Constructing Sentiment Sensitive Vectors for Word Polarity Classification

Speaker: Johann Chu

Date: 12/5/15

Introduction

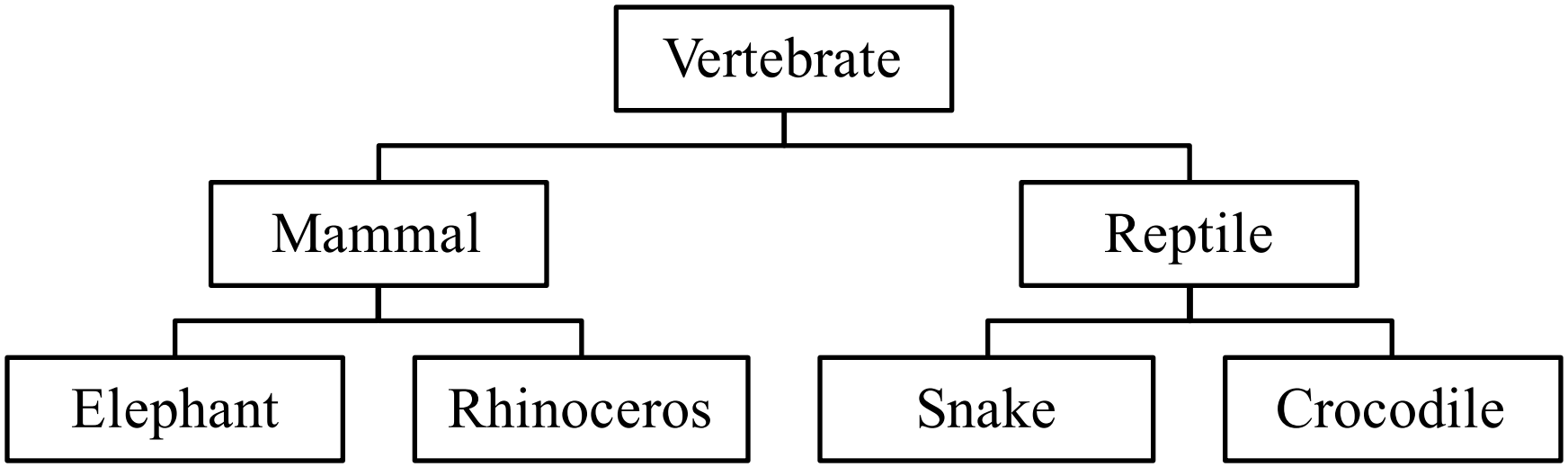
Positive Sentiment Polarity:

I bought this knife set 3 months ago. It was *beautiful* with *unique* patterns on the blade and very *sharp*.

Negative Sentiment Polarity:

Now a month after using, its espresso pump is *loud* and *noisy*, sounding like it's broken.

Introduction (to WordNet)



(For noun and verb only.)

Methodology

- Alternatively, we aim to represent a word with the use of a *vector*, so that it can be quantitatively compared with one another.

- Inspired by Schütze[1] and Patwardhan[2].
- Schütze innovated the use of *word vector*, on which Patwardhan's *gloss vector* was based.

Excellent (adj.) very good; of the highest quality

{good,1; high,1; quality,1} (1-level Gloss Vector)

Unusually *good*

Extraordinarily *good* or great

…

…

Undertaken in *good* faith

Moving at very *high* speed

Full of or showing *high* spirit

…

…

Having a *high* price

Having the power or *quality* of deciding

Rich and superior in *quality*

…

Different in nature or *quality*

Extraordinary	1
Faith	2
Great	1
… …	
Undertake	1

Move	1
Price	1
Speed	1
… …	
Spirit	3

Decide	1
Nature	1
Rich	4
… …	
Superior	2

{(*Faith*, *4*), (*Great*, *8*), (*Rich*, *1*), (*Spirit*, *8*),……,(*Superior*, *13*)}

(2-level Gloss Vector)

Deciding the Depth

1-level Gloss Vector:

Abundant: *present,1; great,1; quantiti,1;*

2-level Gloss Vector:

Abundant: *great,17; quantiti,17; exist,11; time,53; person,30; bodi,11; substanc,12; organ,12; acid,13; extent,25; degre,36; peopl,10; number,45; physic,23; system,11; state,38; pass,10; dai,20; qualiti,21; character,12; measur,35; express,22; volum,11; blood,10; cell,12; form,21; gener,12; amount,21; act,19; unit,47; make,24; parti,10; larg,28; produc,10; ancient,12; northern,15; properti,10; show,14; work,11; perform,10; item,13; denot,15; britain,73; ……*

