

Temporal Word Embeddings for Dynamic User Profiling in Twitter



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

A scalable dynamic user profiling technique for Online Social Networks

Breandán Kerin, Annalina Caputo, Séamus Lawless

kerinb@tcd.ie, annalina.caputo@dcu.ie, seamus.lawless@scss.tcd.ie

Problem Statement

Individuals and their data are dynamic: interests and preferences evolve and vary through time, often following trends, periodicities and spikes.

Temporal variations can often be identified through a user's language usage over time, using NLP techniques like word embeddings.

Most modern word embedding techniques are not capable of modelling this temporal variation.

User profiling: the process of identifying the data about a user interest domain, which can be used to understand more about the user.

Applications: monetisation, targeted advertising, etc..

Word embeddings: representation of semantic properties of term usage in a vector space.

Applications: sentiment analysis, machine translation, information retrieval, etc.

The Landscape

- ✓ Lots of research exists for both word embedding techniques and user profiling
- × Limited research has been conducted into:
 - a) how temporal aspects of users can be captured using user profiling techniques
 - b) the use of word embedding techniques to capture temporal variances in language

Research Gap: Temporal word embeddings for user profiling

- Limited work to-date: A need recognised by previous research
- **Temporal Random Indexing:** Promising results for event detection in news articles (Basile *et al.*, 2016)
- **Temporal word embeddings for user profiling:** Potential demonstrated by temporal extension of the Word2Vec technique, for user profiling (Liang *et al.*, 2018)

Temporal Random Indexing for Dynamic User Profiling

Extension of the Random Indexing word embedding technique to account for temporal variation, such that the interests of users can be modelled based on their use of language over time.

TRI semantic index vector for time period T_k :

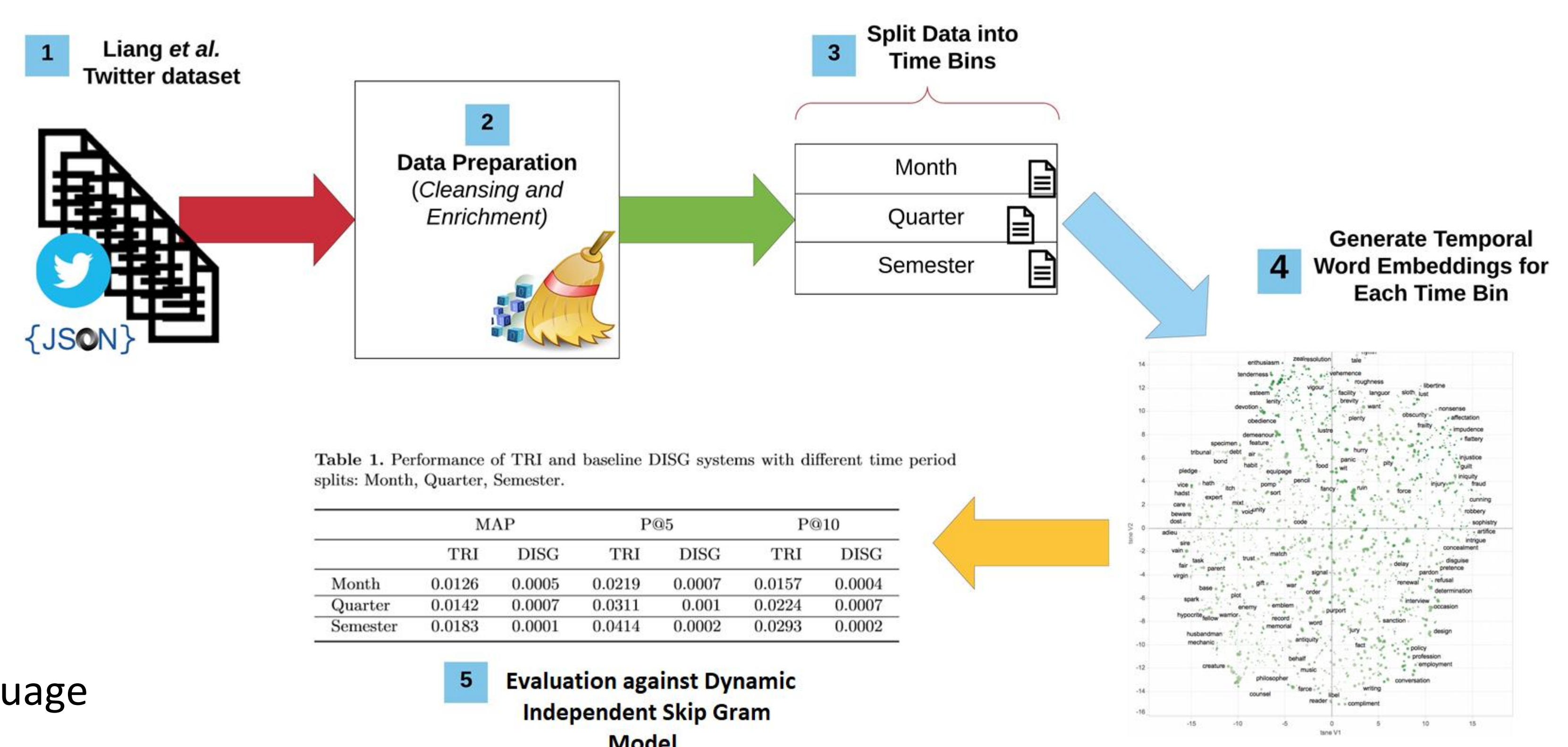
$$SV_{i,T_k} = \sum_{d \in C} \sum_{-j < p < +j} r_p$$

Allows for direct comparison between words within different time periods.

User vector for time period T_k :

$$UV_{i,T_k} = \sum_{p \in T_k} r_p$$

Captures a vectorised representation of a user based on their language usage over different time periods.



Results and Conclusions

A novel temporal user profiling technique, capable of dynamically generating temporal user profiles for short streams of text through the use of Temporal Random Indexing.

- ✓ Successfully modelled Twitter users across multiple granularities of time based on their use of language.
- ✓ Outperformed temporal Word2Vec (DISG model) in the generation of temporal user profiles for the same dataset.
- ✓ Highly performant (memory, time) and scalable enough to be considered for use in live web-scale environments.