Design Review

Name: Piyush Chawla

Word embeddings have recently gained attention in the field of natural language processing. They have become so ubiquitous that most natural language processing pipelines start with an embedding layer or pre-trained embedding dictionaries. Mikolov et al. in their work, “Distributed Representations of Words and Phrases and their Compositionality”, demonstrate the semantic abilities of word-embeddings trained on a skip-gram model (word2vec). Their experiments showed that words which have same linguistic meanings like Apple, Orange (fruits) are closely placed (in clusters) in the embedding space. They also showed that the embeddings capture relationships (like country-capital relationship) between words. This tells that the word embeddings pack-in a lot of semantic information. Since human visualization is limited to only three dimensions, the most commonly used approach used by NLP researchers, to scrutinize these embeddings has been limited to dimensionality reduction algorithms like PCA[1] or t-SNE.

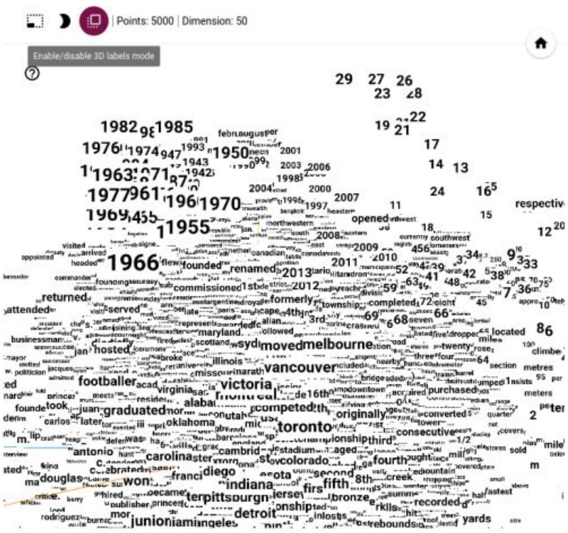


Figure 2: 3D labels view of word embeddings.

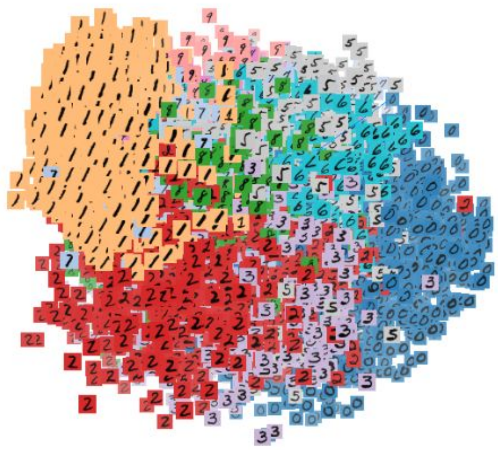


Figure 3: Image view of the MNIST dataset.

Figure A

Figure A

Smilkov et al.[2] give a tool for generating visualizations for embeddings (not limited to words-embeddings). The tools is integrated into tensorflow, a commonly used machine learning library. Figure A shows, two outputs generated by this tool. The

