International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

# Social Media Sentiments Using Latent Correlation Analysis

**Miss. Sonali B. Gaikwad**
Department of Computer Engineering
SND College of Engineering and Research Centre, Yeola, Nashik

**Prof. Vikas N. Dhakane**
Department of Computer Engineering
SND College of Engineering and Research Centre, Yeola,Nashik

***ABSTRACT-*** *Rapidly development of online social networks, millions of social media users like to upload or post or share their daily emotional experiences through text message, images and videos using different social media networks such as twitter, facebook. For the developing various applications the analysis of sentiments in social media users created images are increasingly important. Insufficient investigate has been carried out on the use of text data associate with the images uploaded or posted by the social users, sufficient investigate focusing on the designing of visual features, various conventional methods on image sentiment analysis are used. In this paper a novel approach which exploits latent correlation analysis of multi-views, we will works on sentiment analysis depends on visual and textual contents in social media. From textual and visual contents we were proposing hybrid approach of aggregating sentiments. In this proposed method we employing latent correlation analysis of multi-views (textual and image) to maximized latent correlations between multi-views.*

***KEYWORDS - Correlation analysis, Textual Sentiment, Image Sentiments, Microsoft Cognitive, Multi-view.***

## 1. INTRODUCTION

Sentiment analysis (SA) analyzes the user's emotions, attitudes, opinions, evaluations from the language written by the users in social networks [2]. Now a day's social networks are major platforms of communicating and exchanging the information between users. Large number of textual as well as visual information is provided by social networks which making it possible to detect sentiment. To express user's feelings or emotions they posted images, video and text information [3]. On sentiment dictionary lot of work has been done. To detect the sentiment currently the analysis of textual content technique are used. In mainstream online social network visual content become popular. For daily lives record user using images, we can identify the user emotional wellness depends on the sentiment and emotion from images posted on social networks by the user.

Daily millions of images are posted. The term sentiment analysis first appears in [4] due to exchange of information and communication through social media, researchers give the access of large number of text and images which express users opinions and sentiments. Thus in research sentiment analysis has important impact on Natural Language Processing (NLP), Due to affection by opinions in management sciences, political science, economics, and social sciences they all have profound impact. Users collecting different digital images and posed them on social networks such as Twitter. User uploaded or posted an image that gives the visual aspects of daily lives. Such user created images has information source for analyzing sentiment and opinions, which enables various applications including opinion mining about social activities, marketing of product and human machine interaction affectively [5].

In recent years research attention is increasing due to the received images has automatic inference of the sentiment [6, 7, 8]. For the training sentiment polarity classifiers we have to design effective visual features from conventional methods of image sentiment analysis [6,7,11]. It is very difficult for direct association of the visual features with sentiment labels due to affective gap between high-level human sentiments concepts and low-level visual features. For improving Image content recognition we have to focus on textual features around training images using collaboratively [11, 12].

In this paper, we present a novel sentiment analysis uses latent correlation among multiple views (visual, textual). In this proposed method, we make the separation of features from the pairs of images and text in the URL's and by using Microsoft Cognitive API for making image sentiments and textual sentiments of tweets by using R as positive, negative or neutral. By Latent correlation analysis we employ the rule- base which is nothing but multi-view embedding and defining polarity through the polarity classification and extracting final emotions of tweets.

## 2. LITERATURE REVIEW

Image Sentiment Analysis Using Latent correlations among visual textual, and sentiment views in the paper authors represent a method of image sentiment analysis that using latent correlations among visual, textual, and sentiment views of training images. By finding latent embedding space authors maximized the correlations between multiple views [1].

Yilin Wang, Baoxin Li [2] and You Q. Luo J [3] study the problem of understanding human sentiments from large scale collection of internet images based on both image features and contextual social network information. Despite the great strides in analyzing user sentiment based on text information, the analysis of sentiment behind the image content has largely been ignored. Thus, authors extend the significant advances in text-based sentiment prediction tasks to the higher level challenge of predicting the underlying sentiments behind the images. Authors show that neither visual features nor the textual features are by themselves sufficient for accurate sentiment labeling. Thus, they provide a way of using both of them, and formulate sentiment prediction problem in two scenarios: supervised and unsupervised. Authors develop an optimization algorithm for finding a local-optimal solution under the proposed framework. With experiments on two large-scale datasets, they show that the proposed method improves significantly over existing state-of-the-art methods. In the future, we are going to incorporating more information on the social network and explore sentiment on signed social network.