Sentiment Sensitive Vector

Positive	Gloss Vector
Accessible	{capable,1;easily,1;obtain,1}
Achievable	{capable,1;prove,1;possible,1}
Adorable	{lovable,1;childish,1;naïve,1}
Beautiful	{……}
Nice	{……}
Pretty	{……}
Super	{high,1;quality,1;extreme,1}

Positive Sentiment Sensitive Vector
{capable,2; easily,1; obtain,1;
prove,1; possible,1;……high,1;
quality,1; extreme,1}

Negative	Gloss Vector
Awful	{bad,1;displeasing,1;fear,1, dread,1}
Bad	{undesirable,1;negative,1;regret,1}
Cruel	{dispose,1;inflict,1;suffering,1}
Dreadful	{……}
Sad	{……}
Terrible	{……}
Terrifying	{extreme,1;terror,1}

Negative Sentiment Sensitive Vector
{bad,1; displeasing,1; fear,1; dread,1;
undesirable,1;……extreme,1; terror,1}

Sentiment Sensitive Vector

Positive Sentiment Sensitive Vector

{capable,2; easily,1; obtain,1; prove,1; possible,1;…… high,1; quality,1; extreme,1}

Excellent: {good,1; high,1; quality,1}

Negative Sentiment Sensitive Vector

{bad,1; displeasing,1; fear,1; dread,1; undesirable,1;……extreme,1; terror,1}

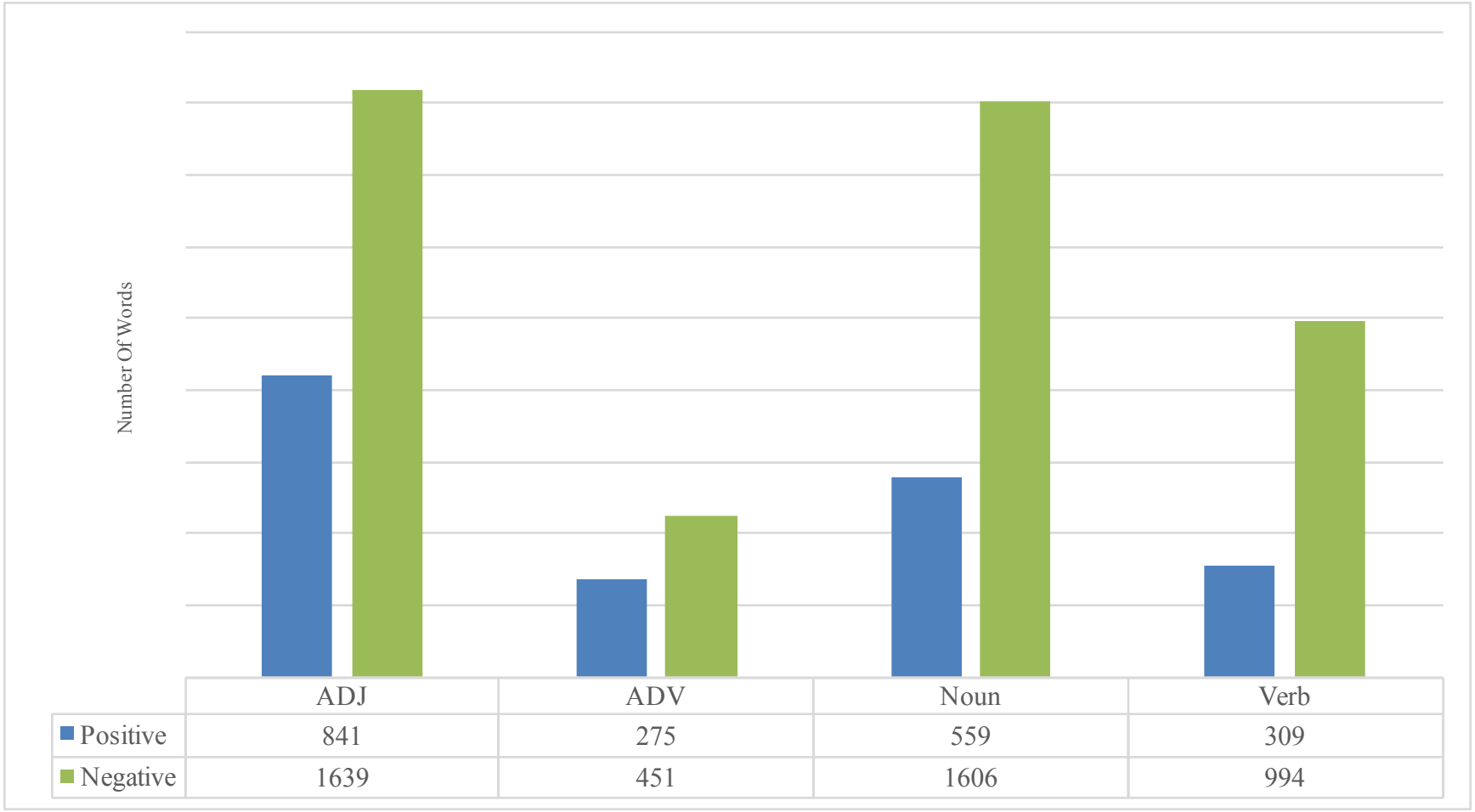
Measuring Similarity

- *Cosine similarity* is used to measure the similarity of the vector with both SSVs.

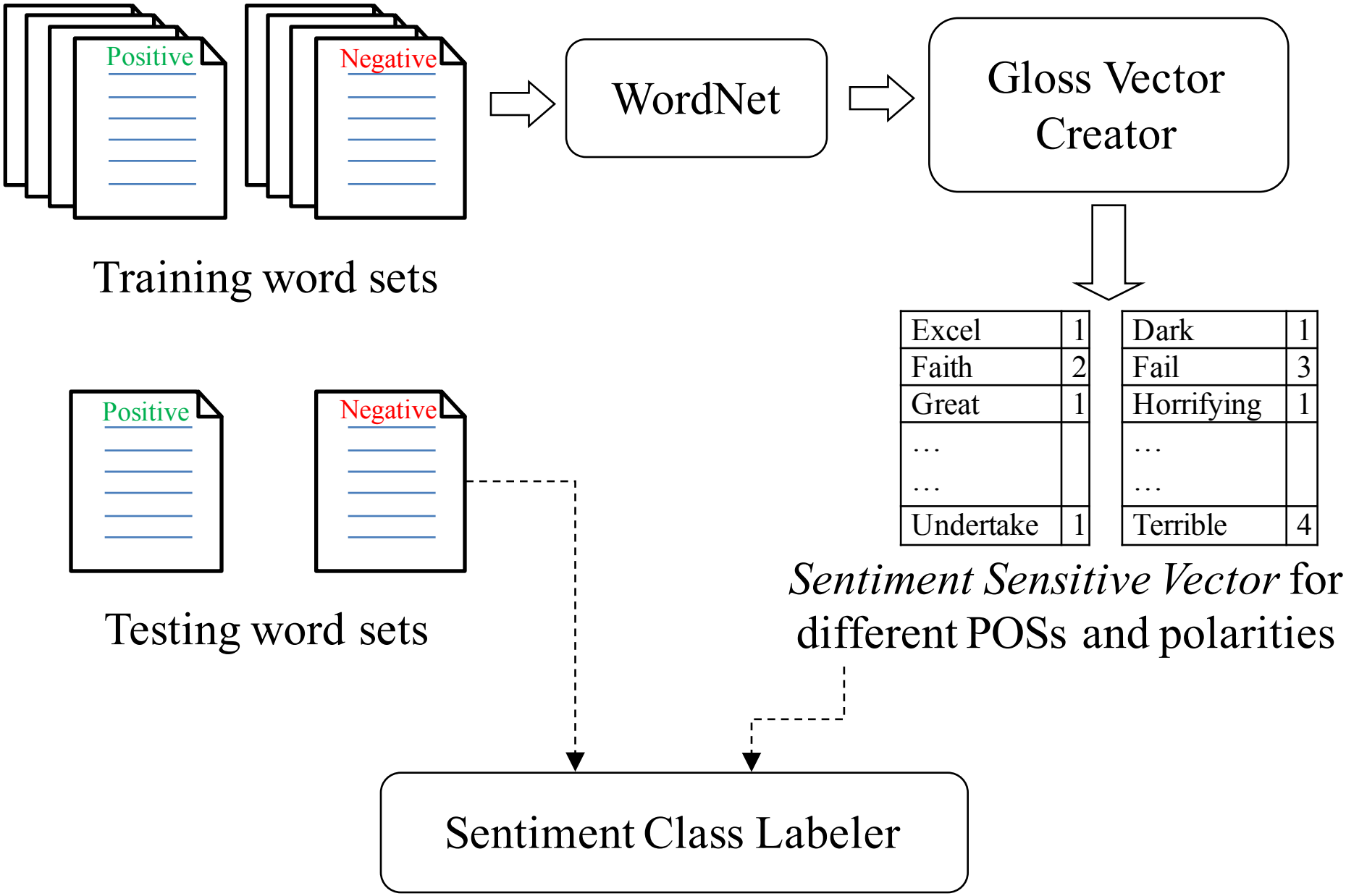
$$Polarity_{w_i} = \arg\max_{SSV \in SSV_p, SSV_n} Sim(w_i, \mathrm{SSV}) = \frac{\vec{V}(w_i) \cdot \vec{V}(\mathrm{SSV})}{|\vec{V}(w_i)||\vec{V}(\mathrm{SSV})|}$$

