Emoji 2021@ICWSM

Modeling Emoji Generation for Emotion Analysis of Social Media Short Texts

Sujatha Das Gollapalli, See-Kiong Ng

Institute of Data Science, NUS, Singapore

This talk

- Motivation
  - Why model emojis using generative models?
- Our proposed EmDMM for capturing emoji clusters
- Analysis of a covid-19 tweet corpus with EmDMM
  - Unsupervised emotion detection

Why Topic Models?

- Topic Mixture Models based on Latent Dirichlet Allocation [1]
  - Widely-used in Information Retrieval, Natural Language Processing research
  - content analysis, topic trends, temporal changes, author modeling, sentiment analysis
- Unsupervised tool for extracting topics for a given document collection with parameters estimated through probablistic sampling
- Underlying assumptions
  - Corpus/document collection: mixture of K topics
  - Each document: multinomial distribution over “latent” topics
  - Each topic : multinomial distribution over “observed” words
- “Explain” how a document/corpus was generated
  - Informally, write a document -> choose a topic -> choose words conditioned on that topic

Why Topic Models for Emojis ?

Emojis and emoticons widely-used on social media platforms for various purposes

Complementing messages: Add emphasis (you look hot! 🔥)

Condensing messages: 🤭

Expression emotions in messages: 🥰 🤣 😒

Recent studies on emojis: emoji2vec for similarity between emojis, emoji prediction and recommendation challenges, personality modeling, sentiment and polarity detection

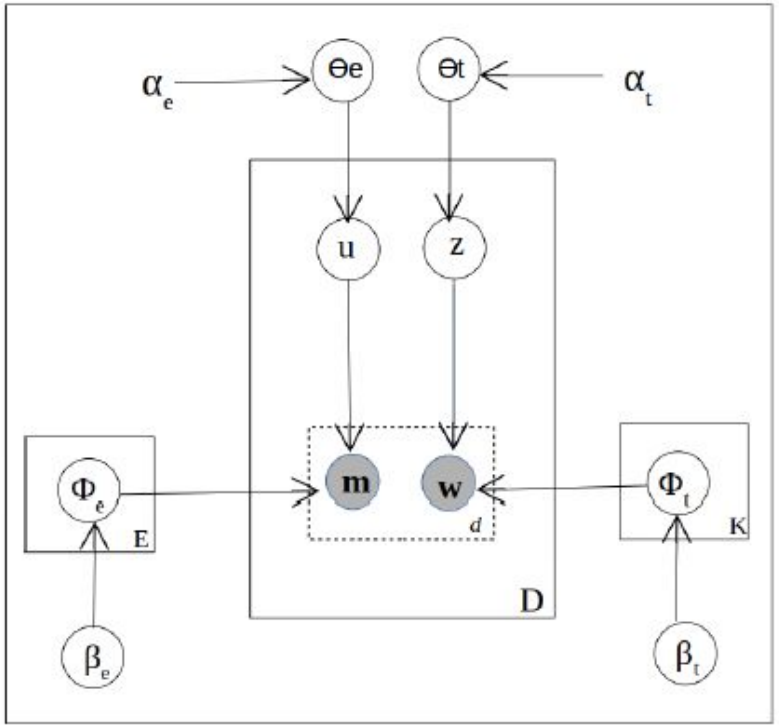
Question in this paper: Just as words are generated based on a “latent” topic in topic models,

Are emojis generated based on a “latent” emotion?

EmDMM: A Dirichlet Multinomial Mixture model for Emojis

- Usage of Emojis in microblogs (Social Media Short Texts)
- Extend DMMs [2] instead of LDA (Single topic for each “document”)
- Each tweet comprises of a
  - “latent” topic and a “latent” emotion and these variables affect what words and emojis are expressed the tweet

After modeling a corpus with EmDMM, we obtain automatic clustering along two dimensions



Experiments: Analyzing covid-19 tweets with EmDMM

- Characterize emotions expressed during the coronavirus pandemic with EmDMM
  - covid19_twitter dataset collected by Banda, et al [3]
  - Social Media Mining Toolkit[1] for processing the content
  - 333, 937 tweets with emojis over the period: March 01, 2020 to Sept 13, 2020
  - The Mallet toolkit from Umass[2] for implementing EmDMM
    - Tune for number of emoji and number of topic clusters and set to 12 and 60, respectively

