estimate the probability of an object label based on its previous probability estimate and the features of its neighborhood. The algorithm stops when global label assignment stays constant over multiple consecutive iterations.

In Baseline approach, they first clean the tweet by removing all special characters, targets (@), hash tags (#), URLs, emotions, etc. Then create unigram (token) with positive and negative probabilities. they sum up the positive and negative probability scores of all the constituent unigrams and use difference (positive-negative) to find the overall score of the tweet. If tweet score is >0 then it is positive tweet otherwise negative.

In Emoticons and Punctuations handling approach, they uses agrwal et al.2011 list of emotions. This list is built from Wikipedia's emoticons and tagged with five classes (extremely positive, positive, neutral, negative, and extremely negative). This paper has merged extremely positive into positive and neutral, extremely negative into negative. This paper gives +1 score to positive and -1 to negative emoticon. Exclamation (!) mark gets 0.1 positive score and Question mark (?) gets 0.1 negative score.

In Stemming approach, they use Porter Stemmer to stem the tweet words, or to remove plurals, -Ed and -ing from words.

In Stop Word Removal approach, they remove the stop word like he, she, at, on, a, the etc. Because stop word don't carry any sentiment information and removes words of ≤ 2 length.

In Spell Correction approach, they use the spell correction algorithm from (Bora 2012).they replace a word with any character repeating more than twice with two words, one in which the repeated character is placed once and second in which the repeated character is placed twice. In this paper doesn't apply phonetic level spell correction for words like "thr" which has been used in place of "there".

In Senti Features approach, authors have used Twitrratr.com. Twitrratr.com provides list of most commonly used positive and negative words. This paper uses these lists to check each unigram (token word) of twit. Paper than score the token + / - 1 depending on the list in which it exist.

In Noun Identification approach, noun words don't carry sentiment and thus this paper ignores them. It uses part of speech tag in English Word Net to identify word as noun.

In Popularity Score, This scoring method boots the score of the words are most commonly used as positive or negative. It calculates popularity factor (pf) on basic of occurrence frequency.

In this paper, they use two datasets; The Stanford dataset was built automatically using emoticons as noisy labels. It has around 1.6 million tweets, equal number of positive and negative. And The Mejaj Dataset contains around 1.4million tweets plus contains set of 40 words which were manually categorized into the positive and negative. After applying above score calculation methods achieves 87% accuracy on Stanford dataset and 88% accuracy on Mejaj Dataset.

3.5 Domain Independent Model for Product Attribute Extraction from User Reviews using Wikipedia [5]

This paper use Wikipedia to identify words from customer reviews as product attributes instead of applying any natural language processing. They propose a novel supervised domain independent

model for product attribute extraction from user reviews. For any given product, they use one approach to attribute extraction. Authors have collected customer reviews of give product from E-Commerce sites, Wikipedia and web. They have than removed stop words. Compute features that we have defined, for the remaining words. Identification of possible attribute words using classification model trained on these features.

Most Frequent Items – MFI feature boots the importance of attribute words by their frequency occurrence in customer reviews. The set of words $\{z_1, z_2, z_3, \dots, z_m\}$ used for this feature are obtained from customer reviews of a given product, after stop word removal is done. MFI score for word is calculated as its frequency/total of every words frequency.

Context Relation using Wikipedia – CR, This paper checks each and every word from previous step to find out article in Wikipedia. If there is article present then it marks word as Wikipedia word otherwise it marked as nonwikipedia word.

Role of surrounding window – SW, This feature helps in identifying sub-attributes (attributes of attributes) by considering left and right words of Wikipedia for given word.

Web search engine reference – WR, The WR score measures the association of a particular word from customer reviews of a product with that product on the internet. This paper have used Bing search engine API and calculate WR as number of instances where the word and the product name both occur within the text snippets given as search results divide by total search results.

In this paper they have used two datasets, Review9-products datasets and Customer Reviews dataset and calculate precision and recall. They presented a domain independent approach for automatic discovery of product attributes from user reviews.

3.6 Sentiment Classification Using Sentence-level Semantic Orientation of Opinion Terms from Blogs [6]

In this paper, sentences are split into subjective and objective ones based on lexical dictionary. Subjective sentences are classified as positive, negative or neutral opinion. A rule based lexicon method is used for the classification of subjective and objective sentences. From Subjective sentences, the opinion expressions are extracted and their semantic scores are checked using the SentiWordNet [16] directory.

The final weight of each individual sentence is calculated after considering the whole sentence structure, contextual information and word sense disambiguation.

Sentiment analysis overall process work as, first split reviews into sentences and make a bag of sentences (BOS). Remove noise form sentences using spelling correction, convert special characters and symbols (phonetics) to their text expression. Use POS [13] for tagging each word of the sentence and store the position of each word in the sentence. Second, make a comprehensive dictionary (feature vector) of the important feature with its position in the sentence. Third classify the sentences into objective and subjective sentences using lexical approach. Fourth, using a lexical dictionary as knowledge base, check the polarity of the subjective sentence as positive, negative or neutral. Fifth, check and update polarity using the sentence structure and contextual feature of each term in the sentences.