$$= \frac{\sum \vec{V}(w_i)_{gt} * \vec{V}(\mathrm{SSV})_{sgt}}{\sqrt{\sum_{gt \in GlossVector} \vec{V}(w_i)_{gt} * \vec{V}(w_i)_{gt}} \sqrt{\sum_{sgt \in SentimentSensitiveVector} \vec{V}(\mathrm{SSV})_{sgt} * \vec{V}(\mathrm{SSV})_{sgt}}}$$

Experiment

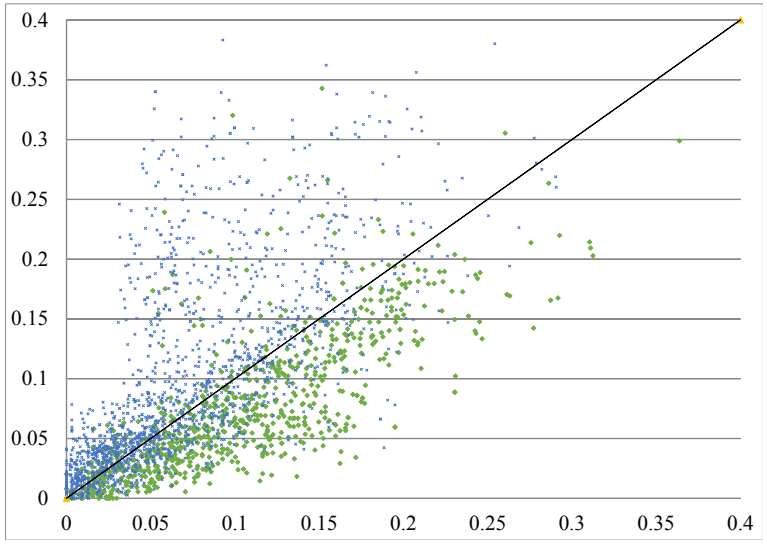


	ADJ	ADV	Noun	Verb
Positive	841	275	559	309
Negative	1639	451	1606	994

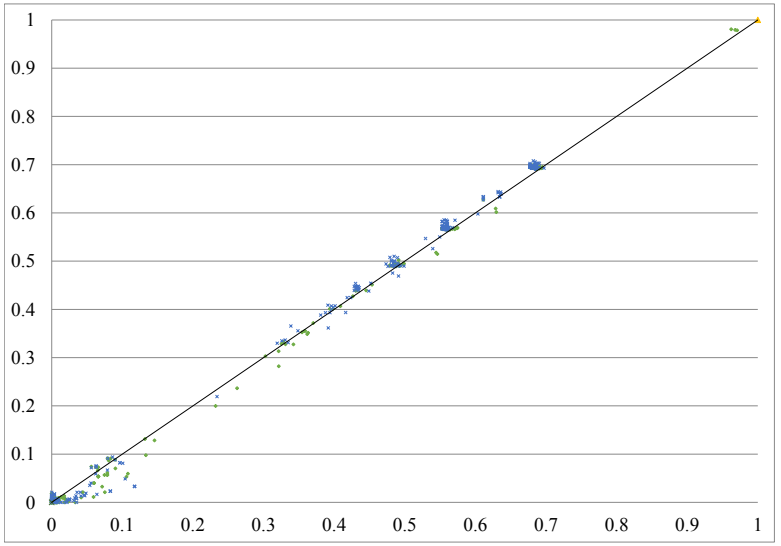

Positive
Negative
Training word sets
WordNet
Gloss Vector Creator
Excel 1
Faith 2
Great 1
… …
Undertake 1
Dark 1
Fail 3
Horrifying 1
… …
Terrible 4
Sentiment Sensitive Vector for different POSs and polarities
Positive
Negative
Testing word sets
Sentiment Class Labeler

