NTU NLP Lab System at SemEval-2018 Task 10: Verifying Semantic Differences by Integrating Distributional Information and Expert Knowledge

Yow-Ting Shiue[1], Hen-Hsen Huang[1], and Hsin-Hsi Chen[1,2]
[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
[2]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
orina1123@gmail.com, hhhuang@nlg.csie.ntu.edu.tw, hhchen@ntu.edu.tw

ABSTRACT

This paper presents the NTU NLP Lab system for the SemEval-2018 Capturing Discriminative Attributes task. **Word embeddings**, **pointwise mutual information (PMI)**, **ConceptNet edges and shortest path lengths** are utilized as input features to build binary classifiers to tell whether an attribute is discriminative for a pair of concepts. Our neural network model reaches about **73% F1** score on the test set and ranks the **3rd** in the task. Though the attributes to deal with in this task are all visual, our models are not provided with any image data. The results indicate that visual information can be derived from textual data.

Introduction

- Modern semantic models: good at capturing semantic similarity and relatedness
- Ability to **distinguish** one concept from another similar concept is also core to linguistic competence; important mechanism for teaching and learning
  - "*subway* is a kind of *train* that runs underground"
  - *plate* v.s. *bowl*: "a *plate* is flatter" / "a *bowl* is deeper"
  - A *discriminative attribute* applies to one of the concepts but does not apply to the other.
- Notation: $< w_1, w_2, a >$ + / -

Distributional Information

[Concatenation of Word Embeddings]

Feature vector = [emb(w_1) emb(w_2) emb(a)]

- Train F1: 0.65 / Test F1: 0.34
- Overfit to training vocab ⇔ test: **all attributes unseen**

[Embeddings Similarity Difference]

If w has a, then it tends to be more similar to a than other words without a. ($\text{sim}_1 = \text{sim}(w_1, a)$, $\text{sim}_2 = \text{sim}(w_2, a)$)
→ Rule: **$\text{sim}_1 > \text{sim}_2$** / Feature: **$\text{sim}_1 - \text{sim}_2$**

Model	Embedding	Pos. F1	Neg. F1	Macro F1
Rule	1. W2V	0.6512	0.5648	0.6080
Rule	2. fastText	0.6435	0.5565	0.6000
Rule	3. Numberbatch	**0.7142**	0.5964	**0.6553**
Rule	4. GloVe	0.6594	**0.6022**	0.6308
Rule	5. Sense-closest	0.6609	0.5068	0.5838
Rule	6. Sense-first	0.5597	0.6013	0.5805
MLP	[**sim x4**] 1. – 4.	0.6572	**0.6521**	0.6546
MLP	[**sim x6**] 1. – 6.	**0.6609**	0.6520	**0.6564**

[PMI Difference]

- PMI is calculated in exact matching manner → no propagation of similarity as word vectors
- E.g. high PMI(*red*, *yellow*) & high PMI(*apple*, *banana*) → **does not imply** high PMI(*red*, *banana*)
- → less prone to confusion of similar concepts

Model	Features	Pos. F1	Neg. F1	Macro F1
Rule	PMI: 10 words context	0.6986	0.5968	0.6477
Rule	PMI: 20 words context	**0.7013**	**0.5948**	**0.6481**
Rule	PMI: 30 words context	0.6959	0.5896	0.6427
MLP	PMI(10+20+30)	0.7026	0.5723	0.6375
MLP	sim x6 + PMI x3	**0.7039**	**0.6432**	**0.6735**

Expert Knowledge from ConceptNet

- Binary features based on ConceptNet graph

[Edge Connection]	[Shortest Path Length]
• Is there an edge from w_1 to a? • Is there an edge from a to w_1? • Is there an edge from w_2 to a? • Is there an edge from a to w_2? r: ConceptNet relation type e.g. HasProperty, CapableOf	• No path from w_i to a • $dis(w_i, a) = 1$ • $dis(w_i, a) = 2$ • $dis(w_i, a) = 3$ • $dis(w_i, a) = 4$ • $dis(w_i, a) \geq 5$

Model	Rule / Features	Pos. F1	Neg. F1	Macro F1
Rule	$w_1 \rightarrow a$ & $w_2 \nrightarrow a$	0.5140	**0.7009**	0.6074
MLP	$w_1 = w_2 \overset{r}{\leftrightarrow} a$ for **each** r	0.4785	0.6376	0.5581
MLP	$w_1 = w_2 \overset{r}{\leftrightarrow} a$ for **any** r	0.4931	0.6661	0.5796
Rule	$dis(w_1, a) < dis(w_2, a)$	0.5740	0.6742	0.6241
MLP	One-hot shortest path lengths	**0.6984**	0.6223	**0.6603**

Submitted Systems

Model	Acc.	Pos. F1	Neg. F1	Macro F1	Rank
[1] Rule: $\text{sim}_1 > \text{sim}_2$ & $\text{PMI}_1 > \text{PMI}_2$	0.7047	0.6944	0.7143	0.7044	(4~5)
[2] MLP: sim x6 + PMI(10,20,30) + ConceptNet	**0.7303**	**0.7138**	**0.7451**	**0.7294**	**3**

Error Analysis

- **Ambiguous concept**

<mouse, squirrel, plastic> **+**

mouse here: "computer device" instead of "animal"
→ answer dependent on which **sense** is selected

- **Vague or ambiguous attribute**

Attribute expressed only with **single word** → sometimes hard to tell attribute meaning even from human perspective

<philanthropist, lawyer, active> **-**

[Positive interpretation]
philanthropist: engage in philanthropy actively
lawyer: handle matters under authorization of someone

- **Relative attribute**

w_1 does **not necessarily** but only **more likely** to have a

<father, brother, old> **+**

father might not be *old* when considered **isolatedly**

<banker, lawyer, rich> **-**

→ When to evaluate an attribute relatively?

Conclusions

- Corpus distribution statistics + expert knowledge
- **Word embedding** & **PMI** complement each other
- **ConceptNet** features (sensitive to **negative** class) complement
  corpus-based features (sensitive to **positive** class)
- F1 = **0.7294** [**3rd** in official run]
- **No image features** → possible to learn substantially about visual attributes solely from text
- Limited advancement of learning-based over rule-based → design mechanism of "comparing" features of two concepts