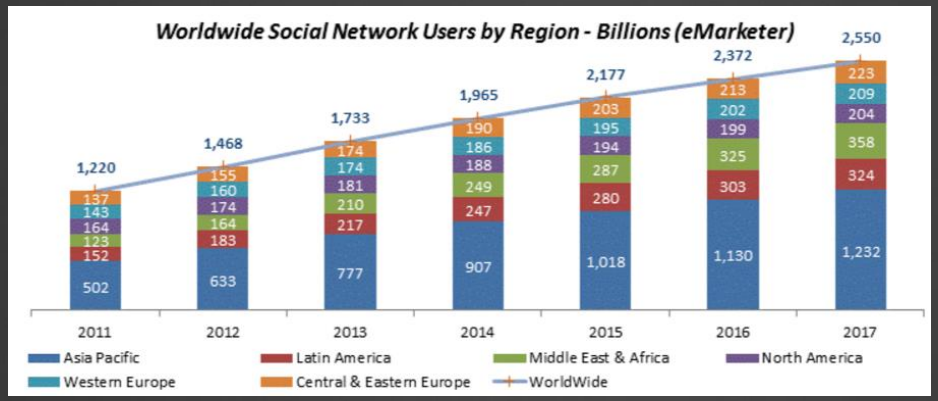
Multilingual Sentiment Analysis in Social Media Based on Image and Text Data

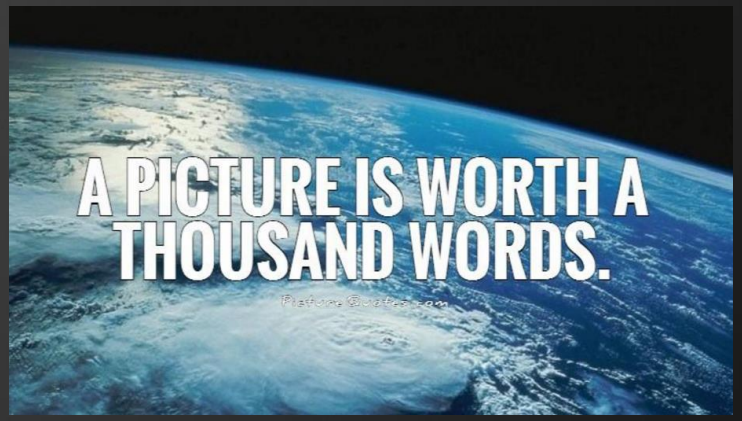
Hongyi Liu
Department of Electrical Engineering
May. 5th, 2016

1. Introduction

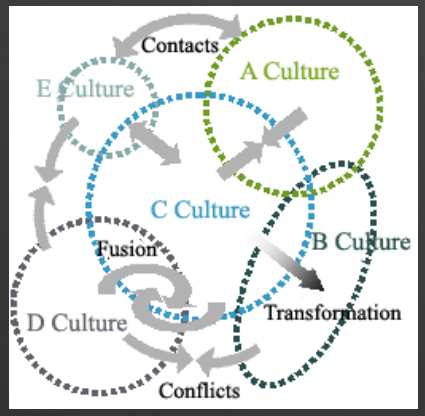
1. Social Media has become a significant part of our daily life



2. Visual Information is also important



3. Cultural Blending Phenomenon



2.Related Work

1. Text-Based Sentiment Analysis

Document Level, Sentence Level, Attribute Level

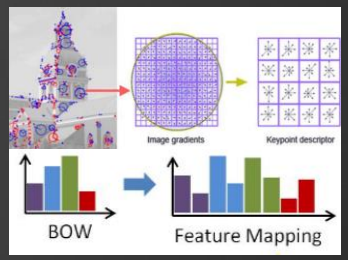
Naïve Bayes, KNN, SVM…

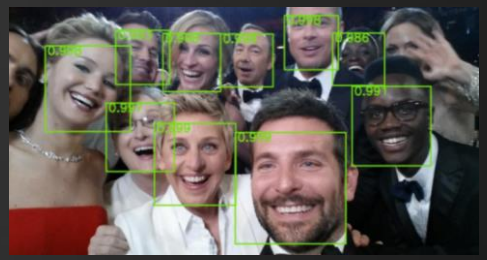
I was no older than three when my father put that first baseball in my hands. But it was later that I started to play football (for which I made a career from).

