Educational Research (ISSN: 2141-5161) Vol. 12 (4) Available online @http://www.interesjournals.org/ER



*Short Communication*

# Using Natural Language Processing to Understand Us

**Sabrina Tiun**
CAIT, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia

**Abstract**

The latest popular approach in natural language processing is the so-called word embedding (WE). The word embedding (WE) model is a neural network based on the distributional semantic model. The distributional hypothesis states that semantically similar words tend to have similar contextual distributions [1][3]. In the WE context, if two words have similar vectors, then they have the same distribution. An application of WE based on periodical is called temporal word embedding or dynamic word embedding (DWE).

We have explored the use of WE and DWE to mine lifestyle, sentiment and evolution of trends and policy. In the work of WE on Malaysian Twitter corpus, we explore the possibility of viewing Malaysia's lifestyle on where they spend most of their time for social meetings and analyse the sentiment of a public figure. The TWE model applications fall under two circumstances: For a TWE with a time-series corpus of a more extended period (or also known as a diachronic corpus), usually the related research concerns on language topics, i.e., the study on the meaning of words over time. A time-series corpus with a shorter period (called temporal corpus) is more on mining texts for culture semantic shift, i.e., detecting event for an actionable purpose [3]. In our work on DWE [3]

## Biography

Sabrina Tiun is a senior lecturer at the Faculty of Information Science and Technology (FTSM) in UniversitiKebangsaan Malaysia (UKM), Malaysia. Her research interests range from natural language processing, computational linguistic, and speech processing to social science computing. She teaches several courses in natural language processing and data science at both undergraduate and postgraduate levels.

**Recent Publication:**

- Sabrina Tiun &Yeoh Zi Liang, Pendekatan Word Embedding Dalam Kajian Sosio-Budaya Pada Korpus Sosial Media Melayu, Prosiding Simposium Kemanusiaan Digital 2019. Bangi Golf Resort, Selangor, 16 Oktober 2019
- Sabrina Tiun, Nor Fariza Mohd Nor, Azhar Jalaludin & Anis Nadiah Che Abdul Rahman. 2020. Word Embedding for Small and Domain-specific Malay Corpus. In: Alfred R., Lim Y., Haviluddin H., On C. (eds) Computational Science and Technology. Lecture Notes in Electrical Engineering, 603:435-443
- Kutuzov, A. Øvrelid, L., Szymanski, T. & Velldal, E. 2018. Diachronic Word Embeddings and Semantic Shifts: A Survey. Proceedings of the 27th International Conference on Computational Linguistics, pp. 1384-1397

**Cite this article** : Sabrina Tiun , CAIT, Faculty of Information Science and Technology, UniversitiKebangsaan Malaysia, Webinar on Artificial Intelligence and Robotics, April 29, 2021,London, UK