Joint Topical Word Embedding for Detecting Drift in Social Media Text

VIJAYARANI J (viji.cs66@gmail.com)
CEG, Anna University, Chennai

Geetha T.V.
CEG, Anna University, Chennai

Research Article

Keywords: topic drift, hashtag, geotag, Langevin dynamics, word embedding, topic2vec

DOI: https://doi.org/10.21203/rs.3.rs-90835/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Social media texts like tweets and blogs are collaboratively created by human interaction. Fast change in trends leads to topic drift in the social media text. This drift is usually associated with words and hashtags. However, geotags play an important part in determining topic distribution with location context. Rate of change in the distribution of words, hashtags and geotags cannot be considered as uniform and must be handled accordingly. This paper builds a topic model that associates topic with a mixture of distributions of words, hashtags and geotags. Stochastic gradient Langevin dynamic model with varying mini-batch sizes is used to capture the changes due to the asynchronous distribution of words and tags. Topical word embedding with co-occurrence and location contexts are specified as hashtag context vector and geotag context vector respectively. These two vectors are jointly learned to yield topical word embedding vectors related to tags context. Topical word embeddings over time conditioned on hashtags and geotags predict, location-based topical variations effectively. When evaluated with Chennai and UK geolocated Twitter data, the proposed joint topical word embedding model enhanced by the social tags context, outperforms other methods.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the latest manuscript can be downloaded and accessed as a PDF.