Improving Context-Aware Semantic Relationships in Sparse Mobile Datasets

Peter Hansel, Nik Marda, William Yin
{pwhansel, nmarda, wyin}@stanford.edu

Overview
➢ Cutting-edge NLP techniques often fail to capture semantic context
➢ Microblogging (and many other types of mobile datasets) have inputs other than text
➢ How do we make relationships between sentences more semantically salient using multimodal data?

Data and Features
➢ Politician Tweets Dataset [1]
  ○ Tweets associated with user locations
  ○ Coordinates collected using GeoPy Nominatim API
  ○ Date/time encoded as cyclical continuous feature
  ○ Data stripped of URLs and NLTK toolkit stopwords
➢ Tweet similarity data labeled by political science students and averaged

Existing Methods
➢ Doc2Vec generates a sentence embedding space allowing for comparison [2]
➢ CoSaT uses contextually significant words in weighted BoW embeddings [3]
➢ Neither incorporates non-textual data

Models
➢ Iterative Minimization - Given embeddings \( a, b, \) and multimodal features \( m_{a,j}, m_{b,j} \), iteratively optimized various distance functions \( d \) for various multimodal features:
  \[
  f(a, b, (m_{a1}, m_{b1}), (m_{a2}, m_{b2}), \ldots) = a \cdot b + d_{1}(m_{a1}, m_{b1}) + d_{2}(m_{a2}, m_{b2}) + \ldots
  \]
➢ PCA for dimensionality reduction of sentence embedding space
➢ t-Distributed Stochastic Neighbor Embedding (t-SNE) for constructing visualizations and determining relative similarity [4]

Iterative Minimization
➢ Manually-annotated comparisons
➢ Distance function
  \[
  d_{i}(m_{a,j}, m_{b,j}) = e^{-|m_{a,j} - m_{b,j}|}
  \]
➢ Iteratively optimizing objective
  ○ Discrete ranking system means no continuous gradient
  ○ Minimizing this function:
    \[
    L(x_1, x_2, \ldots) = \sum_{i,j}[y(i, j) - y(i, j)]^2
    \]
➢ Scaling outputs of distance functions / integrating into \( f \) above

PCA and t-SNE
Excluding Multimodal Features
Including Multimodal Features

Discussion
➢ Multimodal data improves recognition of semantic relationships
➢ Especially valuable when tweets are about the same event but lack textual similarity
➢ Iterative Minimization has an upper bound on performance

Future Directions
➢ Test on tweets from local politicians and see if they differ from national politicians (controlling for location)
➢ Distort the word embedding to directly incorporate information from multimodal features
➢ Beyond Twitter and microblogging: other extended data

References