Leveraging Multimodal Semantics and Sentiments Information in Event Understanding and Summarization

Rajiv Ratn Shah*, Debanjan Mahata, Mayank Meghawat and Roger Zimmermann

1Singapore Management University, Singapore

2Infosys Limited, USA

3Samsung Electronics, India

4National University of Singapore, Singapore

Submission: September 14, 2017; **Published:** September 27, 2017

*Corresponding author: Rajiv Ratn Shah, School of Information Systems, Singapore Management University, Singapore;
Email: rajivshah@smu.edu.sg

Abstract

Social media platforms are the witness of an abundance of user-generated content due to advancements in digital devices and affordable network infrastructures. Such platforms enable people to view, create, and share content with anyone at anytime and anywhere. Often user-generated content is the part of some activities, called events, which include some entities, and happen at some locations and times. Such events represent users' (both creators and viewers') behaviors, opinions, feedbacks, and preferences. Thus, social media companies necessitate an efficient analysis of user-generated content on social media to recommend personalized items such as posts, photos, videos, friends, and events to users automatically. For an efficient search, retrieval, and recommendation, it is necessary to understand semantics and sentiments knowledge structures from social media content (e.g., text posts, photos, and videos) since that help in decision-making, learning, and recommendations. Our exploration of event understanding suggests that multimodal information is very much useful for analyzing social media content since different representations of social media data represent different knowledge structures. In this study, we present our work on semantics and sentiments understanding from social media photos and leverage them in an efficient event detection and summarization.

Introduction

Recent years have been witness to the advent of the user-generated content (text posts, comments, photos, and videos) on social media websites. For instance, over one billion photos have been uploaded to Instagram, a very popular photo-sharing website. This happens due to the increasing popularities of social media websites, advancements in smart phones, and affordable network infrastructures. Social media companies enable users to create content and share with millions of other social media users around the world. Since for emotions (sentiments) aid learning, decision-making, and situation awareness in social media plat forms, it is essential to determine such information efficiently from social media content. However, it is very challenging to derive semantics and sentiments information because real-world user-generated photos are noisy and complex. Over a past few decades, researchers focus on affective computing, an area of artificial intelligence that attempts to provide machines with cognitive capabilities for recognizing, interpreting, inferring, and expressing sentiments and sentiments.

Since the affective computing is an interdisciplinary field and spans computer science, social sciences, psychology, and cognitive science, it is a very difficult problem. Thus, sentiments analysis and emotion recognition become a new trend in social media research and help users in understanding sentiments being expressed on different social media platforms. Similarly, since semantics computing is a field of computing that helps in understanding the intentions, meanings, and mapping of social content (e.g., social media photos), semantics computing is getting much research attentions. Recent work on semantics and sentiments analysis confirm that leveraging multimodal information in the understanding of social media content is very useful due to the distributed nature of social media.

Multimodal information is shown to be very useful in many social media analytics problems [1-3]. For instance, the multimodal information (both content and contextual information) helps in addressing multimedia summarization [4-6] and tag ranking and recommendation [6-9]. It is also useful in preference-aware multimedia recommendation [10-12] and multimedia-based e-learning [13-15]. We provide the detailed overview of multimodal information and their usage in addressing social media analytics problems in the following work [16-19]. Multimodal information is beneficial in semantics and sentiments analysis because different modalities exhibit different suitable concepts. Since it is possible now to collect a huge amount of important contextual information (e.g., temporal, spatial, and preferential information) on social media, we leverage both content and contextual information of user-generated content in our proposed approaches. Figure 1 depicts the framework of our multimodal semantics and sentiments engines for analyzing social media content in addressing several significant social media analytics problems.
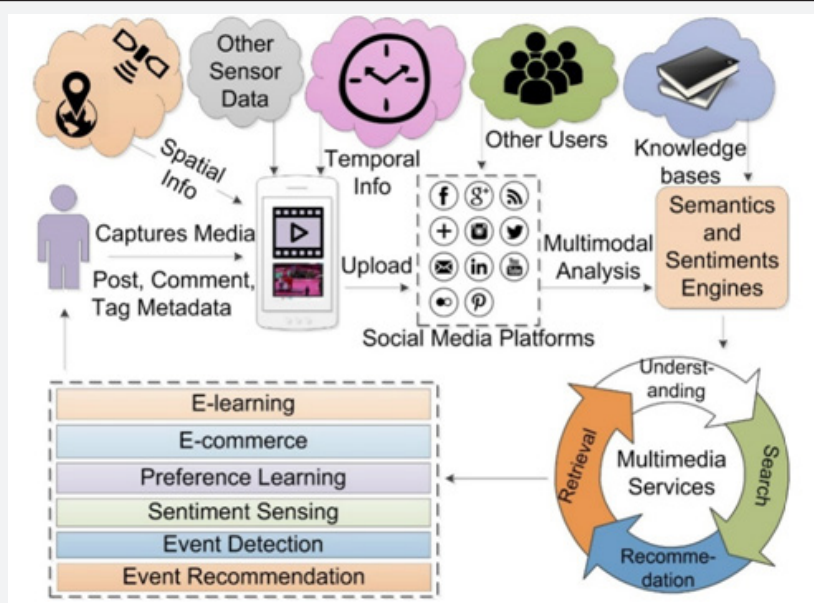

Figure 1: A multimodal framework for event understanding leveraging semantics and sentiments analysis.