Keyword	Sentiment
Excellent	10
Impressed	6
Great	5
OK	1
Meh	0
Boring	-3
Sick	-4
Terrible	-5

2. Image-Based Sentiment Analysis

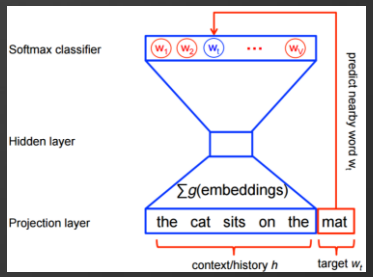
Low-Level Features, Face Detection …

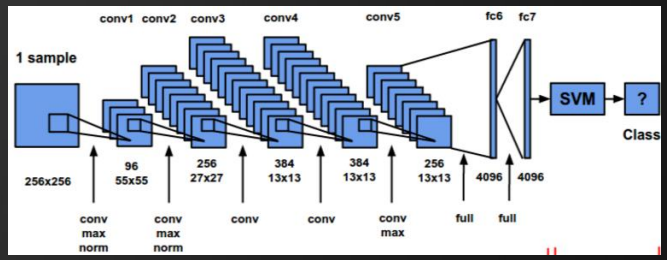




3. Deep Learning Approaches

Word2Vec, RNN, LSTM
CNN …

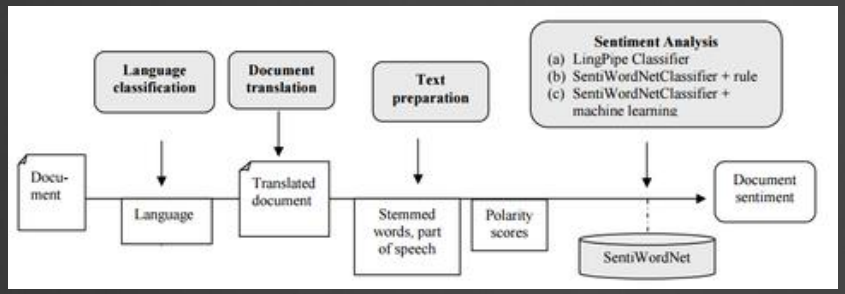




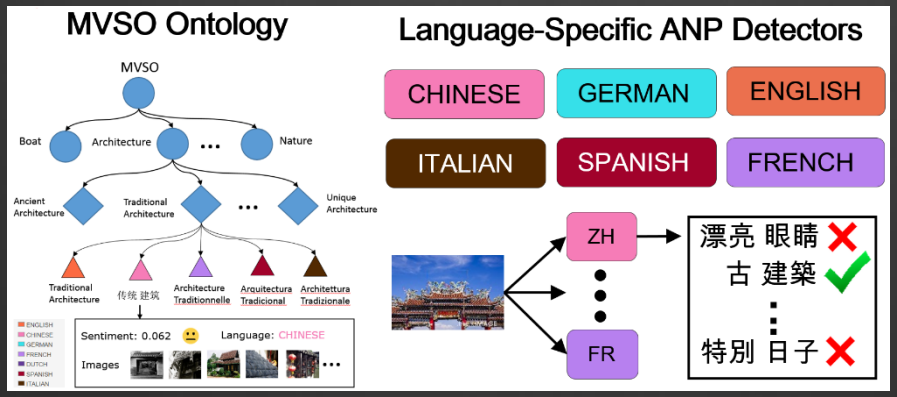
2.Related Work

Multilingual Sentiment Analysis

1. Text-Based Analysis **SentiWordNet[1]**



2. Image-Based Analysis **MVSO[2]**



[1] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." *Proceedings of LREC*. Vol. 6. 2006.

[2] B. Jou*, T. Chen*, N. Pappas*, M. Redi*, M. Topkara* and S.-F. Chang. Visual Affect Around the World: A Large-scale multilingual visual sentiment ontology. ACM MM , 2015.

3.Project Proposal

In this project, we aim to answer the **QUESTIONs**:

1. Will the social media content, for example, a tweet, be viewed coherently or differently across different languages & cultures?
2. If different, what may cause these sentiment discrepancies?
3. And how can we utilize these multilingual observations?

We proposal to complete the following **TODOs:**

1. Construct a multilingual tweets dataset containing texts, images and manually-labelled ground truth.
2. Use different methods to perform sentiment analysis over the dataset.
3. Hopefully find some interesting results that reflect the cross-cultural interconnection in a data-driven way.

4.Methods

4.1 Dataset Construction

1. Crawl tweets data of different languages containing images.
2. Using Amazon Mechanical Turk
   Text-based labeling.
   Image-based labeling.
   Joint text-image based labeling.

The textual data will be translated into differently languages for labeling in order to obtain the multilingual responses for each image tweet.

4.Methods

4.2 Multilingual Sentiment Analysis

1. Text-Based Prediction

Open APIs: SentiWordNet, SentiStrength

2. Image-Based Prediction

Multilingual image detectors from MVSO

3. Joint Image-Text Based Prediction

- Late Fusion of the text-based and image-based prediction scores.
- Embed the image and text into one feature space,
  and then perform sentiment classification based on the fused feature

4.3 Performance Evaluation

- Training / Testing Time, Accuracy …
- Visualization
- Analyze the polarized cases

5.Future Directions

1. Release the collected dataset.

2. Apply the trained models and dataset in real scenarios, such as recommendations, topic trends discovery and search query expansion based on sentiment analysis.

3. Automatically detect users whose tweets are often negative, and perhaps provide kind reminders and tips to help them become optimistic if necessary.

Thank You!