In this paper E. Cambria, B. Schuller, Y. Xia, and C. Havasi discusses relevant techniques and tools of opinion mining and sentiment analysis [5].

In this work S. Siersdorfer, E. Minack, F. Deng, and J. Hare study the visual content in social photo which are sharing on Flickr, also the relation between sentiment images which are presented in the form of metadata [6].

This paper focuses on sentiments to understand visual concepts related to sentiments. For semantic concepts D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang constructed a largescale Visual Sentiment Ontology (VSO) [7].

In this paper with the help of images posted on web depends on visual and textual contents understanding human sentiments. Either visual or textual contents are not sufficient for labeling of sentiment accurately therefore they formulate sentiment prediction problem as supervised and unsupervised, so for finding local optimal solution Y. Wang, S. Wang, J. Tang, H. Liu, and B. Li developed an optimization algorithm [9].

Learning semantic representation to the social media images and text data associated with it by using kernel canonical correlation analysis method. Comparison between the images and text, common representation is provided by the semantic space. From the text query authors retrieving images depends on their content only, also comparing orthogonalization against a standard cross-representation retrieval technique known as the generalized vector space model [13].

In this paper A. Rahimi and B. Recht propose mapping input data to the randomized low dimensional feature space by applying fast liner methods for converting the training of kernel machines. Approximately the inner products of the transformed data is equal to shift invariant kernel specified by user randomized features are designed. They propose simple way for combining advantages of the linear and nonlinear approaches. Two sets of random features are exploring [14].

## 3. PROBLEM STATEMENT

Now a days millions of user tweets by using textual as well as visual sentiments. As visual sentiment are more effective than textual one, so there is need to find out actual correlation between them. From given a tweets, classify whether the tweets sentiment are positive, negative, or neutral. For tweets conveying both a positive and negative sentiment, whichever is the stronger sentiment should be chosen while choosing the stronger sentiment we required, the visual sentiments.

## 4. SYSTEM ARCHITECTURE

In this section we employing latent correlation analysis of multi-views. To maximized latent correlations between multiple views, we find latent embedding space. The System architecture of proposed system is shown in Fig. 1. An overview of the proposed method described in following sections.