Our proposed semantics and sentiments engines perform multimodal analysis of social media content in addition to exploiting existing knowledge bases. First, we present the Event Builder system [20] as our semantics engine that enables people to generate an event summary in real-time automatically. We exploit existing knowledge bases such as Wikipedia as the event background knowledge to obtain more contextual information about an event. Event Builder first performs an event detection leveraging multiple representations of social media content to identify events from social media photos. Shah et al. [21] suggested that exploiting multimodal information in event detection from social media photos helps in dealing with noise efficiently. Next in order to provide an overview of the event, we provide multimedia summarizations for the event. Multimedia summaries for the event include the visualization of photos with high relevance score on the Google map and text summaries for the event. We provide two text summaries for the event by solving an optimization problem with constraints such as the maximum summary length. First, the Flickr summary, which is derived from the descriptions of photos. It indicates what people generally think about the event. Second, the Wikipedia summary, which is derived from the Wikipedia article for the event. It indicates a brief overview of the event from Wikipedia. Subsequently, our semantics engine determines the recommendation and ranking of tags for social media photos.

Our proposed semantics engine computes tag relevance for social media photos. Specifically, first we recommend a list of tags to photos on social media such as Flickr, and subsequently, rank tags of photos based on their relevance to photos. Considering the most of the search-engines are still largely text-based and use tags (i.e., keywords) for searching and indexing multimedia content such as photos, it necessitates to have tag recommendation and ranking systems for social media photos. Social media platforms such as Facebook, Flickr, and YouTube allow users to annotate user-generated content such as photos and videos with tags (i.e., descriptive keywords), which assist in an efficient understanding, searching, and discovering of social media content. However, due to the manual annotation, ambiguous and personalized nature of user tagging, and without any restriction, tags of social media content are in random order. Moreover, often tags are inconsistent and irrelevant to the (both visual and textual) content.

Since manual annotation is a very time-taking process and even cumbersome for many users. Thus, search-engines tend to wrong recommendations and give inaccurate results for a given

user query. Thus, it is essential to perform semantics analysis on user-generated content on social media to compute tag relevance. To this end, first, we present a tag recommendation system, called PROMPT [22] which recommends personalized tags to users for a given photo. PROMPT leverages social contexts to compute relevance scores for photos and exploits personal contexts to provide personalized results. Personalized results are important in search-engines since often users tend to use different words (i.e., synonyms) to describe the same thing. Subsequently, we present a tag ranking system, called CRAFT [23], which ranks tags of social media photos based on voting from neighboring photos derived from multiple sources. Thus, we exploit multimodal information of social media content in our proposed tag relevance computation approaches. Experimental results confirm the effectiveness of our tag recommendation and ranking systems. Finally, we present a sentiments engine, called Event Sensor, which determine emotions and sentiments from social media photos to provide sentiments-based multimedia summaries.

The Event Sensor system [24] aims to solve sentiments understanding and generates a multimedia summary for a given mood and an event. It extracts concepts and mood tags from the visual content and textual metadata of user-generated content and exploits them in supporting several significant multimedia analytics problems such as a musical multimedia summary.

Event Sensor supports sentiment-based event summarization by leveraging Event Builder as its semantics engine component. The Event Builder system, first, determines event summaries and subsequently, the sentiments engine of Event Sensor determines emotional (mood) tags for photos of events. The majority decides the decision of overall mood tags for events. Moreover, users can also get the multimedia summarization for a given mood tag. Finally, to make the viewing experience more interesting, Event Sensor attaches soundtracks to the slideshow of photos derived using the semantics engine for given user inputs such as events or mood tags. We use the following representative emotional tags in our sentiments engine: anger, disgust, joy, sad, surprise, and fear. Experimental results confirm that our semantics and sentiments engines provide good results and helpful in providing an efficient solution for several social media analytics problems.

Knowledge Bases and APIs

Here we present an overview of the following knowledge bases used in our semantics and sentiments understanding: (i) Word Net, (ii) Semantics Parser, (iii) Sentic Net, and (iv) Foursquare Geo-categories. Word Net is a popular lexical database of English [4]. It constructs syn sets by grouping nouns, adverbs, verbs, and adjectives into sets of cognitive synonyms. Thus, each syn set represents a distinct concept and groups words together based on their meanings. Moreover, Word Net labels the semantic relations among words. For example, automobile, railcar, motorcar, railway car and railroad car are the syn sets