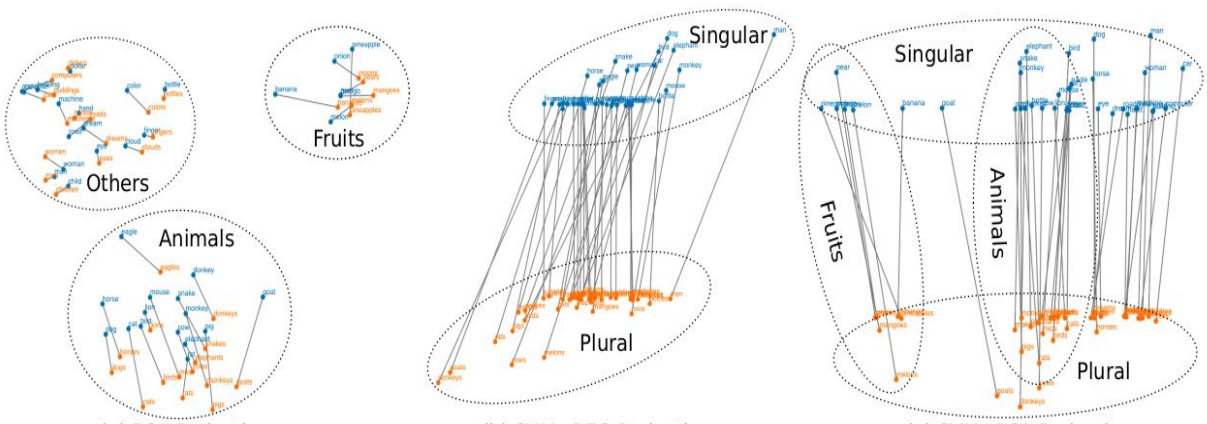
---

[1] Principal Component Analysis

[2] Embedding Projector: Interactive Visualization and Interpretation of Embeddings; Daniel Smilkov, Nikhil Thorat, Charles Nicholson, Emily Reif, Fernanda B. Viégas, Martin Wattenberg; NIPS 2016

first figure shows a 3D projection of word-embeddings trained on a corpus (like Wikipedia). This image captures the variance in the word-embeddings and is a good preliminary visualization, but since word-embeddings are capable more complex inferences, there is still room for improvements. The most noticeable elements are the clusters of words. Years, numbers and place names form three different clusters. An important missing information in the image is the scale. Even though this is a PCA output, the tool supports custom axes inputs as well (where users define their own axes). In such cases scale and ticks might be needed to better analyze the visualization. There is no use of colors. A better visualization would use colors for showing different clusters. Some words like footballer and graduate are closely placed, this might be because of the context of text corpus, but these two words might not be so commonly used together in general. This can lead to misinterpretations.

The second figure is 2D projection of MNIST data. Again, clusters are the most important aspect of this visualization and since MNIST does not contain other semantic information, this is the most important aspect for this dataset. Different colors have been used for different number-classes which helps in discerning the clusters. A 3D projection however, might be more effective because it can be seen that class 3 and class 2 are overlapping. Viewing the same dataset on a 3D plot might give more definitive clusters.



(a) PCA Projection (b) SVM+REG Projection (c) SVM+PCA Projection

Figure B

Figure B

Figure B shows another word embedding visualization technique taken from a study conducted by Liu et al.[3] The image shows two classes of words, singular (blue) and plural (orange). The figure (a) is a simple PCA-based 2D-projection of the word-embeddings. It can be seen that PCA clusters words into different groups just like Figure A. However, this image gives more information than Figure A. It captures the relationship between singular and plural words using different colors. The use of different colors and lines between singular words and their plural counterparts makes this methods effective. (b) and (c) use an approach different from (a). Instead of only capturing the variance in the data, new axes are selected such that more emphasis is put on singular-plural relationship (analogical relationships). The x-axis separates words based on their linguistic features (meanings) and y-axis maintains the distance between singular-plural word pairs. This is achieved by a regression model. This is a better approach than (a) as it captures the domain specific relationship (singular-plural analogy in this example).

Conclusion

Figure B is a better visualization technique than Figure A. Visualizing reduced dimensions give information about the overall variance. Visualizing some trivial datasets, like MNIST, can give an idea about the distribution of different class labels, but a complex dataset like word-embeddings or knowledge-embeddings[4] needs more than just simple variance preservation. Word-embeddings are known to capture complex relationships like gender, hierarchical properties etc. The method used in Figure B, preserves the singular-plural relationship between words rather than just the overall variance of the dataset. Figure B also uses different colors for the two classes and lines that connect singular and plural words, this further facilitates the inference process. One problem with the image is small size of text, which makes reading name of datapoints (embeddings) difficult.

---

[3] Visual Exploration of Semantic Relationships in Neural Word Embeddings; Shusen Liu, Peer-Timo Bremer, Jayaraman J. Thiagarajan, Vivek Srikumar, Bei Wang, Yarden Livnat and Valerio Pascucci; IEEE Transactions on Visualization and Computer Graphics (Volume: 24 , Issue: 1 , Jan. 2018)

[4] Knowledge Graph Embeddings. e.g. Freebase, DBpedia etc.