A) Grab the real time tweets

B) Storage of Twitter Stream

C) Textual opinion mining of tweets

D) Latent correlation analysis

E) Visualization of sentiments

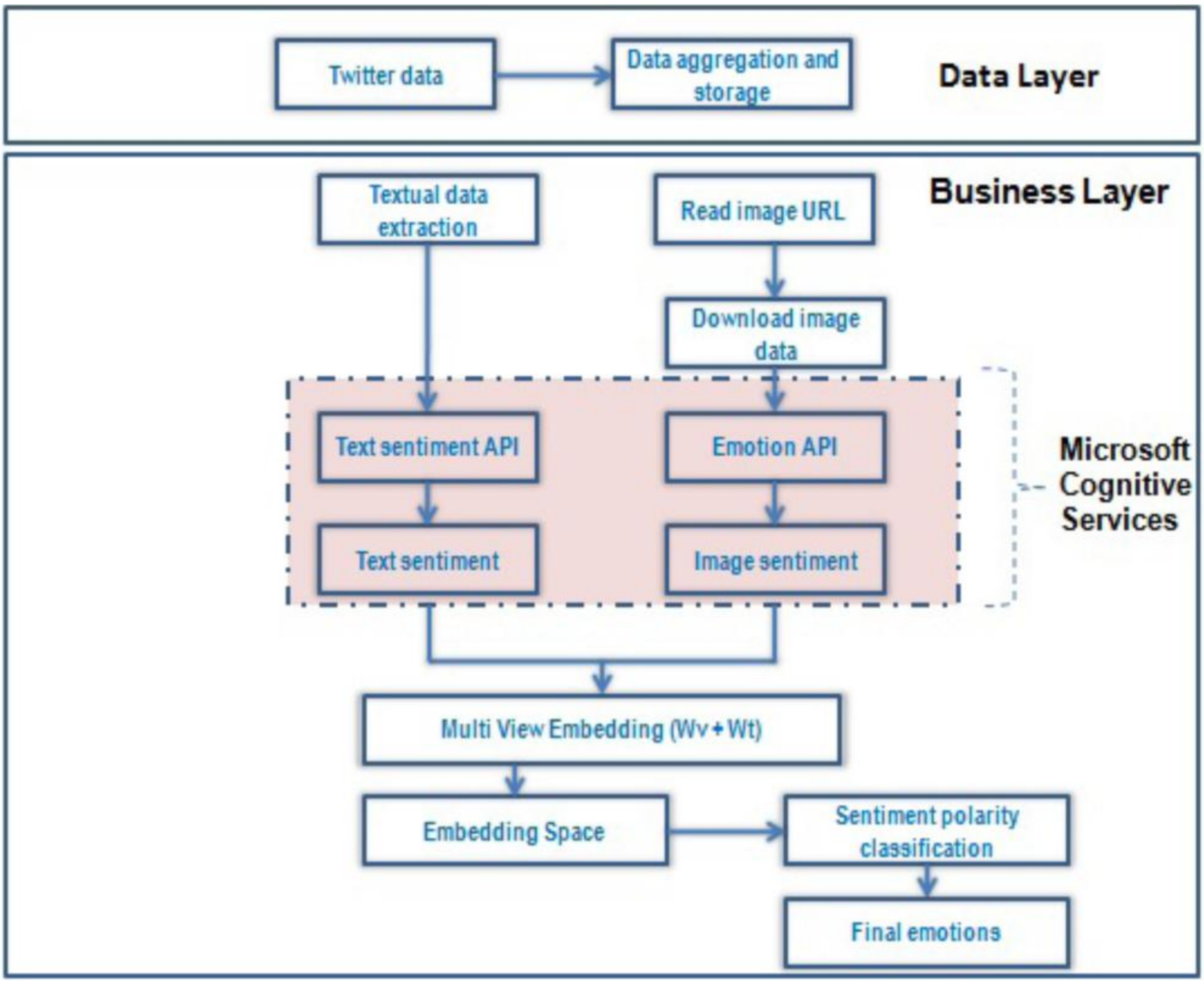


**Fig 1: System Architecture**

### A. Grab the real time tweets

**Step 1:** Getting Twitter API keys: We required four type of information from twitter API key, API secret, access token and access token secret, to access Twitter Streaming API.

**Step 2:** Connecting to Twitter Streaming API and downloading data: For download the data from Twitter we are connecting Twitter Streaming API using Python library called Tweepy and downloading the data.

**Step 3:** Reading and Understanding the data: The downloaded data should be readable and easy to understand so we stored the data in JavaScript Object Notation (JSON) format. These are easy to read and understand the data by humans and easy to read for machines to parse it.

### B. Storage of Twitter Stream

MongoDB is an open source document database that provides high performance, high availability, and automatic scaling. A record in MongoDB is a document, which is a data structure composed of field and value pairs. MongoDB documents are similar to JSON objects. The values of fields may include other documents, arrays, and arrays of documents. The twitter stream is JSON document, so it can be stored and retrieved efficiently using MongoDB database. To store the twitter data stream to MongoDB, we are using python library named pymongo.

### C. Textual Opinion Mining of tweets

While the twitter stream is being stored to MongoDB, we can mine textual sentiments periodically out of the stream data [2]. To implement such textual sentiment analysis we are using well known analytical tool named R Programming. R allows us to employ various text mining approaches, correlation analysis and sentiment analysis or opinion mining. In order to apply sentiment analysis on twitter data stream, stirred into MongoDB, we are using R-.NET connectivity. R implements sentiments analysis.

### D. Latent Correlation Analysis

In this section, we present latent correlations for image sentiment analysis of multi- views (Textual view and Image view). Latent Correlation means correlation among hidden features that are not observed, but inferred. From each view we extracting features then learning multi-view embedding space and to train image sentiment polarity classifier used latent embedding space.