for the word "car". Since a natural language text is a collection word and the combination of words makes phrases or keywords, it is essential to determine semantics concepts from the texts in terms of phrases. Such semantics concepts also carry sentiments information. Thus, researchers often perform semantics parsing of the natural language text for a better semantics and sentiments analysis. They extract concepts from the text by first extracting linguistic patterns from deconstructing text into meaningful pairs such as VERB+NOUN, NOUN+NOUN, and ADJ+NOUN. Subsequently, they exploit commonsense knowledge to decide which pairs are more relevant in the current context. Poria et al. [25] presented a semantics parser that derives phrasesor multi-word concepts from a given natural language text. For example, for a given natural language text, "Google Inc. is an American multinational technology company that specializes in Internet-related services and products.", the semantics parser generates the following concepts: technology company inc., American technology company, multinational technology company, internet related service, and others. To provide the semantics and affective labels for common and commonsense concepts, Poria et al. [26] presented SenticNet-3 which consists of approximately 30000 common and commonsense concepts such as accident happen, party, and food [17,18]. For a given concept, their Sentic Net API provides the following sentiments information: pleasantness, attention, sensitivity, and aptitude [19]. This API also provides five semantically related concepts for each Sentic Net concept. Using Emo Sentic Net, Poria et al. [20] mapped SenticNet-3 concepts to six affective labels such as anger, disgust, fear, joy, sadness, and surprise. For an efficient semantics and sentiments analysis, SenticNet knowledge base and WordNet are merged together [21,22]. In order to derive semantics and sentiments concepts from other modalities, we derive knowledge structures from locations since the most of user-generated content captured by smart phones and modern devices are geo-tagged. Specifically, we derive geo-categories using Foursquare API (https://foursquare.com) which provides a list of geo-categories such as Building, Restaurant, and Park for a given GPS point. Geo-categories help in an efficient semantics understanding (e.g., scenes in user-generated multimedia content) and sentiments understanding (e.g., mood tags, since often location surroundings indicate the emotions of the environment). Foursquare provides include ten high-level categories (e.g., restaurant) and approximately 1300 low-level categories (e.g., Chinese restaurant, Indian restaurant, Thai restaurant and Italian restaurant).

Problems and Solutions

In this study, we mainly focus on semantics and sentiments understanding from social media content such as photos. Since social media content is very much noisy as compared to traditional multimedia content, it is important to reduce noises and derive only relevant knowledge structures. Since our approaches are based on an intuition that different representations of the social media content exhibit different knowledge structures,

we exploit multimodal information of user-generated content in our approaches. Particularly, we exploit information from different modalities such as texts (e.g., descriptions and tags), videos, photos, audio, and context. Our study depicts that leveraging existing knowledge bases such as Sentic Net, Word Net, Foursquare geo-categories to solve different social media analytics problems such as event detection and summarization, tag recommendation, and tag ranking. The following are details of our approaches.

Leveraging Semantics Engine in Event Detection and Summarization

The advent of social media content such as textual posts, photos, and videos on the web has garnered much research attention in the recent years due to the presence of a significant number of people on such websites. Their presence has enabled social media companies to learn users' behaviors, needs, and preferences. Since a significant portion of revenues for social media companies comes from the advertisements, it is essential for them to learn accurately the needs and preferences of users. Thus, such companies invest heavily on semantics and sentiments understanding of user-generated content on social media, which help them in providing the most relevant recommendations to social media users. Due to ubiquitous availability of modern digital devices and affordable Internet infrastructures, people share social media content with anyone at anytime and anywhere. Since such social media content is often the part of some events, and users' preferences, needs, and behaviors exhibited by the attended events, it necessitates identifying events from social media content automatically. To this end, we presented the Event Builder system that performs event detection from social media photos leveraging multimodal information using its semantics engine. Multimodal information helps in dealing with noises and complexities of real-world social media photos. In addition to event detection, derived semantics and sentiments knowledge structures are also useful in multimedia search, retrieval, summarization, and recommendation of social media content and events. Semantics engine in Event Builder produces multimedia summaries for a given event in real-time by leveraging information from different social media platforms such as Wikipedia and Flickr.

Figure 2 shows the framework of the semantics engine in the Event Builder system, which consists of two components: (i) an offline event detection component and (ii) an online processing component that leverages results from the event detection component to visualize and compute events summaries. The event detection component first derives semantics knowledge structures from multimodal information and exploits them in detecting events from social media photos. Specifically, Event Builder leverages both content and contextual information since the only content is not enough to describe social media photos fully. It is based on intuition that the semantics and sentiments of social media content is much affected by the context such as likes, comments, views, temporal, and spatial information.
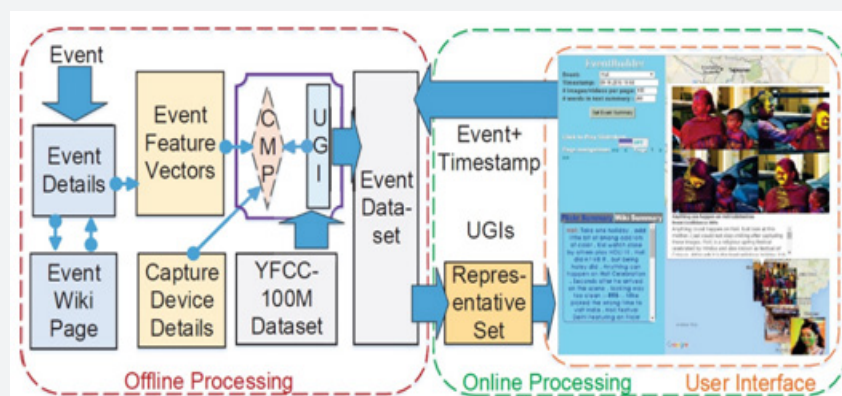

Figure 2: The framework of semantics engine in Event builder for event summarization [16].

