# Entity-Centric Multimodal Aspect-Opinion Mining in Social Media

[1]Nasrin Shah, [2]S N Gite

[1]Lecturer, [2]Assistant Professor
MSSCET Jalna

***Abstract*—This article describes the approach we take to the analysis of social networks, combining the extraction of textual and multimedia opinions (images) and focusing on the recognition of entities and events. We examined a particular use case, which is to help archivists select material to be included in a social media file to preserve community memories, moving towards a structured preservation around semantic categories. The textual approach we consider is a rule-based approach wherein we consider a series of subcomponents, taking into account the problems that are there and inherent in the social media colony such as noisy grammar or misspelled words, oath or other form of speech including sarcasm, and so on. The analysis of multimedia content Complements this work to solve ambiguity and provide other contextual information. We propose two main innovations in this work: first, the new combination of tools to extract information from text and multimedia; And second, the adaptation of NLP tools for the exploration of specific information to the problems of social networks.**

***IndexTerms*— Application, knowledge mining, probabilistic topic model, LDA, Dirichlet**

## I. Introduction

Social Web analysis concerns all users who are actively engaged and generate content. This content is dynamic, reflecting the societal and sentimental fluctuations of the authors as well as the ever-changing use of language. Social networks are pools of a wide range of articulation methods, from simple "Like" buttons to complete articles, their content representing the diversity of public opinion. User activities on social networking sites are often triggered by specific events and related entities (eg sports events, celebrations, crises, news articles) and topics (eg global warming, the financial crisis , Swine flu). With the volume of rapidly growing resources on the Web, archiving this hardware becomes an important challenge. The notion of community memory extends traditional web archives with related data from a variety of sources. To include this information, a semantic and social-based preservation model is a natural way: Web 2.0 exploitation and the wisdom of crowds can make web archiving a more selective and meaning-based process.

Social media analysis can help archivists choose material for inclusion, while social media can enrich the archive by promoting structured preservation around semantic categories. In this article, we focus on the challenges in developing opinion extraction tools from textual and multimedia content. We focus on two very different areas: socially conscious federated social archiving (realized by the national parliaments of Greece and Austria) and web archiving of the socially contextualized broadcaster (produced by two large multimedia broadcasting organizations based in Germany : Sudwestrundfunk and Deutsche Welle). The objective is to help journalists and archivists answer questions such as opinions on critical social events, their distribution, evolution, opinion leaders and their impact and influence. Parallel to natural language, a large number of interactions between participants in the social network include other media, especially images. Determining whether a specific non-textual multimedia element functions as an opinion-forming device in some interaction becomes an important challenge, even more so when the textual content of an interaction is weak or has no strong feelings. Trying to determine a feeling value for an image clearly presents great challenges, and this area of research is still in its infancy. We describe here a work we have undertaken, first of all to try to provide a value of feeling from an image outside of any specific context and, on the other hand, to use the multimodal nature of the social Facilitate the analysis of the feeling of multimedia or text.

## II. Related Work

While much work has recently focused on analyzing social media to get an idea of what people think about current topics of interest, there are still many challenges ahead. Current mining approaches of opinion, which focus on product reviews and so on, are not necessarily adapted to our task, partly because they tend to operate in a single narrow area and partly because, The objective of the opinion is either known in advance or at least to a limited subset (eg film titles, product names, companies, political parties, etc.).

In general, sensing techniques can be roughly divided into lexicon based methods [22] and machine learning methods, [1]. Methods based on the Lexicon are based on a lexicon of feeling, a collection of known and pre-compiled feeling terms. Machine learning approaches use syntactic and / or linguistic characteristics, and hybrid approaches are very common, with feel lexicons playing a key role in most methods. For example, [17] establishes the polarity of the examinations by identifying the polarity of the adjectives that appear in them, with a reported accuracy of about 10% greater than pure machine learning techniques. However, such relatively successful techniques often fail when they are moved to new domains or types of text, as they are inflexible with respect to the ambiguity of the terms of feeling. The context in which a term is used can change its meaning, especially for adjectives in the lexicons of feeling [18]. Several evaluations have shown the utility of contextual information [26] and have identified context words with a high impact on the polarity of ambiguous terms [8].