**Textual features:** With the help of text associated to image textual features are created and from the training dataset creating a vocabulary using bag-of-words approach, which are helpful to count how much time the word occurs in text associated to the image. For counting the number of words shared between two images we use linear kernel for textual features [11, 12]. To train the system the word and word count are used. To reduce the dimensions of the textual feature, we used SVD for large and sparse matrices. Final textual representation set to the dimension of 1, 500 experimentally.

**Visual features:** Recently visual classification methods following feature design used [12, 15, 16]. By using combination of various visual descriptors we represent image appearance. From RGB color channels extract 3256 dimensional histogram, GIST descriptor of 512 dimensional, 1000 word dictionary with a 2-layer spatial pyramid and max pooling using by Bag-of-Words quantized descriptor. Following mid-level features also extracted: attribute features of 2000 dimensional [18]. To approximate the kernel we are using Fourier feature mapping for GIST, attribute, and SentiBank features [14].

Microsoft Cognitive API works on cloud where the training objects contains billions of images gives high accuracy. It is efficient to use and provide fast results. So by using Microsoft Emotion API we fetch emotion from image. To extract image sentiment, we first download the image present in the tweet url, then we pass it to Microsoft cognitive API. The API returns eight emotions as Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness and Surprise with their polarities.

**Multi-Views Embedding:** The multi-views in concern multiple views of one common keyword among the social media. Multi-view are multiple sentiments for a popular keyword like “GST”. As there are multiple sources of information, the sentiments must have multiple views. We broke down those views into textual view and image view.

Textual View: Sentiments from text data.

Image View: Sentiments from image data.

We employ the rule base which is nothing but multi-view embedding (Textual + Image) and extracting final emotions.

**E. Visualization of sentiments**

Grouping of both textual and visual approaches and finalizing the sentiments and visualizing those using Google charts and ASP.NET.

## 5. IMPLEMENTATION

The proposed system work as follows.

**Part 1:** Data Collection and Aggregation

1. Get twitter streams using Tweepy.
2. Convert the real time streams into JSON data.
3. Store JSON data into MongoDB.

**Part 2:** Data Extraction and Sentiment Analysis

A) Textual Data Extraction

1. Apply aggregation function on MongoDB and extract textual data of interest.
2. Remove stop words.
3. Apply Microsoft textual analytics API to extract textual sentiments.

B) Image Data Extraction

1. Apply aggregation function on MongoDB and extract image URLs
2. Download image from web resources.
3. Apply Microsoft emotion API to extract image sentiments.

**Part 3:** Latent Correlation analysis.

**Part 4:** Sentiments Polarity Classification

1. Defining polarity to the sentiments through sentiments polarity classification.

**Part 5:** Finalizing Sentiment

1. Maximum polarity of sentiment specifies the final sentiment of input image.

## 6. EXPERIMENTAL RESULTS

Creating database and collection dynamically providing database and collection name and then providing tweet keyword and tweet count I importing live tweets of corresponding keyword and finally stored in database. Table I shows of Dataset of different Tweet Keyword and corresponding sentiments.

First download dataset from twitter for evaluation performance of social media sentiment classification. We used live twitter dataset for project. Extracting image data by apply aggregation function on MongoDB and extract image URLs and using extracted images, finding image sentiments by using Microsoft Cognitive API, and textual sentiments of tweets by using R as positive, negative or neutral. By Latent correlation analysis We employ the rule-base which is nothing but multi-view embedding and defining polarity through the polarity classification and extracting final emotions of tweets. Table II shows metrics of textual classification and text plus image classification. The Comparison analysis of textual and image plus textual is shown in table III.

Fig. 2. shows comparison graph of comparison analysis of textual and image plus textual. From graph it clears that image plus textual sentiment is better than textual and image sentiments.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