The semantics engine in Event Builder exploits the information from the following sources. First names of events, since they describe events in short phrases or sentences. Second, keywords of events, since most social media platforms enable users to describe events by a few phrases and words. Third, spatial information, since locations of photos often provide important semantics and sentiments information. Fourth, temporal information, since times of events describe the preferences of users and importance of events at any given time. Finally, fifth, device characteristics that captured the photos and videos, since the qualities of social media content are highly dependent on the features of capturing devices.

The semantics engine computes relevance scores of photos for a given event based on these five criteria. It assigns different weights to different modalities and fuses relevant scores from different sources to get the overall similarity scores of social media photos for the given event. Our study indicates that similarity based on names of events is the most salient features in detecting events from social media photos. Since the similarity based on the characteristics of devices are mainly useful in boosting relevant scores for photos of high-quality content, they are mainly useful in visualization of high-quality photos and not much useful in event detection. Thus, we assign only the 5% weight age to relevant scores derived from the characteristics of

devices in overall relevant scores. The overall relevance scores of photos used in offline event detection (see the left block of the Figure 2. The online component takes a list of photos (a representative set) with high relevant scores for the given event to visualize and summarize the event.

Event Builder provides the visualization and text summaries for events from photos in the representative set as the part of multimedia summary for a given event. The right block of the Figure 2 depicts online event summaries for the event, which consists of two sections. First, we visualize photos from the representative set with high relevance scores for events on the Google map. Subsequently, Event Builder solves an optimization problem to generate text summaries for events from the textual descriptions of photos. Experimental results (including both objective and subjective evaluations) confirm that Event Builder outperforms baselines. We have used the YFCC100M (Yahoo Flickr Creative Commons 100 Million) dataset from the Flickr for the evaluation of Event Builder. The YFCC100M dataset is a collection of 100 million photos and videos from Flickr and provides several significant metadata information such as user tags, visual tags (automatically derived tags from the visual content), spatial information, and temporal information social media content in this dataset are from top cities such as Tokyo, New York, London, and Hong Kong. We consider the following

seven representative events in our evaluation: (i) Occupy Movement, (ii) Holi, (iii) Hanami, (iv) Olympics Games, (v) Byron Bay Bluesfest, (vi) Batkid, and (vii) Eyjafjallajkull Eruption. The detailed information of the EventBuilder system is provided inour earlier work [23,24].

Leveraging Semantics Engine in Recommending Tags for Social Media Photos

Most search-engines still use bag-of-words models to represent multimedia content such as photos. They index photos using tags (i.e., keywords or multi-word phrases).However, a significant number of photos on social media is not annotated because manual annotation is cumbersome, time taking, incorrectly spelled, and noisy. Moreover, often people describe the same thing with different tags. Thus, for an efficient working of search engines, it would be useful to predict tags for social media photos automatically. This will ease the life of user's as well social media companies in searching multimedia content from a vast amount of multimedia content on the web. To this end, we present the PROMPT system that predicts personalized tags for a given photo leveraging personal and social contexts. Figure 3 depicts the framework of semantics engine in PROMPT. Particularly, first, PROMPT determines a group of users who have similar tagging behaviors as a user (who capture the photo), and refer this as the personal contexts of the user.
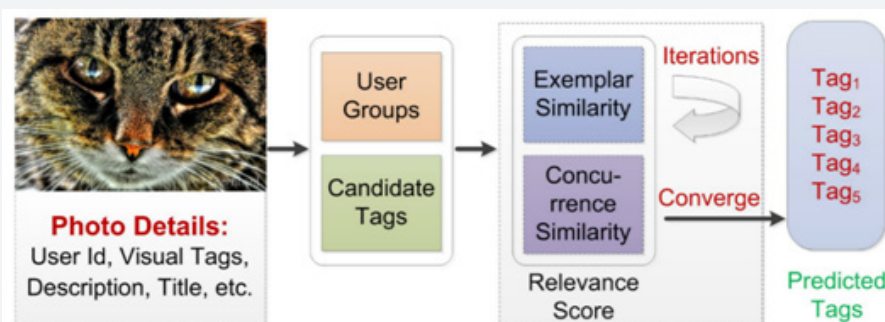

Figure 3: The framework of semantics engine in PROMPT for tag recommendation [23].

Similar to the idea of collaboration filtering, we assume that users who use similar tags to annotate photos should have the same tagging behaviors. Thus, the personal contents of users help our semantics engine in predicting personalized tags for users' social media photos. Next, it determines candidate tags from social contexts of photos leveraging multimodal information such as visual content, textual metadata, and tags of neighboring photos. Finally, we compute relevance scores of candidate tags to predict the five most suitable tags to social media photos. The steps to compute the tag relevance are as follows. First, we initialize relevance scores of candidate tags using asymmetric tag co-occurrence probabilities. Next, we normalize scores of tags after voting from neighboring photos. Next, we perform a random walk to promote the tags that have many close neighbors and weaken isolated tags.

Finally, we predict top five user tags with high-relevance scores to the given photo. Our earlier work [25] describes the

detailed descriptions of PROMT. For evaluation, we consider of the most frequent 1,540 user tags from the YFCC100M dataset from Flickr since recommending tags from an endless pool is a NP-hard problem. Experimental results confirm that tags predicted by PROMPT describe social media photos well.

