Dimensional Sentiment Analysis of Traditional Chinese Words Using Pre-trained Not-quite-right Sentiment Word Vectors and Supervised Ensemble Models

Feixiang Wang[2], Yunxiao Zhou[2], Man Lan[1,2]

1 Shanghai Key Laboratory of Multidimensional Information Processing

2 Department of Computer Science and Technology

East China Normal University Shanghai 200241, P. R. China

Task Definition

- **Shared Task**: Dimensional Sentiment Analysis for Chinese Words

  i.e., to automatically acquire the valence and arousal ratings of Chinese affective words.

Input:

```
1,不可思議
2,不平
3,不甘
4,不安
5,不利
6,不足
7,不定
8,不宜
9,不幸
10,不拘小節
```

Output:

No.,Word,	Valence,	Arousal
1,不可思議	5.4	7.2
2,不祥	2.8	5.2
3,成功	8.2	6.6
4,閒散	4.6	2.2
5,不爽	2.8	7.2
6,不足	4.2	4.2
7,不定	4.2	3.8
8,不宜	3.8	4.4
9,不幸	2.6	6.0
10,不拘小節	6.4	4.2

System Overview

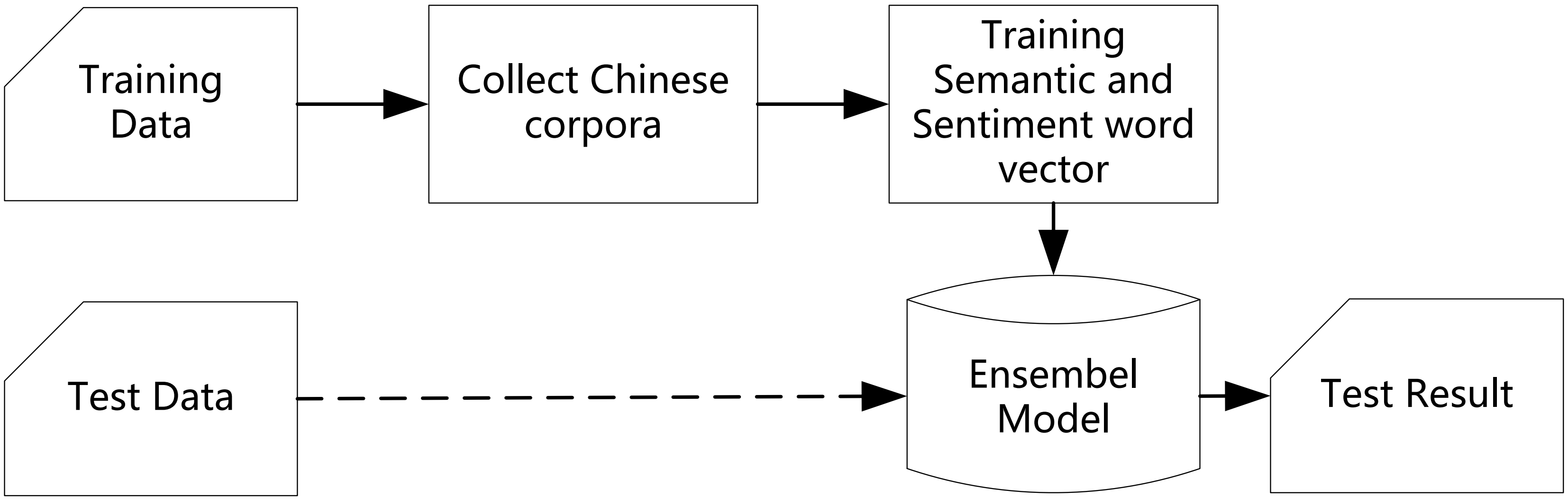

Figure 1. The framework of our System

Data Collection

- **We collected two Chinese corpora to train word vectors:**
  - Wikimedia dumps[1]: 3,736,800 sentences
  - SogouCA[2]: 2,588,466 sentences

- **Preprocessing:**
  - using OpenCC[3] converted the simplified Chinese text into the traditional Chinese text
  - tokenized with jieba[4] tokenizer

1. https://dumps.wikimedia.org/zhwiki/20160501/
2. http://www.sogou.com/labs/resource/ca.php
3. https://pypi.python.org/pypi/OpenCC
4. https://github.com/fxsjy/jieba

Word Vector Training

- **Not-quite-right Sentiment Word Vector (SWV)[1]**:
  - Training corpora: Wikimedia corpus, SogouCA
  - Annotation strategy(take valence label for example):

(1) Seed words: 10% of the most pleasant words and most unpleasant words from CVAW[2]

(2) Pleasant sentence: one or more seed words with pleasant label in a sentence

(3) Unpleasant sentence: one or more seed words with unpleasant label in a sentence

(4) Dropout strategy: dropout the sentence without any seed words

1. There are no gold VA labels for word vectors training data, and it is costly to manually annotate their sentiment labels. To address it, we adopt the some annotation strategy in a low cost but sound reasonable way. Therefore, the sentiment word vectors pre-trained in this way may not quite right.
2. IALP 2016 Shared Task training data set

