Aicyber

Aicyber's System for IALP 2016 Shared Task:Character-enhanced Word Vectors and Boosted Neural Networks

Steven Du , Xi Zhang

Outline

Task overview

- To predict a given Chinese word’s affective states in continuous numerical values (from 1 to 9) on valence-arousal space.
- Supervised regression task:
  - 1653 for training
  - 1149 for testing
- System for supervised learning task
  - Font-end Features
  - Back-end Models (classifiers/regressors)

Features: Overview

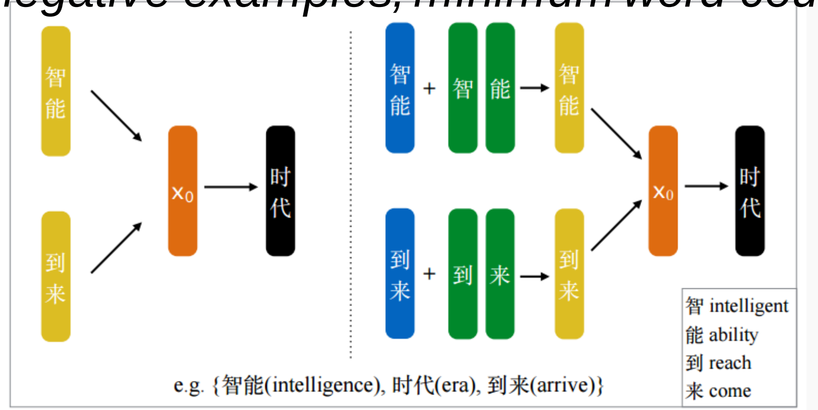
- Given that training and test are all Chinese words, they are unique and no overlapping, thus estimating word representations in a common vector space is required to "link" the training and test set.

- Methods could learn word vectors from collection of text:
  - Vector Space Model
  - Latent Semantic Analysis
  - Latent Dirichlet Allocation
  - Explicit Semantic Analysis
  - Distributed word representation (word2vec)

- Two word embeddings derived from word2vec are used in our system

Features: CWE

- First set of features derived from distributed word representation: Character-enhanced word embedding [1] (CWE)
- Two type of CWE: CWE+P , CWE+L
- Training parameters: *CBOW or Skip-gram , window size of 5, 5 iterations, 5 negative examples, minimum word count of 5*



(A) CBOW (B)Character-enhanced Word Embedding

[1]Xinxiong Chen, Lei Xu, Zhiyuan Liu, Maosong Sun, and Huanbo Luan. 2015b. Joint learning of character and word embeddings. In International Conference on Artificial Intelligence

Features: FastText

- The second type of embedding feature: FastText, train word vectors with character n-grams
- Eg: with min character n-grams set to 3

  hometown -> hom ome met ...own.. home town … hometown
- Training parameters: *CBOW or Skip-gram, window size of 5, 5 iterations, 5 negative examples, minimum word count of 5,* ***min n-grams=1***

Features: training data for word embedding

Following public datasets are used:

1) Chinese Wikipedia Dumps (Time stamp: 2011-02- 05T03:58:02Z), however use of the latest dumps is encouraged.

2) Douban movie reviews

3) Aicyber synthesized 200 sentences . These are intended to cover the out of vocabulary words in the task.

Resulting 445662 Chinese words, mapping of words to remove OOV in the task , 瞭解 > 了解 (understood).

Regressors

- Regressors used in the submitted system:
    - Boosted Neural Network, Boosting algorithm (AdaBoost.R2) applied on neural networks
    - Each neural network has only one hidden layer, and its size is 100, with *relu* activation function, *adam* as its training algorithm and a constant learning rate of 0.001.

Evaluation

- Done locally on valence estimation only
- 3 round of 10 folds cross-validation
- To identify the best feature (and its training parameters) and regression method

word2vec	< Baseline >	Linear-SVR
CWE+P		Boosted Neural Network
CWE+L		Neural Network
FastText		GBM
		XGB

Evaluation of Features

- C: continuous bag-of-words model
- S: Skip-gram
- 100/300:dimensions of feature vectos

Valence MAE by LSVR Baseline				
Embeddings	**C-100**	**S-100**	**C-300**	**S-300**
word2vec	0.936	0.839	0.950	**0.819**
CWE+P	0.940	0.773	0.940	**0.765**
CWE+L	0.953	0.827	0.923	**0.769**
FastText	1.093	0.796	1.343	**0.765**

Table I

A LSVR IS APPLIED TO DIFFERENT TYPE OF EMBEDDING FEATURES, GROUPED BY TRAINING METHODS AND SIZE OF FEATURE VECTOR.

