Optimal Semantic Distance for Negative Example Selection in Grounded Language Acquisition

Nisha Pillai, Francis Ferraro, Cynthia Matuszek
npillai1 | ferraro | cmat @ umbc.edu
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County, Baltimore, Maryland 21250

I. OVERVIEW AND RELATED WORK

Grounded language acquisition, in which the meanings of utterances are learned from and with respect to the physical world, is often treated as a data-driven machine learning problem. For a robot, obtaining negative examples of language referents is a challenging problem: people tend to describe things that are true of a situation, rather than negatives about it [15, 4, 3]. For example, humans may are unlikely to describe an apple as “not a banana.” Previous methodologies to acquire negative examples include explicit prompting [13, 2] or crowdsourcing [14, 5]. Other work selects negatives randomly from the dataset, sometimes omitting those with shared descriptions [12, 1].

Our prior research [9] addressed this in an unsupervised system that learned language using a “words-as-classifiers” approach [7], using semantic similarity to automatically choose negative examples from a corpus of perceptual and linguistic data. This joint model of grounded language acquisition is based on the idea that descriptions of physically similar objects should be nearby in semantic space. We used the well-known paragraph vector (PV) [6, 8] encoding to embed these object descriptions in a semantic vector space. Cosine similarity was then used to discover the semantic distance between vector representations: the angle between the PV representations of descriptions was treated as an object similarity metric. Objects with the largest cosine distance were chosen as negative data points. This work is comparable to the unsupervised label identification of Roy [11], but uses document similarity instead of clustering.

The most immediate outstanding question regarding this approach is how to choose an appropriate distance (angle) to select negative examples; when training classifiers, the least-similar object is often not the ideal choice. Previously, it was chosen empirically; our current research aims to more rigorously find and describe a suitable way of selecting negative samples in the semantic embedding space. Our initial findings show that objects which are semantically closer but with non-overlapping characteristics give better results in our grounded language learning experiments.
II. Preliminary Results

Initial experiments on finding an optimal distance threshold in selecting negative examples were conducted on a dataset containing 72 objects in 18 distinct categories (such as food and toys). We collected 6,000 descriptions of these objects from Amazon Mechanical Turk and chose as negative examples objects whose cosine distance was greater than 80 degrees, an empirically chosen threshold [9]. The language groundings were created based on a joint model of language and vision [7, 10] that jointly trained classifiers to predict linguistic descriptions from an object’s visible features. Here, we aim to understand the implications of using different thresholds point through theoretical and practical approaches.

In this work, we conducted experiments on selecting negative examples from different areas in the semantic distance space, which we divided into quintiles for early experiments. Figure 2 shows the cosine angle distance between our objects on the left to the remaining objects in the dataset. The orange line in the figure marks the threshold selected as part of previous research, where everything to the right of the orange line was selected as a negative example. In contrast, the blue line suggests a possibly more optimal threshold for selecting best negative examples for language acquisition tasks. The objects to the right of the blue line represent the “most different” objects in semantic space.

As an additional step, we trained our visual classifiers on color, shape, and object features with every section of the negative training data selected by dividing the dataset into quintiles. Figure 1 depicts our initial results in which color, shape, and object language acquisition show promising predictive performance when selecting the 2nd and 3rd quintiles of objects as negative samples. These results suggest a more dynamically-chosen threshold may yield improved performance.

III. Discussion and Future Work

Our aim is to build a model which selects the most informative negative data points from the complete negative dataset. Efficient selection of semantic distance will underpin this grounded language learning by providing negative examples without prompting the user explicitly, reducing the number of (possibly repetitive) questions. A thorough evaluation including Mechanical Turk user studies will be conducted to ensure the effectiveness...
of the model. We are also considering alternative methods of semantic similarity to further improve the performance of our model.

REFERENCES