1. https://github.com/thepanacealab/SMMT

2. http://mallet.cs.umass.edu/

Covid_tweets Corpus Composition

- Highly-skewed, long-tail: Top-3 topic/emoji clusters are assigned to almost 45%/41% of the corpus, respectively
- Bottom-30% topics only cover 1% of the corpus
- “positive” emotion assigned to about 26% and “negative” to about 18%

Topic	Top-Hashtags
45 (19%)	#coronavirussa #covididiots #coronapocalypse #riots #covidiot #coronavirusoutbreak
20 (14%)	#fakenews #dumptrump #ccp #theresistance #democrats #fauci #who #billgates
35 (11%)	#savetheworld #billionshields #quarantine #school #quarantinelife #backtoschool
42 (7%)	#africa #blog #leadership #innovation #technology #publichealth #healthcare
32 (6%)	#coronaviruschina #wuhan #covid– #thankyou #coronapocalypse

Emoji clusters extracted by EmDMM on the covid tweet corpus

Emoji clusters extracted by EmDMM on the covid tweet corpus

Joy, Trust, Anticipation, Surprise, Sadness, Fear, Anger, Disgust

Plutchik's basic emotions [5]

EID	Emojis	EID	Emojis
0		1	
2		4	
5		7	 BPB*
8		9	

EID	Emojis	Emotion Labels
3		Disgust
6	 d:	Joy
10		Sadness
11		Joy

Anecdotes comparing Emotion Detection models

TID	*EmDMM*	DepecheMood	ESTeR	Tweet
45	Joy	Happy	Anticipation	Come rain, heat, snow or coronavirus the bank's demands make it through my letter box. :)
45	Disgust	Afraid	Joy	Couldn't agree more. This is a big fuck you from Mother Nature herself 🙌🟫
45	Sadness	Afraid	Surprise	Shit..... Hope he gets well soon 🙏🙏
20	Joy	Inspired	Joy	That's Humanity ❤️#CoronaVirus #COVID
20	Disgust	Inspired	Neutral	Thanks coronavirus for ruining ALL of my plans 🙃
20	Sadness	Inspired	Joy	God!! Something more coming to us!! 💔💔💔
35	Joy	Annoyed	Joy	Share with the boys in your life! 🙌
35	Disgust	Inspired	Fear	🥺🥺🥺reasons i am not flying anywhere
35	Sadness	Happy	Disgust	New Coronavirus definition: Something that is utilized to take away everything that makes life bearable. 😒

EmDMM is compared with two state-of-the-art unsupervised Emotion Prediction models: DepecheMood [7] and ESTeR [6]

Takeaway: When the content is sparse or ambiguous, emojis can indeed provide a valuable signal for detecting the underlying emotion expressed in the tweet

Observations

- ~48% accuracy on a sample set of ~100 tweets
- Need a large-scale evaluation
  - What about other emotions/other corpora?
  - Joy, Trust, Anticipation, Surprise, Sadness, Fear, Anger, Disgust
- Simplistic assumption in EmDMM
  - Separation of text -> topical content
  - and emojis -> emotional content

"OMG corona 😱"

"Doctors and nurses in hospitals. #coronavirus 🌍🌍🌎"

Going Forward

- Distinguishing non-emotion and emotion emojis and their combinations for finer semantics

  ✅ ❤️ 👨‍⚕️🙏

- Sarcasm and Humour

  Yeah right! 😜😂 Yeah right!🙄😒

- User, context, and temporal modeling using topic models

References

1. Blei, et al. (2003) Latent Dirichlet Allocation. In JMLR. .
2. Yin and Wang (2014) A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering. In KDD.
3. Banda, et al (2020) A large-scale COVID-19 Twitter chatter dataset for open scientific research – an international collaboration.
4. Mohammad, et al (2018) SemEval-2018 Task 1: Affect in Tweets. In SemEval-2018.
5. Plutchik (2001) The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. In American Scientist.
6. Gollapalli, et al (2020) ESTeR: Combining Word Co-occurrences and Word Associations for Unsupervised Emotion Detection. In Findings of EMNLP.
7. Araque, et al (2019) Depechemood++: a bilingual emotion lexicon built through simple yet powerful techniques. In IEEE Transactions on Affective Computing.

Thank you! Questions? Email: idssdg@nus.edu.sg