
Understanding Popularity of Social Media Entities: From Hashtags to Question Topics

Suman Kalyan Maity

Dept. of CSE

IIT Kharagpur

sumankalyan.maity@cse.iitkgp.ernet.in

Abstract

With the evolution of social media over the years, entities like hashtags, tags, topics have also evolved and adopted traits that are very similar to various natural language units. One of the interesting and prevalent linguistic phenomena is compounding. Like natural language compounding, hashtag compounding also takes place and we study the factors behind their adoption in social media. Similar compounding phenomena can also be seen in Quora topics and other non-English languages used in social media.

Author Keywords

hashtag popularity; hashtag compounding; topic popularity

ACM Classification Keywords

H.4.m [Information Systems Applications]: Miscellaneous;
J.4 [Computer Applications]: [Social and Behavioral Sciences]

Introduction

Social media creates massive amount of user-generated content. To organize such a huge volume of content, various entities have been created - for example, hashtags in Twitter, Facebook, Instagram, Google+ etc, topics in Quora, tags in Stack Overflow etc. Over the years, these entities have evolved and adopted various traits that are specific to natural language units. We try to focus on the popularity aspects of such linguistic entities in various contexts.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).

CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, USA

ACM 978-1-4503-4688-7/17/02.

<http://dx.doi.org/10.1145/3022198.3024934>

Twitter hastags are widely used to define a shared context for specific events, topics etc. Apart from hashtags concerning well-known topics like sports, music, technology etc., there are a set of hashtags that are primarily used for conversational and personal reasons. These type of conversational and personal themed hashtags are generally called Twitter idioms (#10ThingsAboutMe, #4WordsAfterABreakup, #ThingsMyMamaDo, #ICantForgetAboutYou etc.). Twitter idioms are usually used by people for day-to-day gossip, showing personal feelings etc. Moreover, people join these topics when they see other people post them; in principle, they are motivated by sources that are external rather than internal (from their mind) [3]. Therefore, if one needs to understand human conversation dynamics, one has to study the idioms in isolation. This could open the gateway to understand the opinion [1], sentiments and emotions of people in more focused way instead of looking into the entire gamut of Twitter hashtags which researchers have been doing over the years. Hashtags have been studied extensively by researchers to analyze the competition dynamics, the adoption rate and popularity scores [10, 5, 4, 2]. However, there are very few attempts to study the linguistic aspects of hashtag evolution over large time scales.

One of the interesting and prevalent linguistic phenomena in today's world of brief expressions, chats etc. is hashtag compounding where new hashtags are formed through combination of two or more hashtags together with the form of the individual hashtags remaining intact. For example, #PeoplesChoice and #Awards form #PeoplesChoiceAwards. #KellyRipa and #CelebrationMonth make #KellyRipaCelebrationMonth; #WikipediaBlackout is formed from #Wikipedia and #Blackout; #OregonBelieveMovieMeetup is formed from #Oregon, #BelieveMovie and #Meetup; #Educational, #Ipad, #Apps together make #EducationallpadApps etc.

Another class of social media sites correspond to Q&A forums like Stack Overflow, Quora etc. Apart from being a Q&A site, Quora has a social network backbone that nicely integrates its user base and is a very unique feature compared to the other Q&A sites. People can tag questions with various topics, follow a question, follow a topic, share questions and its answers apart from the basic features like upvoting/downvoting, commenting etc. In Quora's ecosystem of knowledge sharing through question-answering, topics play an important role. People follow topics to get important and valuable content related to a topic of their interest. Similarly, when a user posts a question, he/she can tag it with relevant topics so that the topical experts and people interested in the topics get to know about the question and can provide better answers thus helping to control the content quality in Quora. Further, the users in Quora usually provide compelling answers to the questions in which they are interested. Therefore, topics form an essential organizing tool for Quora's knowledge corpus.