Word Vector Training

- **Not-quite-right Sentiment Word Vector (SWV):**
  - wiki_VSWV[1]/ wiki_ASWV: trained on the Wikimedia corpus
  - sg_VSWV/ sg_ASWV: trained on the SogouCA
  - tool: SWV-C[2]

- **Semantic Word Vector(WV):**
  - wiki_WV: trained on the Wikimedia corpus
  - sg_WV: trained on the SogouCA
  - tool: Google W2V[3]

1. wiki_VSWV represents the SWV trained on Wikimedia corpora with valence orientation
2. L. Man, Z. Zhihua, L. Yue, and J. Wu, “Three convolutional neural network-based models for learning sentiment word vectors towards sentiment analysis,” in *2016 International Joint Conference on Neural Networks IJCNN*, pp. 3172–3179.
3. https://code.google.com/archive/p/word2vec

Ensemble Model

- **Algorithm:**

  \- Support Vector Machine(SVM):
  kernel = linear, c = 0.02 for valence subtask, c =0.01 for arousal subtask

  \- Stochastic Gradient Descent (SGD)

  \- Gradient Boosted Regression Trees (GBRT):
  n_estimator = 200

- **Weighted Ensemble of Algorithms[1]:**

Subtask	SVM weight	SGD weight	GBRT weight
Valence	0.4	0.1	0.5
Arousal	0.4	0	0.6

1. We investigated ensemble learning models for two subtasks and find that the ensemble learning models show superior performance to single model. And Table lists the three algorithms weights which achieved best results.

Datasets & Evaluation

- **Training Datasets:**
  - IALP 2016 Shared Task corpus(CVAW):
    1,653 affective words
- **Evaluation:**
  - $$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - P_i| \qquad (1)$$
  - $$Pearson = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{A_i - \overline{A}}{\sigma_A}\right)\left(\frac{P_i - \overline{P}}{\sigma_P}\right) \qquad (2)$$

Experiments on Training Data

Table 1: Experimental results of different learning algorithms

Algorithms	Valence		Arousal	
	MAE	**Pearson**	**MAE**	**Pearson**
SVM	0.7287	0.8545	0.6681	0.7533
SGD	0.7533	0.8422	0.7009	0.6984
LR	0.8593	0.7825	0.9433	0.5590
DT	1.0069	0.6616	0.9068	0.5784
GBRT	**0.7015**	**0.8558**	**0.6564**	**0.7680**

==> For both valence and arousal orientation, GBRT consistently outperforms SVM, SGD, LR, and DT algorithm with word vector and sentiment word vector.
==> We chose top 3 algorithms to construct the ensemble model.

Experiments on Training Data

Table 2: Experimental results of various word embedding

Word Embedding	Valence		Arousal	
	MAE	**Pearson**	**MAE**	**Pearson**
wiki_WV	0.8576	0.7913	0.8168	0.5683
wiki_VSWV/wiki_ASWV	0.7525	0.8369	0.6603	0.7559
wiki_WV+ wiki_VSWV/wiki_ ASWV	0.7015	0.8558	0.6564	0.7680
sg_WV+ sg_VSWV/sg_ASWV	0.7548	0.8322	0.7124	0.7019
wiki_WV+ sg_WV+ wiki_VSWV/wiki_ASWV	**0.6915**	**0.8669**	0.6529	0.7677
wiki_WV+ wiki_VSWV/wiki_ASWV+ sg_VSWV/sg_ASWV	0.6936	0.8640	**0.6363**	**0.7782**
wiki_WV+ sg_WV+ wiki_VSWV/wik_ASWV+ sg_VSWV/sg_ASWV	0.7036	0.8570	0.6378	0.7749

==> *SWV makes a great contribution to performance improvement.*
==> *The word vectors trained on Wikimedia dumps outperform that trained on SogouCA.*
==> *The best performance is achieved by using the combination of diverse word vectors.*

Experiments on Test Data

Table 3: Performance of our systems and the top-ranked systems on both two subtasks.

Subtask	System	MAE	Pearson
Valence	**ECNUCS Run1** **ECNUCS Run2**	**0.577(1)** 0.639(11)	0.811(17) 0.771(21)
	CKIP	0.583(4)	0.862(3)
	Aicyber	0.577(1)	0.848(8)
Arousal	**ECNU Run1** **ECNU Run2**	1.33(25) 1.333(27)	0.617(9) 0.558(13)
	NCTU+NTUT	1.165(5)	0.631(4)
	Aicyber	1.212(8)	0.671(1)

- The numbers in the brackets are the official ranks.
- Ensemble learning model is named ECNUCS_Run1.
- ECNUCS_Run2 only uses SGD algorithm.

==> Run1 on valence ranked the 1st on the MAE evaluation .

==> Ensemble learning model shows superior performance to single algorithm.

Conclusion & Future Work

- **Conclusion**
  - – For both valence and arousal subtasks, SWV is capable of complementing sentimental information, which is critical for our task.
  - – Ensemble learning model outperforms the model using a single algorithm.

- **Future Work**
  - – Further exploring the difference between valence and arousal to improve our ensemble learning model.
  - – Attempt to apply the constructed resources to sentence level or document level sentiment analysis.