

MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets

Jichang Zhao

Beihang University
jichang.zhao@gmail.com

Li Dong

Beihang University
donglixp@gmail.com

Junjie Wu

Beihang University
wujj@buaa.edu.cn

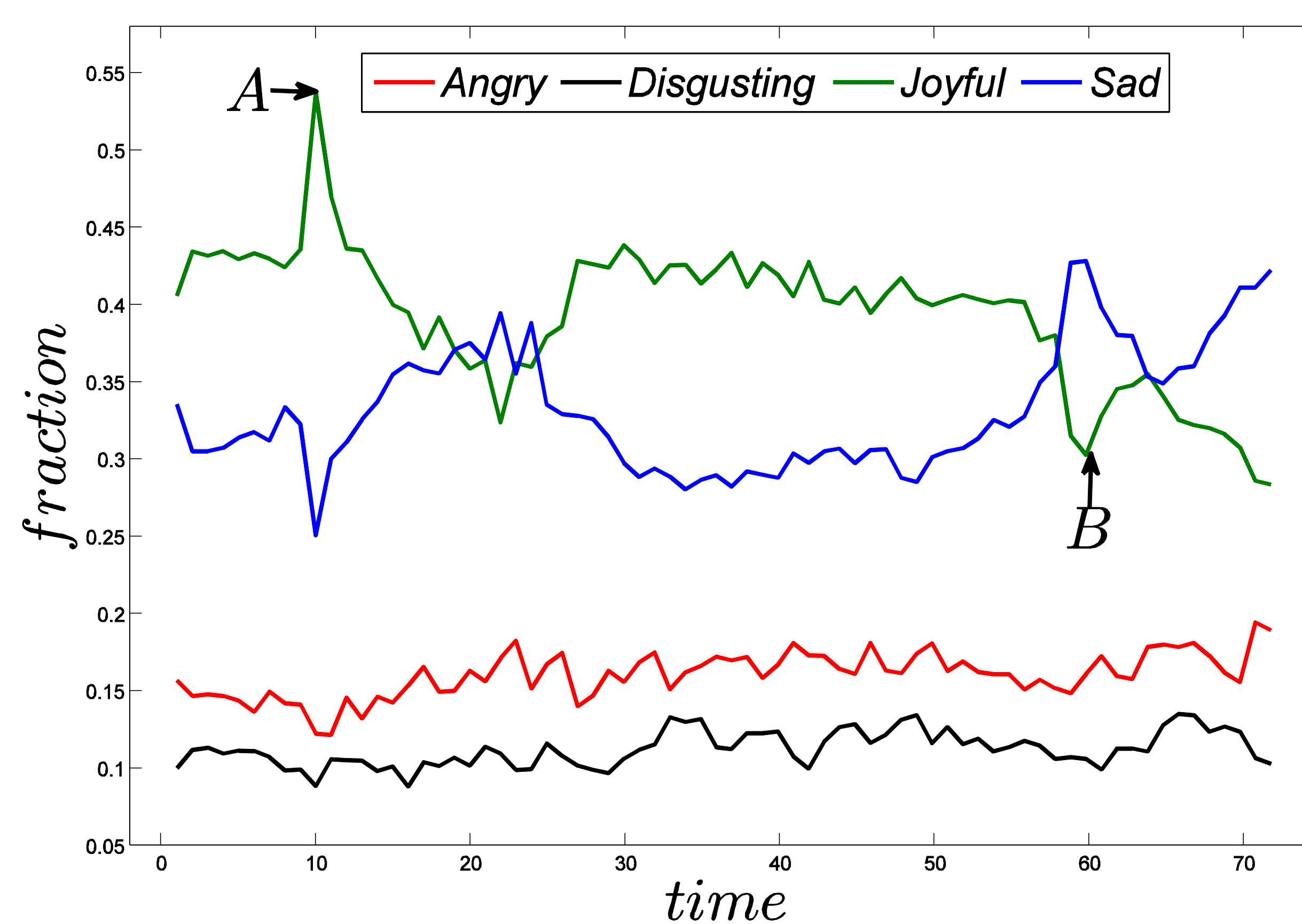
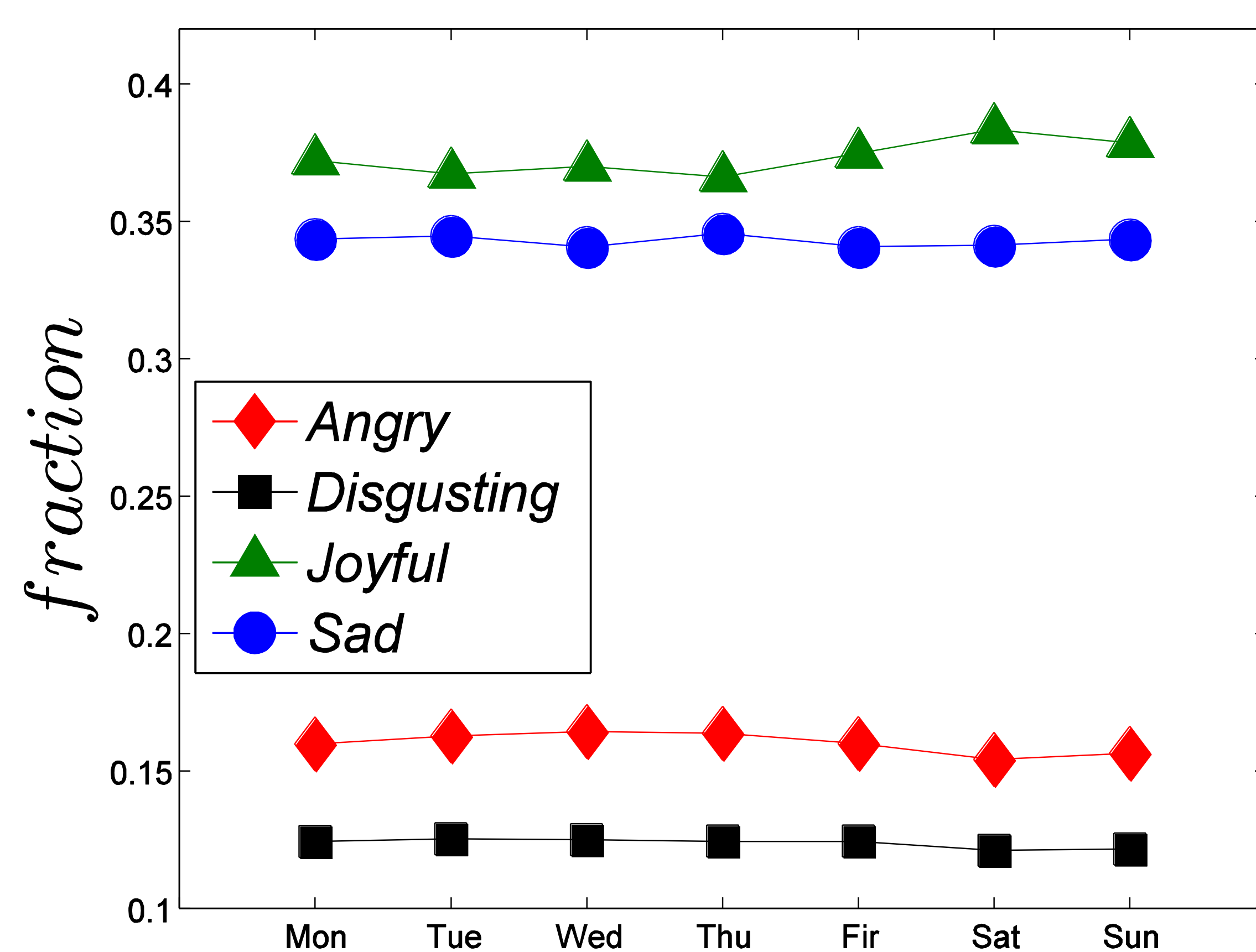
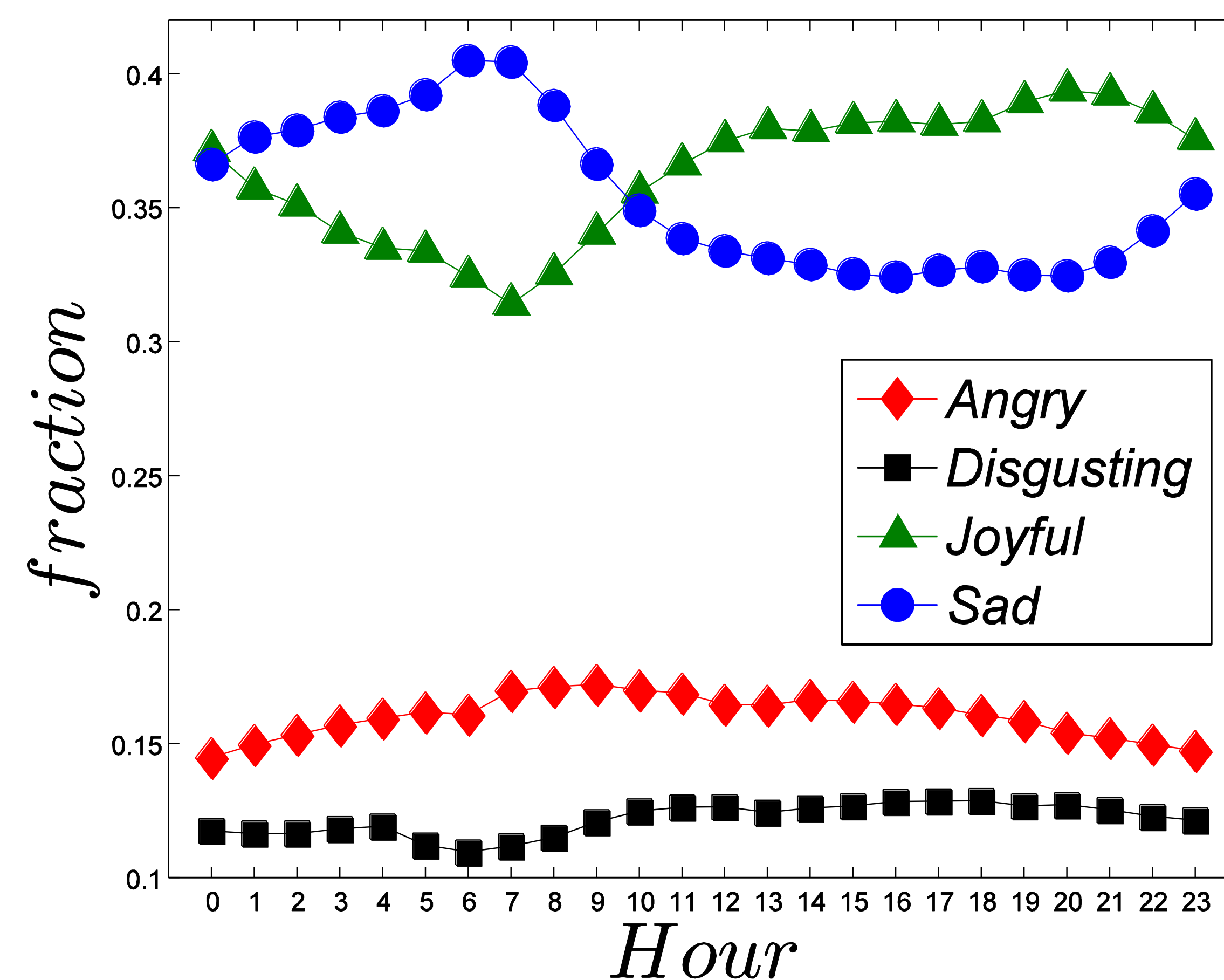
Ke Xu