Another bottleneck is the long creation of these sentiment dictionaries, although solutions have been proposed in the form of crowd sourcing techniques.3 Almost all of the work on the use of Twitter opinion has used techniques of Machine learning. [19] was to classify arbitrary tweets on the basis of a positive, negative, and neutral feeling, by constructing a simple binary classifier that used n-gram and POS functions and formed on instances that had been annotated according to The existence of positive and negative emotions. Their approach has much in common with a prior sentiment classifier constructed by [9], which also used unigrams, bigrams and POS tags, although the first one demonstrated by analysis that the distribution of some POS tags varies between positive posts And negative. One reason for the relative lack of linguistic techniques for extracting opinions on social networks is probably due to difficulties in using NLP on poor quality text [7]; for example. The Stanford NER falls from 90.8% F1 to 45.88% when applied to a corpus of tweets [14].

There have been a number of recent work to detect sarcasm in tweets and other user-generated content [23, 13, 20, 5], with typically about 70-80% accuracy. They mainly train on a set of tweets with the #sarcasm and / or the #irony markers, but simply try to classify whether a sentence or tweet is sarcastic or not (and sometimes in a set of types Of predefined sarcasms). However, none of these approaches exceeds the initial classification phase and therefore can not predict how sarcasm will affect the feeling expressed.

Extracting the feeling from images is still an area of research that is in its infancy and is not yet published prolificly. However, published ones often use small sets of data for their truth on the ground to build SVM classifiers. Evaluations show that systems often respond a little better than chance to emotions formed from general images [27]. The implication is that the selection of features for such a classification is difficult. [25] used a set of color functions to classify their small set of ground truth data, also using SVM and publishing an accuracy of about 87%. In our work, we develop this color-based approach to use other features and also use the wisdom of the crowd to select a large set of ground truth data. Other articles have begun to indicate the multimodal nature of image sentiment on the Web. Previous work, such as [11], involves a similar multimodal image annotation, but not specifically for sentiment. They use latent semantic spaces for the correlation of image features and text in a single feature space. In this article we describe the work we have undertaken to use text and images to create a sense for social media.

## III. Opinion Mining from Text

The ambiguity is a particular problem for tweets because we can not easily use coreference information: unlike articles and blog comments, tweets usually do not follow a thread and appear much more in isolation from other tweets . They also have much more linguistic variation and frequently use emoticons, abbreviations and hashtags, which can be an important part of meaning. Typically, they also contain intensive use of irony and sarcasm, which is particularly difficult for a machine to detect. On the other hand, their sensitivity can also be beneficial in focusing subjects more explicitly: it is very rare that a single tweet is linked to more than one subject, which can thus help with disambiguation by focusing On the situational relationship. In longer publications such as blogs, commentaries on news articles and so on, another challenge is raised by tracking the changing and contradictory interpretations in the discussions. We are studying the first steps towards a coherent model for identifying opinion holders and targets in a thread (taking advantage of the information about the extracted entities involved). We refer the reader to [2] for our specific IE work on Twitter, which we use as a pre-processing for the notice extraction described below. It's not just the tweets that are problematic; The sarcasm and noisy language of other forms of social media also have an impact. In the next section, we demonstrate some ways in which we deal with this.

Opinion Mining Application Our approach is based on a rule, similar to the one used by [22], focusing on the construction of a number of subcomponents that all have an effect on the score and polarity of a sub- a feeling. On the other hand, our opinion extraction component finds opinions about entities and events previously identified in the text. The main component of opinion exploitation is described in [15], so we need only give an overview, and we focus on some social media issues that have not been addressed in this paper. Such as the detection of sarcasm and the decomposition of hashtag. Detection of real opinion is carried out in different phases: the detection of positive, negative and neutral words, the identification of factual or operative points against questions or doubtful statements, identification of negatives, sarcasm and l Irony, hashtag analysis and detection of extra-linguistic clues such as smilies. The application involves a set of grammars that create annotations on segments of text. Grammar rules use directory information combined with language features (POS tags), as well as contextual information to create a set of annotations and features that can be changed at any time by other rules. Once the words of feeling have been adapted, we find a linguistic relationship between them and an entity or event in the sentence or phrase. A Sentiment annotation is created for this entity or event, with characteristics indicating the polarity (positive or negative) and the polarity score. The scores are based on the initial score of the word sentiment, and are intensified or diminished by modifiers such as swear words, adverbs, negations, sarcasms, etc., as explained below. The sworn words are particularly prolific on Twitter, especially on topics such as popular culture, politics and religion, where people tend to have a very strong vision. To cope with this, we associate ourselves with a list of matching words and phrases, which was manually created from various lists found on the web and a manual inspection of the data, including words acquired by collecting Tweets with sworn words like hashtags (which often contain more sworn words in the main text of the tweet). Useful information about feeling is contained in hashtag tags, but this is problematic to identify because hashtags usually contain several words in a single token, for example #not really. If a hashtag is camelcased, we use the capitalization information to create separate tokens. Second, if the hashtag is in lowercase or uppercase, we try to form a combination of tokens against the Linux dictionary. Working from left to right, we search for the longest match against a known word, then continue from the next shift. If a combination of matches can be found without interruption, the individual components are converted into tokens. In our example, #notreally would be correctly identified as "not" + "really". However, some hashtags are ambiguous: for example,