Valence MAE by Boosted Neural Network				
Embeddings	**C-100**	**S-100**	**C-300**	**S-300**
word2vec	0.878	0.757	0.952	**0.756**
CWE+P	0.823	0.702	0.837	**0.670**
CWE+L	0.823	0.741	0.816	**0.662**
FastText	0.876	0.695	0.947	**0.668**

Table II

BOOSTED NEURAL NETWORK REGRESSION METHOD APPLIED TO DIFFERENT TYPE OF FEATURES.

- Summery:
  - Skip-gram always outperform CBOW
  - **Skip-gram, 300 DIM are the best training parameters**

Evaluation of Regression Methods

Done on S-300 Embedding

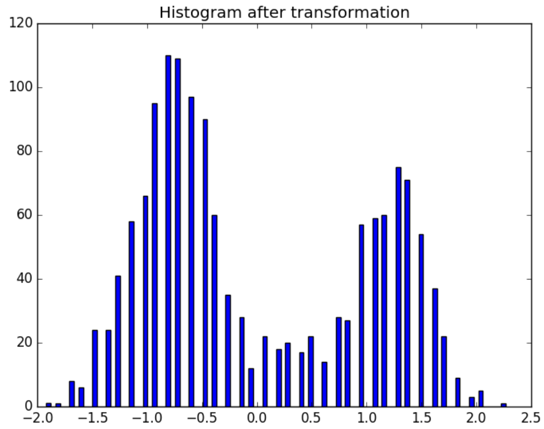
Embedding S-300	Valence MAE by Different Regression Methods				
	LSVR	GBM	XGB	NN	BNN
word2vec	0.819	0.829	0.881	0.801	**0.756**
CWE+P	0.765	0.757	0.809	0.729	**0.670**
CWE+L	0.769	0.795	0.860	0.730	**0.662**
FastText	0.765	0.791	0.847	0.711	**0.668**

Table III

EVALUATION OF DIFFERENT REGRESSION METHODS APPLIED TO S-300 EMBEDDING.

Evaluation of Regression Methods

- Done on S-300 Normalize target value leads to better MAE and PCC



Embedding S-300-PCA100	**Valence MAE**		**Valence PCC**	
	BNN	BNN_Norm	BNN	BNN_Norm
word2vec	0.702	**0.686**	0.858	**0.867**
CWE+P	0.678	**0.623**	0.874	**0.891**
CWE+L	0.671	**0.627**	0.873	**0.891**
FastText	0.662	**0.639**	0.879	**0.889**

Table IV
NOTABLE IMPROVEMENT IN MAE AND PCC MADE BY NORMALIZED BNN APPROACH.

Final Submission

- Run1 = Average of BNN_Norm systems trained on CWE+P, and FastText
- Run2 = BNN_Norm system trained on CWE+P
- Same approaches are used for valence and arousal estimation
- Due to a "bug" in the script CWE+L is not included in the ensemble.

Run#	Valence MAE(rank)	Valence PCC(rank)
Run1	**0.577 (1)**	0.848 (8)
Run2	0.581 (3)	0.843 (11)

Run#	Arousal MAE(rank)	Arousal PCC(rank)
Run1	1.212 (8)	**0.671 (1)**
Run2	1.215 (9)	0.662 (3)

Conclusion & Discussion

- A boosted neural networks build on character-enhanced word vectors, it works “well” on given data.

- Performance of the system should further investigated by repatriation train/test data set.

- Further improvement:
  - Text , speech , vision are just communication channels
    - word2vec
    - Acoustic word vector
    - Visual word vector (Google DeepMind:Lipnet)

Q & A

https://github.com/StevenLOL/ialp2016_Shared_Task

steven@aicyber.com
zhangxi@aicyber.com