Temporal Word Embeddings for Dynamic User Profiling in Twitter

Breandán Kerin
M.A.I. Computer Engineering

The research conducted under this title focused on exploring the domain of user profiling, a nascent technology which has been steadily attracting increased interest from the research community as its potential in the provision of personalised digital services is realised.

An extensive review of related literature revealed that there has been limited research conducted to-date into how temporal aspects of users can be captured using user profiling techniques. Coupled with the notable lack of research into the use of word embedding techniques to capture temporal variances in language, an opportunity was identified to extend the Random Indexing word embedding technique such that it could model the interests of users based on their use of language.

To achieve this, the work completed concerned itself with extending an existing implementation of Temporal Random Indexing to model Twitter users across multiple granularities of time. The product of this was a novel approach to producing a set of vectors describing the evolution of each Twitter user's interests over time through their use of language. These vectors were evaluated against another state-of-the-art word embedding technique, the Word2Vec Skip-gram model, where it was found that Temporal Random Indexing outperformed Word2Vec in the generation of temporal user profiles.

The major contribution of this research has been the development of a novel temporal user profiling system, capable of generating temporal user profiles for short streams of text through the use of Temporal Random Indexing. The fact that Twitter user data can be processed, enriched, and used to produce both user and word embeddings in a single application is a significant accomplishment of this research, and a meaningful contribution to the knowledge of the research community in the domain of user profiling.