Leveraging Semantics Engine in Ranking Tags of Social Media Photos

As described earlier, most search-engines still use bag-of-words models to represent multimedia content. Thus, searching social media photos using only bag-of-words models are not sufficient to provide accurate results for a given query. Since not all tags (words) of photos are of the same importance, it is important to rank keywords of social media photos in the order of their importance. The ranked tags help in an efficient searching, retrieval, and recommendation of social media content. To this end, we present the CRAFT system that ranks tags of a given

photo based on voting from its neighboring photos derived from multimodal information such as visual content, spatial information, and semantics information.

Particularly, CRAFT finds photo neighbors leveraging multimodal information such as geo, visual, and semantics concepts derived from spatial information, visual content, and textual metadata, respectively. Fig. 4depicts the framework of the semantics engine in the CRAFT system. We leverage the proposed high-level features instead traditional low-level features to rank tags of photos because high-level representations carry more semantics information than low-level representations of multimedia content. For the evaluation of CRAFT, we consider 200,000 photos (that have multimodal information such as geo, visual, and semantics concepts) from the YFCC100M dataset. We leverage the Foursquare API to determine geo concepts (e.g., Hotel, Park, and Beach) from social media photos for a given GPS point (i.e., the location where the photo is captured). Next, visual concepts (or tags) are automatically generated tags from the visual content of social media photos using deep learning models such as Google Cloud Vision.
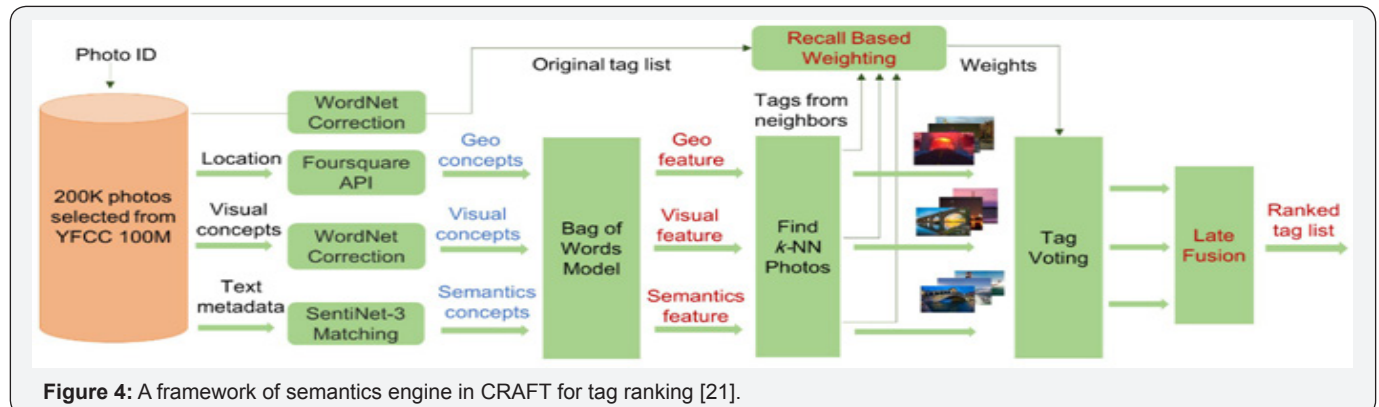

Figure 4: A framework of semantics engine in CRAFT for tag ranking [21].

For instance, the Google Cloud Vision API (https://cloud.google.com/vision/) able to derive semantic and sentiment concepts from photos. However, visual concepts for photos are provided as the part of the YFCC100M dataset. Finally, semantics concepts are common and commonsense tags from the SenticNet-3 knowledge base. Semantics concepts for social media photos are derived from mapping semantics concepts (extracted using the semantics parser) to SenticNet-3 concepts. Next, we find k-nearest neighbors from the dataset with 200,000 photos and perform tag voting from the neighboring photos. Finally, we perform the late fusion to get the ranked list of tags for photos. Our earlier work [26] describes the full details of CRAFT. Experimental results confirm that CRAFT outperforms state-of-the-art models and the produced ranked tags significantly improve the understanding of social media photos. Moreover, our neighbor voting for tag ranking helps in removing non-relevant and noisy tags (Figure 4).
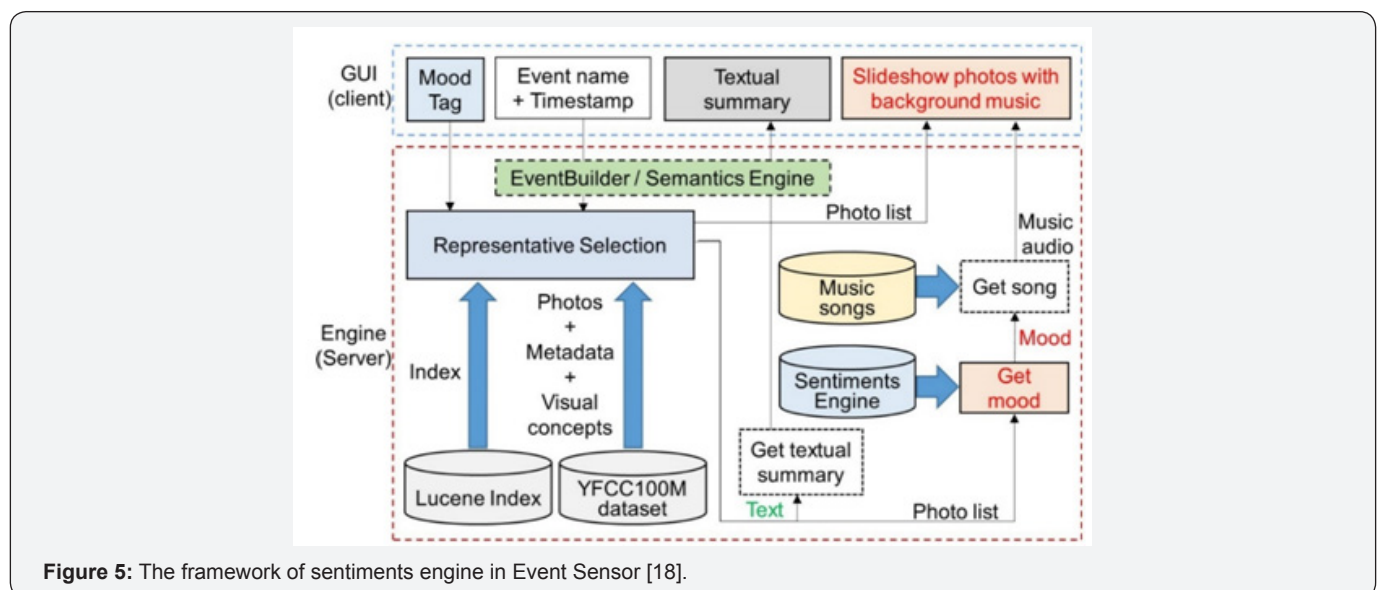