Research overview and Work done

A stratified learning approach for predicting the popularity of Twitter Idioms

One of the first steps to understand the rich data of conversation dynamics in social media is to know what are the popular idioms, the factors that make these kind of hashtags popular. [9] observed that different topical categories (sports, music, idioms) of hashtags have different propagation patterns. Therefore, there is a natural intuition that they would gain popularity through different mechanisms as the underlying spreading pattern is different. These observations formed the central motivation of this work where we investigate the detailed mechanics of the spread of the idioms and, subsequently, develop a model that can automatically predict those idioms that are going to be popular in future. We also perform predictions at different time points.

Benefits of participation to the Doctoral Colloquium

I believe that attending the CSCW 2017 Doctoral Colloquium can benefit my research in various ways. First and foremost, it will provide me an unique opportunity to present and discuss my research work and receive constructive comments from my fellow researchers and professionals from the various backgrounds. Since I plan to finish my PhD works within a couple of years, I think those feedbacks and advices would certainly better shape my PhD work and help me achieve my desired goal. Secondly, I believe that my participation into the Colloquium is not unilateral. I expect to share my views, thoughts on my fellow students' PhD thesis works which might benefit them as well. I attended the CSCW 2016 conference last year where I could mingle and interact with the fellow researchers, eminent academicians and professionals which was certainly satisfying and enriched me a PhD Student. Lastly, I see the colloquium as a great platform to socialize and to be acquainted with the CSCW community better.

Prediction of popular idioms at different time points can be directly applied to identify the temporal scope of sentiments/opinions in the associated tweets and the way the scope changes. This also helps us understand the community sentiment as people usually follow the trending idioms. At first, we categorize the set of hashtags into two different classes: idioms and non-idioms. We then contrive a set of specialized popularity features on the stratified sample to independently predict the popularity of Twitter idioms. **This work was had been presented at ICWSM 2015 [6].**

Early Prediction of Popular Hashtag Compounds

Like general hashtags, predicting popular hashtag compounding is also an important and interesting task. There are marketing strategic needs, needs for fulfilling communicative intents (affective expression, political persuasion, humor etc.) as well as spontaneous needs for use of hashtag compounds. For example, the e-commerce company Amazon used #AmazonPrimeDay to promote the discounted sale of its product. The hashtag is a compound of #Amazon and #PrimeDay whereas the individual hashtag #PrimeDay was also popular. So, there is a trade-off whether to use hashtag compounds or the uncompounded constituents. Similarly, assume another scenario where an event is taking place, say the premiere of a movie 'The Imitation Game'. Here one can use both the hashtags #TheImitationGame and #Premiere or can use a hashtag compound #TheImitationGamePremiere. In this context, one needs to identify which version one should use so that the hashtag being used gains a higher frequency of usage in the near future. Hashtag compounds also serve the communicative intents like political campaign hashtags (#PresidentTrump = #President + #Trump : hashtag to support Donald Trump for the 2016 US Presidential election). Hashtag compounding can also happen spontaneously. These hashtags are generally conversational or personal

themed hashtags like #TheBestFeelingInARelationship (#TheBestFeeling + #InARelationship), #ThrowbackThursday (#Throwback + #Thursday), #ComeOnNowDontLie (#ComeOnNow + #DontLie). Our prediction framework is different from existing popularity prediction/trend identification algorithms/frameworks in the following ways. Most popularity prediction frameworks deal with the problem of predicting whether a hashtag will become popular or not among a competing hashtag pool consisting of all hashtags across various topics from the data stream and filtered by the time window in which the prediction is being made. However, in our framework, we attempt to predict whether the hashtag compound or the individual constituent hashtags would become popular. Therefore, our competition space is smaller and topically more well-defined. In other words, the predictor attempts to indicate whether users would adopt the compounded hashtags or not. **This work was presented at CSCW 2016 and was selected as an honorable mention paper [7].**

Analysis and prediction of question topic popularity in Quora

In this work, we have moved from hashtags in Twitter to topics in Q&A sites like Quora. We study the dynamics of topic growth and evolution in Quora over time; in other words, we focus on how the Quora knowledge base is changing over time with the influx of new topics, growth or decay of older topics. One of the primary interests of this study is to identify factors that have a direct impact on the growth of popularity of the question topics. Understanding the popularity of topics is important because it helps us identifying trending topics. This study has a direct application in recommendation of the trending topics to various users in Quora. To the best of our knowledge, this is the first rigorous and in-depth measurement study on a large dataset spanning over a period of more than four years that focuses on the prediction of popular question topics and can potentially have a significant impact on a service like trending topic recom-

Challenges in Indian Language Context

Indian-ness: These posts are usually in standard English but are relevant in the Indian context only.