Beihang University
kexu@nlse.buaa.edu.cn

Abstract

Recent years have witnessed the explosive growth of online social media. Weibo, a Twitter-like online social network in China, has attracted more than 300 million users in less than three years, with more than 1000 tweets generated in every second. These tweets not only convey the factual information, but also reflect the emotional states of the authors, which are very important for understanding user behaviors. However, a tweet in Weibo is extremely short and the words it contains evolve extraordinarily fast. Moreover, the Chinese corpus of sentiments is still very small, which prevents the conventional keyword-based methods from being used.

In light of this, we build a system called *MoodLens*, which to our best knowledge is the first system for sentiment analysis of Chinese tweets in Weibo. In *MoodLens*, 95 emoticons are mapped into four categories of sentiments, i.e. *angry*, *disgusting*, *joyful*, and *sad*, which serve as the class labels of tweets. We then collect over 3.5 million labeled tweets as the corpus and train a fast Naive Bayes classifier, with an empirical precision of 64.3%. *MoodLens* also implements an incremental learning method to tackle the problem of the sentiment shift and the generation of new words. Using *MoodLens* for real-time tweets obtained from Weibo, several interesting temporal and spatial patterns are observed. Also, sentiment variations are well-captured by *MoodLens* to effectively detect abnormal events in China. Finally, by using the highly efficient Naive Bayes classifier, *MoodLens* is capable of online real-time sentiment monitoring. The demo of *MoodLens* can be found at <http://goo.gl/8DQ65>.



Sentiment	#Emoticons	Typical emoticons
Angry	9	
Disgusting	14	
Joyful	50	
Sad	22	