**TABLE I: TEST DATASET**

| Sr. No | Tweet Keyword | Sentiment from Text | Sentiment from Image | Final Sentiment | Actual Sentiment | Textual Classification | Textual + Image Classification |
|---|---|---|---|---|---|---|---|
| 1 | Mangalyaan | Positive | Neutral | Positive | Positive | TP | TP |
| 2 | Barak Obama | Positive | Positive | Positive | Positive | TP | TP |
| 3 | Amarnath | Neutral | Negative | Negative | Negative | FP | TN |
| 4 | Sachin Tendulkar | Neutral | Positive | Positive | Positive | FN | TP |
| 5 | Anna Hazare | Positive | Positive | Positive | Positive | TP | TP |
| 6 | Zika | Negative | N/A | Negative | Negative | TN | TN |
| 7 | MH 370 | Negative | Negative | Negative | Negative | TN | TN |
| 8 | Earth Climate | Positive | N/A | Positive | Positive | TP | TP |
| 9 | Tsunami | Positive | N/A | Positive | Positive | TP | TP |
| 10 | Narendra Modi | Neutral | Positive | Positive | Positive | FN | TP |
| 11 | Terrorism | Positive | N/A | Positive | Negative | FP | FP |
| 12 | Facebook | Positive | N/A | Positive | Positive | TP | TP |
| 13 | Data Analytics | Positive | N/A | Positive | Positive | TP | TP |
| 14 | Computer science | Positive | N/A | Positive | Positive | TP | TP |
| 15 | Donald Trump | Neutral | Positive | Positive | Positive | FN | TP |
| 16 | GST | Neutral | N/A | Neutral | Positive | FN | FN |
| 17 | Bhopal GAS Tragedy | Neutral | Negative | Negative | Negative | FP | TN |
| 18 | Hiroshima Nagasaki | Neutral | Negative | Negative | Negative | FP | TN |
| 19 | Machine Learning | Positive | N/A | Positive | Positive | TP | TP |
| 20 | Swine Flue | Negative | N/A | Negative | Negative | TN | TN |
| 21 | Silicon Valley | Positive | N/A | Positive | Positive | TP | TP |
| 22 | Hollywood | Positive | N/A | Positive | Positive | TP | TP |
| 23 | Indian Institute of Technology | Positive | N/A | Positive | Positive | TP | TP |
| 24 | Indian Rail | Negative | N/A | Negative | Negative | TN | TN |
| 25 | Indian defense | Positive | N/A | Positive | Positive | TP | TP |
| 26 | Agriculture Ministry | Negative | N/A | Negative | Negative | TN | TN |
| 27 | SWAT | Positive | N/A | Positive | Positive | TP | TP |
| 28 | Air Pollution | Negative | N/A | Negative | Negative | TN | TN |
| 29 | Cancer | Neutral | N/A | Neutral | Negative | FP | FP |
| 30 | Ocean Temperature | Positive | N/A | Positive | Positive | TP | TP |

**TABLE II: METRICS**

| Sr. No | Classification | Textual Classification | Text + Image Classification |
|---|---|---|---|
| 1 | Sum (TP) | 15 | 18 |
| 2 | Sum (TN) | 6 | 9 |
| 3 | Sum (FP) | 5 | 2 |
| 4 | Sum (FN) | 4 | 1 |

**TABLE III: COMPARISON**

| Metric Classifier | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| **Textual** | 0.7000 | 0.7500 | 0.7800 | 0.7647 |
| **Image + Textual** | 0.9000 | 0.9000 | 0.9473 | 0.9243 |