Results

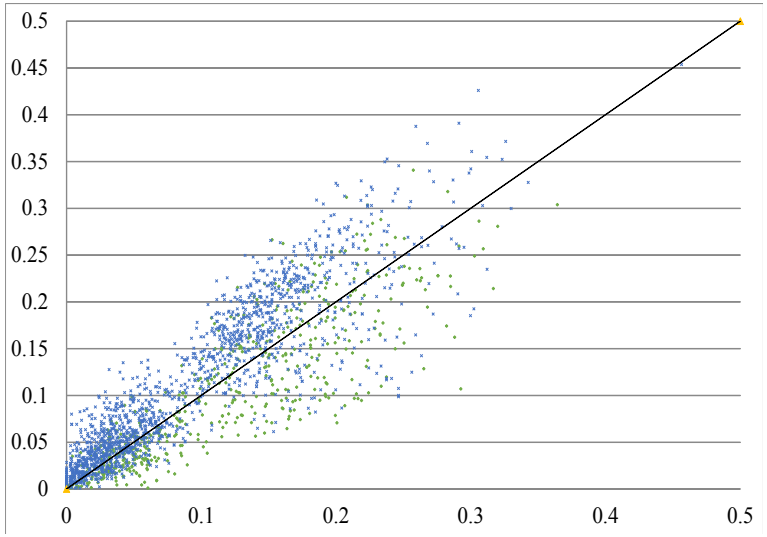
POS	***Accuracy (%)*** ***Positive / Negative / Overall***	
	Comparison	**Our method**
Adjective	80.15 / 27.45 / 45.37	**82.41** / **67.68** / **72.68**
Adverb	**37.09** / 71.16 / 58.26	28.73 / **80.70** / **61.01**
Noun	52.37 / 50.20 / 50.76	**65.47** / **75.39** / **72.83**
Verb	41.11 / 67.20 / 61.02	**64.58** / **81.68** / **77.62**
A^{μ}	59.90 / 48.21 / 51.68	**67.42** / **74.54** / **72.42**
A^{M}	52.68 / 54.01 / 53.34	**60.30** / **76.36** / **68.33**



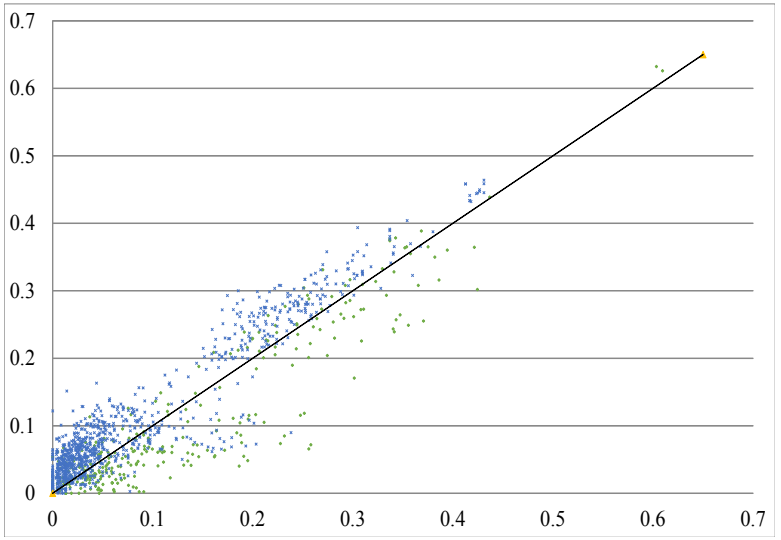
(a) adjective



(b) adverb



(c) noun



(d) verb

Reference

1. H. Schutze, “Automatic Word Sense Discrimination”, 1998.
2. S. Patwardhan, “Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts”, 2003.
3. B. Liu, M. Hu and J. Cheng, “Opinion Observer Analyzing and Comparing Opinions on the Web”, 2005.

Attempt on Chinese Opinion Words

- Same method has been applied to Chinese opinion words with the help of eHowNet.
- Each Chinese term in eHowNet has an associated English gloss, which can be used to generate gloss vector.
- Average accuracy: 70.95%