When people in #flood-ravaged #Uttarakhand went hungry during 2013 disaster, state govt officers relished mutton, gulab jamuns #shame

Code-mixed/code switched tweets:

Quite often the Indian language words are embedded in English posts and vice versa. *Huge traffic restrictions for PM's visit to #blast site mean deserted roads in #Hyderabad. "Itna sanaata kyon hai bhai?" (code-switch)*

MMS will go to #Hyderabad-Blast site to take jayeja of area & say Hazaaron Jawabon Se Acchi Hai Meri Khamoshi #ThikHai (code-mix)

Indian language and script: Post may also be completely in an Indian language and often typed in devanagari unicode.

mentation in Quora. **This work was presented at ICWSM 2015 [8].**

Future direction and Expected Contribution

Quora topic merging

Like hashtag compounding which is a prevalent form of compositionality in social media, we also observe compounding of Quora topics. In Quora, topics are organized via parent-child relationships thus shaping up an unique ontological structure. Preliminary investigation suggests existence of topic merges, where one topic gets merged into another topic. For example, 'Paper Money' merged into 'Banknotes', 'Chinese Chess' into 'Xiangqi game' and so on. We here try to figure out how such a phenomena happens. The most important question is whether such merges are outcomes of linguistic or social pressures.

Indian language compounding

Though English language is a prevalent medium of tweeting, a significant number of tweets are posted in other languages like Japanese, Spanish, Portugese, Hindi etc. We are currently focusing on identifying and studying hashtag compounding for Indian languages. The task is difficult due to unavailability of state-of-the-art linguistic tools (named entity disambiguator, POS tagger etc.) The final goal of this project is to develop a end-to-end system for hashtag compounding and recommendation.

REFERENCES

1. Marco Bastos, Rodrigo Travitzki, and Cornelius Puschmann. 2012. What Sticks With Whom? Twitter Follower-Followee Networks and News Classification (*ICWSM '12*).
2. Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and Krishna P. Gummadi. 2010. Measuring user influence in Twitter: The million follower fallacy (*ICWSM'10*).

3. Anqi Cui, Min Zhang, Yiqun Liu, Shaoping Ma, and Kuo Zhang. 2012. Discover Breaking Events with Popular Hashtags in Twitter (*CIKM '12*). ACM, New York, NY, USA, 1794–1798.
4. Evandro Cunha, Gabriel Magno, Giovanni Comarela, Virgilio Almeida, Marcos André Gonçalves, and Fabrício Benevenuto. 2011. Analyzing the Dynamic Evolution of Hashtags on Twitter: A Language-based Approach (*LSM '11*). 58–65.
5. Zongyang Ma, Aixin Sun, and Gao Cong. 2013. On predicting the popularity of newly emerging hashtags in Twitter. *JASIST* (2013), 1399–1410.
6. Suman Kalyan Maity, Abhishek Gupta, Pawan Goyal, and Animesh Mukherjee. 2015. A Stratified Learning Approach for Predicting the Popularity of Twitter Idioms (*ICWSM '15*).
7. Suman Kalyan Maity, Ritvik Saraf, and Animesh Mukherjee. 2016. #Bieber + #Blast = #BieberBlast: Early Prediction of Popular Hashtag Compounds. In *Proc. of CSCW*.
8. Suman Kalyan Maity, Jot Sarup Singh Sahni, and Animesh Mukherjee. 2015. Analysis and Prediction of Question Topic Popularity in Community Q&A Sites: A Case Study of Quora (*ICWSM '15*).
9. Daniel M. Romero, Brendan Meeder, and Jon Kleinberg. 2011. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In *Proc. of WWW*. 695–704.
10. Jaewon Yang and Jure Leskovec. 2010. Modeling Information Diffusion in Implicit Networks (*ICDM '10*). 599–608.