Figure 5: The framework of sentiments engine in Event Sensor [18].

For sentiments understanding from social media photos, we present the Event Sensor system. The sentiments engine in Event Sensor produces a sentiment-based multimedia summary for a given event or mood tag. It derives emotional or mood tags from visual content and textual metadata of social media photos. Event Sensor exploits such knowledge structures (i.e., concepts

and sentiments) in addressing several significant multimedia analytics problems such as in providing the musical multimedia summary. Event Builder and Event Sensor together build a novel system that provides both semantics and sentiment services from social media content. For instance, Figure 5 shows the framework of the Event Sensor system that Event Builder as its semantics engine.

The Event Sensor system consists of the following two components: (i) a GUI client and (ii) a backend server. To display the sentiments-based multimedia summarization to users, first, the GUI client accepts users' inputs such as mood tags, timestamps, and dates. Next, the GUI client communicates with the backend component to get the representative set of photos for given inputs using the semantics engine of Event Sensor (i.e., Event Builder). Subsequently, the sentiments engine finds mood tags for each photo in the representative set. Finally, the sentiments engine in the Event Sensor system selects the

soundtrack for the most frequently occurred mood tags for social media photos and builds a slideshow of photos with a recommended soundtrack. Since music enhances the viewing experience of multimedia content, the produced musical soundtrack enhances the viewing experience of multimedia summaries for given events or queries.

Figure 6 shows the framework of the sentiments (also known as sentics) engine in the Event Sensor system. It is useful in providing sentiment-based multimedia-related services to users from social media content accumulated on social media websites. In our proposed sentiments engine, we leverage information from both multimedia content and contextual information since determining sentiments and emotions from social media photos is very difficult due to the following reasons. First, it is very subjective and second, relying only on one modality does not give very accurate sentiments prediction.
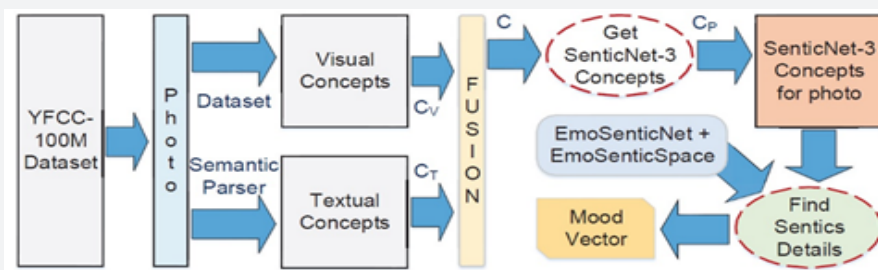

Figure 6: The framework of affective computing in the Event Sensor system [18].

The steps to determine emotional labels from social media content are as follows. First, Event Sensor determines keyword phrases (i.e., semantics concepts) from multimodal information (i.e., both content and contextual information) of social media photos using a semantics parser [27], and fuses them. However, such derived semantics concepts result in an endless pool of concepts and it is not feasible to consider all such possible concepts produced by the semantics parser in the sentiments analysis. Thus, we map all derived concepts to a fixed set of 30,000 common and commonsense concepts of SenticNet-3. SenticNet-3 is a well-known knowledge base and consists of semantics and sentiments information. Specifically, it bridges the conceptual and affective gap. Next, Event Sensor exploits the Emo Sentic Net knowledge base [28], which maps concepts of SenticNet-3 to the following six affective tags such as anger, disgust, joy, sad, surprise, and fears. Leveraging these six emotional tags, we buildsix-dimensional sentiments vectors from the multimodal information of social media photos. We select the most dominant mood tags from all social media multimedia content (e.g., photos) in the representative set to construct a musical slideshow. Similar to the evaluation of Event Builder, we used the YFCC100M dataset in the evaluation of our Event Sensor system. We used the ISMIR'04 dataset of 729 songs from the ADVISOR system [29] to generate a musical multimedia summary (i.e., sentiments based summaries for events and mood tags). We use the 20 most frequent mood tags from the Last.

fm, a well-known website for music. Evaluation results confirm that Event Sensor outperforms baselines. Our earlier work [30] describes the full details of Event Sensor.

