



VISUALIZING WORD EMBEDDINGS

KATE REISS

WORD EMBEDDINGS

- Words represented as vectors
- Closer vectors = similar meaning
- Differ between corpora and models



CAN WORD EMBEDDINGS BE REPRESENTED IN 3D?



DATA

- 50-dimensional Gensim pre-trained word embeddings
 - Wikipedia 2014
 - 6B tokens
 - 400,000 words in vocabulary
-



CONCLUSION

- Dimensionality reduction can maintain patterns in original data
- App provides visual representation of how machines learn semantics



THANKS!

FUTURE WORK

- Contextual word embeddings; RNN
- Add different models to app
- Reduce loading time

TensorBoard Embedding Projector

