HYBRID DISTRIBUTIONAL AND DEFINITIONAL WORD VECTORS

Haiyuan Mei and Ranjani Iyer
Stanford University, Department of Computer Science

OVERVIEW

- **Motivation**: Out Of Vocabulary (OOV) problem - exploration of word definitions in downstream NLP tasks.
- **Prior methods**: Def2Vec, on-the-fly embeddings capable of capturing OOV words, and limited usage exploration. [1]
- **Approach**: HybridVec Generate word embedding from word definitions, combine it with distributed representations, and explore the possibility or improving downstream NLP tasks.
- **Evaluation**: Intrinsic word embedding benchmarks and Extrinsic NMT evaluation, shown to improve translation perplexities and capture complementary aspect of word regarding distributed representation.

RESULTS

- **Word embeddings benchmarks for GloVe, LSTM Baseline and Seq2Seq model. LSTM baseline model is roughly at the level of distributional method; Seq2Seq model shows very limited evidence of such capability[2]:**

<table>
<thead>
<tr>
<th></th>
<th>BLESS</th>
<th>ISSUU 14</th>
<th>MEN</th>
<th>MTurk</th>
<th>RG65</th>
<th>SL199</th>
<th>W1453</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glove</td>
<td>0.82</td>
<td>0.75</td>
<td>0.237465</td>
<td>0.633182</td>
<td>0.769525</td>
<td>0.3705004</td>
<td>0.543526</td>
</tr>
<tr>
<td>Baseline glove</td>
<td>0.55</td>
<td>0.509991</td>
<td>0.51071</td>
<td>0.4256407</td>
<td>0.664302</td>
<td>0.367968</td>
<td>0.489056</td>
</tr>
<tr>
<td>Baseline rand</td>
<td>0.52</td>
<td>0.519636</td>
<td>0.444708</td>
<td>0.3185651</td>
<td>0.6449086</td>
<td>0.3281222</td>
<td>0.556069</td>
</tr>
<tr>
<td>s2s en mean</td>
<td>0.275</td>
<td>0.527272</td>
<td>0.106169</td>
<td>0.1707244</td>
<td>0.0908722</td>
<td>0.186833</td>
<td>0.051959</td>
</tr>
</tbody>
</table>

- **GloVe: WEB benchmark for GloVe vectors**
- **Baseline glove: WEB benchmark for LSTM baseline model initialized from GloVe**
- **Baseline rand: WEB benchmark for LSTM baseline model initialized randomly**
- **s2s en mean: WEB benchmark for Seq2seq model with encoder output mean as the def vec.**

- OpenNMT compare performance improvements using LSTM baseline vector and GloVe:

<table>
<thead>
<tr>
<th></th>
<th>No pretrained</th>
<th>LSTM Baseline</th>
<th>GloVe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train PPL</strong></td>
<td>7.47</td>
<td>6.69</td>
<td>4.84</td>
</tr>
<tr>
<td><strong>Train ACC</strong></td>
<td>56.29</td>
<td>58.4</td>
<td>64.32</td>
</tr>
<tr>
<td><strong>BLEU</strong></td>
<td>0.0093</td>
<td>0.0137</td>
<td>0.0199</td>
</tr>
</tbody>
</table>

*10k nmt training 10 epochs, one on each of nmt val sentences, similar result as above perplexity and accuracy.

ANALYSIS

- **LSTM baselines vectors tend to cluster in feature space. Need to train from a broader source.**
- **GloVe makes use of feature space more efficiently, grasp more subtle meaning of words.**

3D tsNE: LSTM Baseline vectors is likely to cluster

GloVe uses feature space more efficiently

FUTURE WORK

- **Additional plans for model: greater regularization, inputting multiple definitions, inputting sentence structure, try other embeddings.**
- **Continue the exploration of combining different word vectors for downstream NLP tasks.**

REFERENCES


MODEL

- **Baseline LSTM**: A two-layer LSTM encoder, Simple linear decoder and NLL loss, where the encoder layer hidden output denotes the final definitional word vector.
- **Seq2Seq**: A two-layer LSTM encoder with dropouts plus a two layer LSTM decoder without attention.
- **Variational AutoEncoder**: Adapted VAE with single-layer LSTM encoder and decoder with Gaussian prior regularizer[3].

TRAINING

- **Dataset**: GloVe [4]. All models are trained on pretrained 300d GloVe vectors based on a crawl of 2014 Wikipedia. Definitions retrieved from WordNet[5].
- **HybridVec Implementation**: Pytorch, Adam optimizer, Xavier initialization, hidden size 150, learning rate of 1e-4, batch size 64, 15/20 epochs.
- **Intrinsic evaluation**: Word embedding benchmarks [2]

<table>
<thead>
<tr>
<th></th>
<th>OpenNMT-py demo (10k)</th>
</tr>
</thead>
</table>

NMT Dataset: OpenNMT-py demo (10k) dataset. Only for comparison between GloVe and HybridVec.