Future Research Directions

Since an event is an activity that takes place at some time, in some location, and involves entities such as people, animals, scenery, and building, people often capture these activities through social media content. For example, users capture texts (e.g., tweets, blogs, and posts), photos, and videos at some location and share with other users any time on social media platforms such as Twitter, Facebook, and Flickr.

This becomes possible due to advancements and affordability of digital devices, network infrastructure, and social media websites. As described earlier in this study, the multimodal information of social media content shows users' behaviors and preferences on the web, which is very useful in providing personalized search, retrieval, and recommendation? Thus, in addition to event detection, summarization, and recommendation, the multimodal information has shown its usefulness in news videos uploading and map matching to benefit the society. However, in this study, we mainly focus on events from social media content. Thus, in the future, we would like to discover event-specific informative content from social media to facilitate event analysis. In our current frameworks for

event detection and summarization, we ignored tags of social media photos that indicate semantics and sentiments knowledge structures of a great importance. In the future, we would like to integrate our semantics and sentiments engines for events with that of tag recommendation and ranking to uncover hidden knowledge structures for an efficient event understanding.

Mahata, Talburt [31-33] proposed a framework for collecting and managing entity identity information from social media. Such entity identity information can help us in an efficient event search, retrieval, and recommendation since the characteristics of different events are different in nature. Since the mining of blogosphere from a socio-political perspective yields good results [34], we would like to exploit event-specific information from different perspectives and sources to address event-related problems in social media. Moreover, since deep neural networks (DNN) have produced immense success in several research areas of computer science such as computer vision, NLP, and speech processing, we would like to explore deep learning techniques to derive useful information from the multimedia content in event analysis [35,36].

We believe that DNN-based new representations for multimedia content should be very useful in addressing several multimedia analytics problems. Inspired by the SMS based FAQ systems, we would also like to build a messaging service for FAQs for events in the future. For an efficient mapping of events and users' queries, we would like to leverage multimodal information such as location, content, and contextual information. In our proposed event FAQ-based system, there would be the following three components. First, an analyzer that analyzes the received SMS, MMS, and other user queries. Second, a controller that computes similarities leveraging multimodal information. Finally, third, a recommender component that recommends a list of events after correlating them with users' preferences and queries. In the long term, we would like to improve our semantics engine for tag recommendation and ranking, we would like to use DNN-technologies since a recent study Zheng et al. [37] confirms that DNN technologies are very useful in improving recommendation. Particularly, we would like to build a general DNN-based recommendation framework that should be able to support all kind of recommendation problems on social media. In this proposed framework, we would like to fuse information efficiently from multiple modalities to provide state of the art recommendations for different social media problems.

Conclusion

This study presented a multimodal semantics and sentiments computing on social media photos. We presented our novel frameworks to address several significant research problems. Specifically, our proposed semantics engine helps in event detection, event summarization, tag recommendation, and tag ranking. Moreover, our proposed sentiments engine assists in providing sentiments-based multimedia summarization. Experimental results confirm that leveraging multimodal

information is very useful in deriving knowledge structures from social media content such as photos. Particularly, we exploited contextual information in addition to the content information in an efficient semantics and sentiments understanding. Finally, we proposed several interesting social media analytics problems to readers as future research directions.

References

1. Shah RR (2016) Multimodal Analysis of User-Generated Content in Support of Social Media Applications. In International Conference on Multimedia Retrieval pp: 423-426.

2. Shah RR (2016) Multimodal-based Multimedia Analysis, Retrieval, and Services in Support of Social Media Applications. In Multimedia Conference 1425-1429.

3. Shah R, Zimmermann R (2017) Multimodal Analysis of User-Generated Multimedia Content.

4. Shah RR, Shaikh AD, Yu Y, Geng W, Zimmermann R, Wu G (2015) Eventbuilder: Real-time multimedia event summarization by visualizing social media. In International Conference on Multimedia 185-188.

5. Shah RR, Yu Y, Verma A, Tang S, Shaikh AD, Zimmermann R (2016) Leveraging multimodal information for event summarization and concept-level sentiment analysis. In Knowledge-Based Systems 108: 102-109.

6. Shah R, Zimmermann R (2017) Event Understanding. In Multimodal Analysis of User-Generated Multimedia Content pp: 59-99.

7. Shah RR, Yu Y, Tang S, Satoh SI, Verma A, Zimmermann R (2016) Concept-Level Multimodal Ranking of Flickr Photo Tags via Recall Based Weighting. In Workshop on Multimedia COMMONS 19-26.

8. Shah RR, Samanta A, Gupta D, Yu Y, Tang S, Zimmermann R (2016) PROMPT: Personalized User Tag Recommendation for Social Media Photos Leveraging Personal and Social Contexts. In International Symposium on Multimedia (ISM) pp: 486-492.

