*Data article*

**Title:** Arabic Sentiment Embeddings

**Authors:** Nora Al-Twairesh, Hadeel Al-Negheimish

**Affiliations:** College of Computer and Information Sciences, King Saud University

**Contact email:** twairesh@ksu.edu.sa

**Abstract**

*Includes sentiment-specific distributed word representations that have been trained on 10M Arabic tweets that are distantly supervised using positive and negative keywords. As described in the paper [1], we follow Tang's [2] three neural architectures, which encode the sentiment of a word in addition to its semantic and syntactic representation.*

**Specifications Table**

| | |
|---|---|
| Subject area | *Natural Language Processing* |
| More specific subject area | *Arabic Sentiment Embeddings* |
| Type of data | *text files* |
| How data was acquired | *Training Tang's [2] models on an Arabic tweets dataset that was independently collected.* |
| Data format | *Raw* |
| Data source location | *Not applicable* |
| Data accessibility | |

**Value of the data**

- May replace hand-engineered features for sentiment classification.
- Can be used for benchmarking other Arabic sentiment embeddings.
- The Arabic sentiment embeddings can be used for other NLP tasks where sentiment is important.

**Data**

We include three files, each corresponding to one of the models which are described in detail in [1]:

1. embeddings_ASEP.txt: the **A**rabic **S**entiment **E**mbeddings built using the **P**rediction model.
2. embeddings_ASER.txt: the **A**rabic **S**entiment **E**mbeddings built using the **R**anking model.
3. embeddings_ASEH.txt: the **A**rabic **S**entiment **E**mbeddings built using the **H**ybrid model.

Each of the files contains 212,976 lines, starting with the word in the vocabulary, followed by a space, and then 50 decimal numbers separated by spaces (which represent the word vector).

**Acknowledgements**
The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through the Research Project No R17-03-69.

**References**

1. *N. Al-Twairesh, H. Al-Negheimish, Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets , in submission.*
2. *D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, B. Qin, Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification, in: Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap., Association for Computational Linguistics, Baltimore, Maryland, 2014: pp. 1555–1565. http://www.aclweb.org/anthology/P14-1146 (accessed May 18, 2018).*