"#greatstart" does not share well in both "greats" + "tarta" tokens. These problems are difficult to solve; In some cases, we could use contextual information to help you. We conducted an experiment to measure the accuracy of the hashtag decomposition, using a corpus of 1000 randomly chosen tweets among the US elections that we undertook in the project. 944 histograms were detected in this corpus, 408 of which were identified as multi-word chips (we included combinations of letters and numbers as multi-words but no abbreviations). 281 were camelled and / or combinations of letters and nubards, 27 were foreign words, and the other 100 had no distinguished characteristics. The evaluation of hard-to-recognize cases (multiple-word chips without camels) produced scores of 86.91% accuracy, 90% recall and 88.43% F measurement. Since these difficult-to-solve combinations make up about a quarter of the multi-word hashtags in our corpus, and we fully succeed in decomposing the remaining hashtags, this means that the overall accuracy for hashtag decomposition is much higher.

In addition to using the sentiment information of these hashtags, we also collect new hashtags that usually indicate sarcasm, as often more than one sarcastic hashtag is used. For this, we used the GET list collector to collect pairs of hashtags where one knew to be sarcastic and examined the second hashtag manually. From this we could identify another set of hashtags indicating sarcasm, like #thanksdude, #yay etc. Further research must be done on these to ascertain how often they actually indicate sarcasm when used alone. Finally, emoticons are treated like other words bearing the feeling, according to another list of nomenclature, if they occur in combination with an entity or an event. For example, the tweet "They all voted Tory :-(" would be annotated as negative to the target "Tory." Otherwise, as for swear words, if a sentence contains a smiley but no other entity or event, Annotated as sentiment-bearing, with the value of that of the smiley from the list of the nomenclator. Once all subcomponents have been passed to the text, a final output is produced for each feeling segment, with A polarity (positive or negative) And a score, based on the combination of the individual scores of the different components (eg, the negation component usually returns to the polarity, the adverbial component increases the strength of the feeling, etc.) Aggregation Of feeling happens for all mentions of the same entity / event in a document, so that summaries can be created.

## IV. PROPOSED SYSTEM

Opinions can be expressed about anything such as a product, a service, or a person by any person or organization. We use the term entity to denote the target object that has been evaluated. An entity can have a set of compo-nents (or parts) and a set of attributes. Each component may have its own sub-components and its set of attributes, and so on. Thus, an entity can be hierarchically decomposed based on the part-of relation (Liu, 2006).

**Definition (entity):** An entity e is a product, service, person, event, organ-ization, or topic. It is associated with a pair, e: (T, W), where T is a hierar-chy of components (or parts), sub-components, and so on, and W is a set of attributes of e. Each component or sub-component also has its own set of attributes.

**Example**: A particular brand of cellular phone is an entity, e.g., iPhone. It has a set of components, e.g., battery and screen, and also a set of attributes, e.g., voice quality, size, and weight . The battery component also has its own set of attributes, e.g., battery life, and battery size.

Based on this definition, an entity can be represented as a tree or hierar-chy. The root of the tree is the name of the entity. Each non-root node is a component or sub-component of the entity. Each link is a part -of relation. Each node is associated with a set of attributes. An opinion can be ex-pressed on any node and any attribute of the node.

Figure 1 System Architecture


## V. LATENT DIRICHLET ALLOCATION (LDA)

In the LDA model, each document is seen as a mixture of themes that are present in the corpus. The model proposes that each word in the document is attributable to one of the topics in the document.

For example, consider the following set of documents as the corpus:

Paper 1: I had a peanut butter sandwich for breakfast.

Paper 2: I like to eat almonds, peanuts and nuts.

Paper 3: My neighbor has a small dog yesterday.

Document 4: Cats and dogs are deadly enemies.