4. Applications

When faced with tremendous amounts of online information from various online forums, information Seekers usually find it very difficult to yield accurate information that is useful to them. This has motivated the research on identification of online forum hotspots, where useful information is quickly exposed to those seekers. Nan Li (2010) used sentiment analysis approach to provide a comprehensive and timely description of the interacting structural natural groupings of various forums, which will dynamically enable efficient detection of hotspot forums. In order to identify potential risks, it is important for companies to collect and analyze information about their competitors' products and plans. Sentiment analysis find a major role in competitive intelligence (Kaiquan Xu , 2011) to extract and visualize comparative relations between products from customer reviews, with the interdependencies among relations taken into consideration, to help enterprises discover potential risks and further design new products and marketing strategies.

Opinion summarization summarizes opinions of articles by telling sentiment polarities, degree and the correlated events. With opinion summarization, a customer can easily see how the existing customers feel about a product, and the product manufacturer can get the reason why different stands people like it or what they complain about. Ku, Liang, and Chen (2006) investigated both news and web blog articles. Algorithms for opinion extraction at word, sentence and document level are proposed. The issue of

relevant sentence selection is discussed, and then topical and opinionated information are summarized. Opinion summarizations are visualized by representative sentences. Finally, an opinionated curve showing supportive and non-supportive degree along the timeline is illustrated by an opinion tracking system. Other applications includes online message sentiment filtering-mail sentiment classification, web blog author's attitude analysis etc.

5. Conclusion

Sentiment detection has a wide variety of applications in information systems, including classifying reviews, Summarizing review and other real time applications. There are likely to be many other applications that is not discussed. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that different methods are used to find out feature form customer reviews, different types of Features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

In future, more work is needed on further improving the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for sentiment analysis are advancing fast, however, a lot of problems in this field of study remain unsolved. The main challenging aspects exist in use of other languages, dealing with negation expressions; produce a summary of opinions based on product features/attributes, complexity of sentence/document, handling of implicit product features, etc. More future research could be dedicated to these challenges.

References

- [1] Hu, Minqing and Bing Liu. 2004. "Mining and summarizing customer reviews." SIGKDD '04, pages 168-177, NY, USA. ACM.
- [2] Weishu Hu, Zhiguo Gong and Jingzhi Guo, "Mining Product Features from Online Reviews." IEEE 2010.
- [3] Ana-Maria Popescu and Oren Etzioni, "Extracting Product Features and Opinion from Reviews", Proceeding of Human Language Technology Conference and Conference on Empirical Methods in Natural Language, ACL, Vancouver, October 2005, pp 339-336.
- [4] Akash Bakliwal, Vasudeva Varma, "Mining Sentiment from Tweets", Association for Computational Linguistics (ACL) 2012.
- [5] Sudheer Kovelamudi, Sethu Ramalingam, Arpit Sood and Vasudeva Verma, "Domain Independent Model for Product Attribute Extraction from User Reviews using Wikipedia", 5th International Joint Conference on Natural Language Processing, page 1408-1412, Thailand, Nov-2011.
- [6] Aurangzeb Khan, Baharum Baharudin, "Sentiment Classification Using Sentence-level Semantic Orientation of Opinion Terms from Blogs", IEEE, 2011.
- [7] NLProcessor - Text Analysis Toolkit. 2000. <http://www.infogistics.com/textanalysis.html>
- [8] Hu, M., and Liu, B. 2004. "Mining Opinion Features in Customer Reviews." To appear in AAAI'04, 2004.
- [9] Aggrawal, R. & Srikant, R. 1994. "Fast algorithm for mining association rules." VLDB'94, 1994.
- [10] Liu, B., Hsu, W., Ma, Y. 1998. "Integrating Classification and Association Rule Mining." KDD'98, 1998.
- [11] Bruce, R., and Wiebe, J. 2000. "Recognizing Subjectivity : A Case Study of Manual Tagging." Natural Language Engineering
- [12] Miller, G., Beckwith, R, Fellbaum, C., Gross, D., and Miller, K. 1990. "Introduction to WordNet: An on-line lexical database." International Journal of Lexicography (special issue), 3(4) : 235-312.
- [13] Manning C. and Schütze H., "Foundations of Statistical Natural Language Processing." MIT Press, May 1999.
- [14] OpenNLP, open source toolkit for natural language processing <http://opennlp.sourceforge.net/>
- [15] Jokinen P., and Ukkonen E., "Two algorithms for approximate string matching in static texts", Mathematical Foundations of Computer Science, 1997.
- [16] Andrea Esuli and Fabrizio Sebastiani, "Sentiwordnet: A publicly Available Lexical Resource for Opinion Mining", In Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation, Genova, IT. 2006, pp 417-422.
- [17] Etzioni, M. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, an A. Yates. 2005. "Unsupervised named-entity extraction from the web: An experimental study." Artificial Intelligence, 165(1) : 91-134
- [18] D. Turney. 2001. "Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL." In Procs. Of the Twelfth European Conference on Machine Learning (ECML-2001), pages 491-502, Freiburg, Germany.

[19] R.A.Hummel and S.W.Zucker. 1983. "On the foundations of relaxation labeling processes." In PAMI, pages 267-287

IJERT