Stock2Vec User Guide

The purpose of those Jupyter Notebook:

* Comparison between the R2 and other metrics of prediction models, and they will get back the comparison chart to you as well
  + - prediction\_employee
    - prediction\_esg
    - prediction\_market capital
* PCA chart in 4D, 3D, 2D and other versions
  + - Pca
    - train\_embeddings - word2vec\_t-sne
* dimensionality reduction with autoencoder instead of PCA, but we don’t need to reduce the dimensions after word embeddings model. It will need if we build the model using neural network model. Thus, just an autoencoder implementation
  + - train\_embeddings - word2vec\_autoencoder + prediction
* fine tuning and LSTM model implementation, but we haven’t tuned the parameters.
  + - train\_embeddings\_updated - word2vec- fine tuning
    - train\_embeddings\_updated - word2vec - LSTM tuning

thoughts for the future work

* new input is needed, we think the input should be the matrix like this:



X: to measure the position change of the stock in the sentence moves between 2 days. E.g. first day: [APL, …..], second day: [A, APL,….], then should be [-1, 0, 0]

Glove embeddings: to get the vector representation of the position change in the glove pretrained model. It’s a method to make the computer read ‘-1’ is smaller than ‘1’

One-hot embedding: to measure the ordering of the company name. it should not be changed even the day change. Because the order is mixed.

* New model is needed, you probably can try to use LSTM with attention instead of word2vec
* Word2vec and FastText are getting similar results just because they are similar models. FastText has one more N-gram feature in term of word2vec