Document 5: You should not feed peanuts to your dog.

The LDA model discovers the different topics that documents represent and how much of each topic is present in a document. For example, LDA can produce the following results:
Topic 1: 30% peanuts, 15% almonds, 10% breakfast ... (you can interpret this as referring to food)
Topic 2: 20% of dogs, 10% of cats, 5% of peanuts ... (this topic may be interpreted as being about pets or animals)
Documents 1 and 2: 100% Theme 1
Documents 3 and 4: 100% Theme 2
Document 5: 70% Theme 1, 30% Theme 2

## VI. PSUEDOCODE

Check each document and randomly assign each word of the document to one of the topics K (choose K in advance)

This random assignment provides topical representations of all documents and word distributions of all subjects, although it is not very good. Therefore, to improve on them:

• For each document d, check each word w and calculate:

• P (subject t | document d): proportion of words in document d assigned to topic t

• P (word w | topic t): proportion of assignments to topic t, in all documents d, coming from the word w

• Reassign the word w to a new topic t ', where we choose the topic t' with probability

• P (theme t '| document d) * p (word w | theme t')

• This generative model predicts the probability that the topic t 'generates the word w

• Repeating the last step a large number of times, we arrive at a stable state where the assignments of subjects are quite good. These assignments are used to determine the thematic mixes of each document.

• After repeating the previous step a large number of times, eventually you will reach an almost stable state where your assignments are quite good. Use these assignments to estimate thematic mixes of each document (counting the proportion of words assigned to each topic within that document) and the words associated with each topic (counting the proportion of words assigned to each topic in general).

## VII. KEY INDEX PARAMETERS FOR RESULT CLASSIFICATION

In information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also called sensitivity) is the fraction of the relevant instances that are retrieved. Precision and recall are therefore based on understanding and measuring relevance. In simple terms, high accuracy means that an algorithm returns significantly more relevant than irrelevant results, while a high recall means that an algorithm has yielded the most relevant results.

The most important category measurements for binary categories are:

Table 1 Precision, Recall, FMeasure

| Precision | Recall | F Measure |
|---|---|---|
| $P = TP/(TP + FP)$ | $R = TP/(TP + FN)$ | $tp + tn/tp + tn + fp + fn$ |

## VIII. GRAPH BASED SCORE CALCULATION

Graph-based classification algorithms are a way of deciding the importance of a vertex within a graph, based on the information derived from the complete graph. The basic idea, implemented by a graphics-based classification model, is to "vote." When one vertex is linked to another, it is basically casting a vote for that other vertex. The greater the number of votes cast for a vertex, the greater the importance of the vertex. On the other hand, the importance of the vertex emitted by the voting determines how important is the vote itself, and this information is also taken into account by the classification model. Hence, the score associated with a vertex is determined on the basis of the votes cast for it, and the score of the vertices emitting these votes. Here, G = (V, E) is a directed graph with a set of vertices V and a set of edges E. For a given vertex, Vi, In (Vi) denotes the number of edges toward that vertex and Out (Vi) It denotes the number of outer edges of that vertex. D is the damping factor that is set at 0.85. Now, to allow the application of this model to natural language texts, we follow the steps:

1. Identify the text units that best define the task at hand and add them as vertices in the graph.
2. Identify the relationships that connect these units of text, and use these relationships to draw edges between vertices in the graph. The edges can be directed or not directed, weighted or unweighted.
3. Iterate the graphing-based classification algorithm to convergence.
4. Order the vertices according to their final score. Use the values attached to each vertex for classification / selection decisions.

This method allows us to obtain key phrases relevant to each document in the collection. Therefore, to obtain relevant themes from the entire collection, we apply the same procedure, where each vertex of the graph denotes relevant key phrases of the document.

With the explosive growth of social media on the Web, organizations are increasingly relying on opinion mining methods to analyze the content of these media for their decision-making. Look-based opinion mining, which aims to obtain detailed information on opinions, has attracted a great deal of attention from the scientific community and industry. The extraction of aspects and the extraction of entities are two of their main tasks. In this chapter, we have reviewed some representative works for the extraction of aspects and the extraction of entities from the opinion documents. For the extraction of aspects, existing solutions can be grouped into three main categories: (1) using language dependency rules, eg double propagation (Qiu et al., 2011). These methods use the relationships between aspects and words of opinion or other terms to perform the extraction of aspects. The approaches are not supervised and are independent of the domain. Therefore, they can be applied to any domain.

