attr2vec: Jointly Learning Word and Contextual Attribute Embeddings with Factorization Machines
Fabio Petroni, Vassilis Plachouras, Timothy Nugent, Jochen L. Leidner
https://github.com/thomsonreuters/attr2vec

Motivation
The use of word embeddings is considered a “secret sauce” for many recent algorithms.

Popular word embeddings models:
- code.google.com/archive/p/word2vec
- nlp.stanford.edu/projects/glove
- github.com/facebookresearch/fastText

Modeling input data
We model the input data in terms of a target vector \( Y \in \mathbb{R}^m \) and feature matrix \( X \in \mathbb{R}^{m \times n} \), where:
- each row \( x_i \in \mathbb{R}^n \) corresponds to a feature vector;
- variables \( V \) are the set of all considered words and contextual attributes;
- in \( X \) there are as many columns as the number of variables;
- columns grouped according to the type of the variables;
- target value \( y_i \in Y \) represents the number of times the feature vector \( x_i \) has been observed.

We consider a two-fold way to compute the co-occurrence count of a feature vector \( x_i \):
1. linear bag-of-words
2. dependency-based

Linear Bag-of-Words
We compute the co-occurrence count from dependency tree:
- the number of features is increased;
- computational cost increased w.r.t. context-agnostic models (although linear).

In future work:
- we aim to investigate the effect of adding different contextual information;
- we plan to test the resulting models in various applications.