9. Shah R, Zimmermann R (2017) Tag Recommendation and Ranking. In Multimodal Analysis of User-Generated Multimedia Content pp: 101-138.

10. Shah RR, Yu Y, Zimmermann R (2014) Advisor: Personalized video soundtrack recommendation by late fusion with heuristic rankings. In International Conference on Multimedia 607-616.

11. Shah RR, Yu Y, Zimmermann R (2014) User preference-aware music video generation based on modeling scene moods. In Multimedia Systems Conference 156-159.

12. Shah R, Zimmermann R (2017) Soundtrack Recommendation for UGVs. In Multimodal Analysis of User-Generated Multimedia Content pp: 139-171.

13. Shah RR, Yu Y, Shaikh AD, Tang S, Zimmermann R (2014) ATLAS: automatic temporal segmentation and annotation of lecture videos based on modelling transition time. In International Conference on Multimedia 209-212.

14. Shah RR, Yu Y, Shaikh AD, Zimmermann R (2015) TRACE: Linguistic-Based Approach for Automatic Lecture Video Segmentation Leveraging Wikipedia Texts. In International Symposium on Multimedia (ISM), 217-220.

15. Shah R, Zimmermann R (2017) Lecture Video Segmentation. In Multimodal Analysis of User-Generated Multimedia Content pp: 173-203.

16. Shah R, Zimmermann R (2017) Introduction. In Multimodal Analysis of User-Generated Multimedia Content 1-30.

17. Shah R, Zimmermann R (2017) Literature Review. In Multimodal Analysis of User-Generated Multimedia Content pp: 31-57.

18. Shah R, Zimmermann R (2017) Conclusion and Future Work. In Multimodal Analysis of User-Generated Multimedia Content pp: 235-260.

19. Miller GA (1995) WordNet: a lexical database for English. In Communications of the ACM 38(11): 39-41.

20. Poria S, Cambria E, Gelbukh A, Bisio F, Hussain A (2015) Sentiment data flow analysis by means of dynamic linguistic patterns. In Computational Intelligence Magazine 10(4): 26-36.

21. Poria S, Gelbukh A, Hussain A, Howard N, Das D, et al. (2013) Enhanced SenticNet with affective labels for concept-based opinion mining. In Intelligent Systems 28(2): 31-38.

22. Cambria E, Olsher D, Rajagopal D (2014) SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In AAAI Conference on Artificial Intelligence.

23. Cambria E, Poria S, Gelbukh A, Kwok K (2014) Sentic API: a common-sense based API for concept-level sentiment analysis.

24. Cambria E, Livingstone A, Hussain A (2012) The hourglass of emotions. In Cognitive Behavioral Systems 144-157.

25. Poria S, Gelbukh A, Cambria E, Hussain A, Huang GB (2014) EmoSentic Space: A novel framework for affective common-sense reasoning. In Knowledge-Based Systems 69: 108-123.

26. Poria S, Gelbukhs A, Cambria E, Das D, Bandyopadhyay S (2012) Enriching SenticNet polarity scores through semi-supervised fuzzy clustering. In International Conference on Data Mining Workshops (ICDMW) 709-716.

27. Poria S, Gelbukh A, Cambria E, Yang P, Hussain A, et al. (2012) Merging Sentic Net and Word Net-Affect emotion lists for sentiment analysis. In International Conference on Signal Processing (ICSP) 2: 1251-1255.

28. Shah RR, Hefeeda M, Zimmermann R, Harras K, Hsu CH, Yu Y (2016) NEWSMAN: Uploading Videos over Adaptive Middleboxes to News Servers in Weak Network Infrastructures. In International Conference on Multimedia Modeling pp: 100-113.

29. Shah R, Zimmermann R (2017) Adaptive News Video Uploading. In Multimodal Analysis of User-Generated Multimedia Content pp: 205-234.

30. Yin Y, Shah RR, Zimmermann R (2016) A general feature-based map matching framework with trajectory simplification. In SIGSPATIAL International Workshop on Geo Streaming p. 7.

31. Mahata D, Agarwal N (2012) What does everybody know? Identifying event-specific sources from social media. In International Conference on Computational Aspects of Social Networks 63-68.

32. Mahata D, Talburt JR, Singh VK (2015) From Chirps to Whistles: Discovering Event-specific Informative Content from Twitter. In ACM Web Science Conference 17.

33. Mahata D, Talburt J (2014) A Framework for Collecting and Managing Entity Identity Information from Social Media. In MIT International Conference on Information Quality 216-233.

34. Shaikh AD, Jain M, Rawat M, Shah RR, Kumar M (2013a) Improving accuracy of SMS based FAQ retrieval system. In Multilingual Information Access in South Asian Languages pp. 142-156.

35. Shaikh AD, Shah RR, Shaikh R (2013b) SMS based FAQ retrieval for Hindi, English and Malayalam. In Workshops of the Forum for Information Retrieval Evaluation p. 9.

36. Singh VK, Mahata D, Adhikari R (2010) Mining the blogosphere from a socio-political perspective. In International Conference on Computer Information Systems and Industrial Management Applications (CISIM) pp. 365-370.

37. Zheng L, Noroozi V, Yu PS (2017) Joint deep modeling of users and items using reviews for recommendation. In International Conference on Web Search and Data Mining 425-434.

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php