## IX. Conclusion

For the extraction of entities, supervised learning has also been the dominant approach. However, semi-supervised methods have come to the forefront lately. As in opinion mining, users often want to find competing entities for opinion analysis, they can provide some knowledge (eg entity instances) as seeds for semi-supervised learning. In this chapter, we have introduced the learning of PU and Bayesian Sets based on semi-supervised extraction methods. For evaluation, measures commonly used for extracting information, such as accuracy, recall and F-1 scores, are also frequently used in the extraction of aspects and entities. The results of the current F-1 score range from 0.60 to 0.85 depending on domains and data sets. Therefore, the problems, especially the extraction of aspects, remain very difficult. We expect future work to significantly improve accuracy. We also believe that semi-supervised and unsupervised methods will play a greater role in these tasks.

## References

[1] Quan Fang, Changsheng Xu, Fellow, IEEE, Jitao Sang, M. Shamim Hossain, Senior Member, IEEE, and Ghulam Muhammad, Member, IEEE Word-of-Mouth Understanding: Entity-Centric Multimodal Aspect-Opinion Mining in Social Media

[2] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis,” in Mining Text Data. New York, NY, USA: Springer, 2012, pp. 415–463.

[3] K. L. Keller, “Conceptualizing, measuring, and managing customer based brand equity,” J. Marketing, vol. 1, no. 1, pp. 1–22, 1993.

[4] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har’El, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, “Personalized social search based on the user’s social network,” in Proc.CIKM, 2009, pp. 1227–1236.

[5] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. KDD, 2004, pp. 168–177.

[6] S. Moghaddam and M. Ester, “On the design of lda models for aspectbased opinion mining,” in Proc. CIKM, 2012, pp. 803–812.

[7] Y. Fang, L. Si, N. Somasundaram, and Z. Yu, “Mining contrastive opinions on political texts using cross-perspective topic model,” in Proc.WSDM, 2012, pp. 63–72.

[8] X. Meng, F. Wei, X. Liu, M. Zhou, S. Li, and H. Wang, “Entity centric topic-oriented opinion summarization in Twitter,” in Proc. KDD, 2012, pp. 379–387.

[9] J. Dodge, A. Goyal, X. Han, A. Mensch, M. Mitchell, K. Stratos, K. Yamaguchi, Y. Choi, H. D. , III, A. C. Berg, and T. L. Berg, “Detecting visual text,” in Proc. HLT-NAACL, 2012, pp. 762–772.

[10] A. Sun and S. S. Bhowmick, “Quantifying visual-representativeness of social image tags using image tag clarity,” in Proc. Social Media Modeling Comput., 2011, pp. 3–23.

[11] J. Bian, Y. Yang, and T.-S. Chua, “Multimedia summarization for trending topics in microblogs,” in Proc. CIKM, 2013, pp. 1807–1812.

[12] Q. Hao, R. Cai, C. Wang, R. Xiao, J.-M. Yang, Y. Pang, and L. Zhang, “Equip tourists with knowledge mined from travelogues,” in Proc. WWW, 2010, pp. 401–410.

[13] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W. H. Hsu, and H.-Y. M. Liao, “Personalized travel recommendation by mining people attributes from community-contributed photos,” in Proc. ACM Multimedia, 2011, pp. 83–92.

[14] Q. Fang, J. Sang, and C. Xu, “Giant: Geo-informative attributes for location recognition and exploration,” in Proc. ACMMultimedia, 2013, pp. 13–22.

[15] Q. Li, J. Wu, and Z. Tu, “Harvesting mid-level visual concepts from large-scale internet images,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2013, pp. 851–858.

[16] L. Xie, A. Natsev, J. R. Kender, M. L. Hill, and J. R. Smith, “Visual memes in social media: Tracking real-world news in youtube videos,” in Proc. ACM Multimedia, 2011, pp. 53–62.

[17] G. Kim and E. P. Xing, “Visualizing brand associations from web community photos,” in Proc. WSDM, 2014, pp. 623–632.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” J. Mach. Learning Res., vol. 3, pp. 993–1022, 2003.

[19] D. M. Blei and M. I. Jordan, “Modeling annotated data,” in Proc. SIGIR, 2003, pp. 127–134.
[20] K. Barnard, P. Duygulu, D. A. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, “Matching words and pictures,” J. Mach. Learning Res., vol. 3, pp. 1107–1135, 2003.