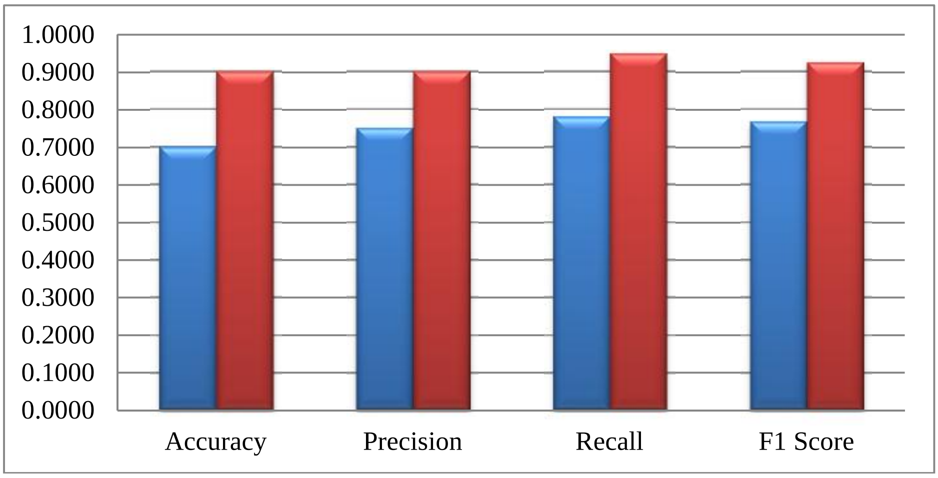


**Fig. 2: Comparative Analysis**

## 6. CONCLUSION

In this section results are based on preliminary work. Either user expresses their emotions using images or texts. It is difficult to predict the emotions of social network users. The results in this section are encouraging for using the multimedia information for sentiment analysis. To determine sentiment both the textual and visual features are important. Sentiment expressed by text and image we discover the correlation between them. We propose a correlation analysis that can exploit visual, textual sentiment of the input keyword related tweets. As compare to text sentiment and image sentiment analysis text and image sentiments are much better than them.

## 8. ACKNOWLEDGEMENT

With all respect and gratefulness, I would like to thanks all people who have helped me directly or indirectly for the paper presentation. I am grateful to my guide and PG Coordinator, Prof. V. N. Dhakane, for his guidance and support. I wish to express my sincere thanks to the HOD Prof. I. R. Shaikh for their support. Lastly I would like to thank staff member of Department of Computer, S.N.D. College of Engineering and Research Center, Nashik for making all the requirements possible and simply available whenever required. I thank the anonymous reviewers for their comments.

## REFERENCES

[1] Marie Katsurai and Shinichi Satoh, Image Sentiment Analysis Using Latent Correlations among Visual Textual, And Sentiment Views 978-1-4799-9988-0/16 2016 IEEE.
[2] Liu B (2012), Sentiment analysis and opinion mining. Synthesis Lectures Hum Lang Technologies 5(1):1167.
[3] You Q, Luo J (2013), Towards social imagematics: sentiment analysis in social multimedia In: Proceedings of the thirteenth international workshop on multimedia data mining, MDMKDD13. ACM, New York, pp 3:13:8.
[4] Nasukawa T Yi J (2003), Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the 2nd international conference on knowledge capture. ACM, pp 7077
[5] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intelligent Systems, vol. 28, no. 2, pp. 1521, 2013.
[6] S. Siersdorfer, E. Minack, F. Deng, and J. Hare, Analyzing and predicting sentiment of images on the social web, in Proc. Int. Conf. Multimedia (MM), 2010, pp. 715718.
[7] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, Largescale visual sentiment ontology and detectors using adjective noun pairs, in Proc. Int. Conf. Multimedia (MM), 2013, pp. 223232.
[8] J. Yuan, S. Mcdonough, Q. You, and J. Luo, Sentribute: Image sentiment analysis from a mid-level perspective, in Proc. Int. Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), 2013, pp. 10:110:8.
[9] Y. Wang, S. Wang, J. Tang, H. Liu, and B. Li, Unsupervised sentiment analysis for social media images, in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), 2015.
[10] A. Esuli and F. Sebastiani, SentiWordNet: A publicly available lexical resource for opinion mining, in Proc. Int. Conf. Language Resources and Evaluation (LREC), 2006, pp. 417422.
[11] M. Guillaumin, J. Verbeek, and C. Schmid, Multimodal semisupervised learning for image classification, in Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR), June 2010, pp. 902909.
[12] Y. Gong, Q. Ke, M. Isard, and S. Lazebnik, A multi-view embedding space for modeling internet images, tags, and their semantics, International Journal of Computer Vision, vol. 106, no. 2, pp. 210233, 2014.
[13] D Hardoon, S Szedmak, and J Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, Neural Computation, vol. 16, no. 12, pp. 2639 2664, Dec 2004.
[14] A. Rahimi and B. Recht, Random features for large-scale kernel machines, in Proc. Neural Information Processing Systems (NIPS), 2007.
[15] F. Perronnin, J. Sanchez, and Y. Liu, Large-scale image categorization with explicit data embedding, in Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR), June 2010, pp. 22972304.
[16] Y. Gong, L. Wang, M. Hodosh, J. Hockenmaier, and S. Lazebnik, Improving image-sentence embeddings using large weakly annotated photo collections, in Computer Vision ECCV 2014, vol. 8692 of Lecture Notes in Computer Science, pp. 529545. Springer International Publishing, 2014.