Transformation between Right and Left News Based on Text Similarity and News Structure

Yu-Wei Lin

Stanford University, yuweilin@stanford.edu

Abstract

The left–right political spectrum is a common way of classifying political issues along a one-dimensional political spectrum. The perspective of Left vs. Right is a binary interpretation of complex questions, but it also highly affects the way we look at the world, just like a filterin front our eyes, even causes some distirtions. From media, like news, can provide a good way to see how the different side of spectrum look at the same thing. it 's also fun to see how the opposite side present a simliar idea in different. So what I do is base on several known news and the news strucure, inverted pyramid, transfer a right news to left one, and vice versa.

1 Introduction

Traditionally, many translation/transformation methods in nlp are only based on words, because the structure is too sparsely in natural language to decide. But the information of structure can still be used in some specific types of articles. As we may notice, for news stories, the most popular structure is "inverted pyramid". In the inverted pyramid, the information is arranged in descending order of importance. The most important material is placed at the beginning of the story, and less important material follows. Succeeding paragraphs explain and support the lead. Since then, in this project, the transformation method will give highest score to the following chosed paragraph to reflect the phenomena of "inverted pyramid".

Also, it's quite similar to machine translation, except changing words to topic(I use paragraph in this project). This project is involved with differents topics of nlp, like Text summarization and segmentation, topic detection, sentence similarity, article generation. Mainly, I use WordNet[6], WordNet::Similarity[7], and Brown Corpus as my developing tools, which I'll go through them more detail in following.

The basic idea of this project is to transform a small, same topic from right to left(vice versa), by the similarity and the information of structure. The next section provides a short explanation of data structure in this program. Section 3 presents the method to measure text similarity. Section 4 gives the overall

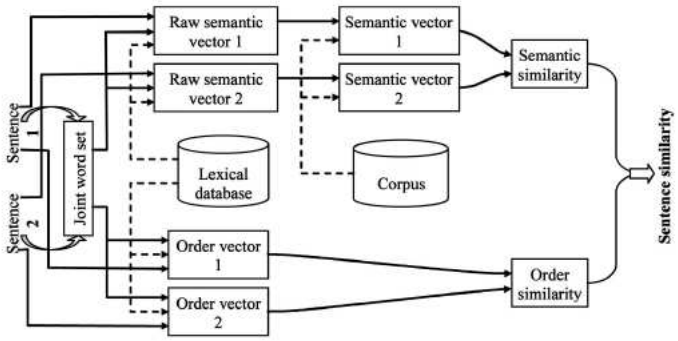

Figure 1:

transformation method. Section 5 briefly shows the experimental result. Section 6 talks the alternatives that I've tried and explains why they don't work. Finally, Section 7 summarizes the work, draws some conclusions, and proposes future related works.

2 Data Structure

The data structure in the program is very simple:

1. news data(list of news, containing the training data with storing right/left news in different list)
2. news(list of paragraphs, storing the information of news, like tile, position, provider, etc.)
3. paragraph(list of sentences)
4. sentece(list of words, containing the mapping between words and tags, but not used at the end)
5. words.

3 Text Similarity Method

Traditionally, techniques for detecting similarity between long texts (documents) have centered on analyzing shared words. Such methods are usually effective when dealing with long texts because similar long texts will usually contain a degree of co-occurring words. However, in short texts, word co-occurrence may be rare or even null. This is mainly due to the inherent flexibility of natural language enabling people to express similar meanings using quite different sentences in terms of structure and word content. Since such surface information in short texts is very limited, this problem poses a difficult computational challenge. The focus of this paper is on computing the similarity between very short texts, primarily of sentence length.

In this project, the method of text similarity baiscslly comes from the idea of . The method derives text similarity from semantic and syntactic information

contained in the compared texts. Fig. 1 [1]shows the procedure for computing the sentence similarity between two candidate sentences. In the following, I'll jsut give the equations, and the detail please look at the refernce.

3.1 Semantic Similarity between Words[8]

When cuculating the word similarity, I use the function in Word::Similarity: lin[], with the equation:

$$lin = 2 \times \frac{IC(lcs)}{IC(w1) + IC(w2)}$$

, where: lcs = the least common subsumer of w1 and w2; IC = the Information Content (of a synset).

$$wordSimilarity = lin \times I(w1) \times I(w2)$$

, where $I(w) = -\frac{\log p(w)}{\log N}$, $p(w)$ is the relative frequency of word w from statistics of Brown Corpus.

3.2 Sentence Similarity

3.2.1 Semantic Similarity[1]

for sentence $T1$ and $T2$, the word set $T = T1 \cup T2$, which means T contains unique wors in $T1$ and $T2$.

vector s_i^1 equals to the max word similarity of ith word in T with one word in $T1$, and s_i^2 is similar.

so the semantic similarity of senetnce

$$S_s = \frac{s^1 \bullet s^2}{\|s^1\| \, \|s^2\|}$$

for example, $T1 = \{I, am, a, person\}$, $T2 = \{She, is, happy\}$, then $T = \{i, am, a, person, she, is, haapy\}$

$$s^1 = [1, 1, 1, 1, w\,(she, i)\,, w\,(is, am)\,, w\,(happy, x)]$$

, x need to be determined by which pair with *happy* will have maximun word similarity.

3.2.2 Word order Similarity[1]

Very similar to semantic similarity, but with the elements be the location in each sentence.

$$S_r = 1 - \frac{\|r^1 - r^2\|}{\|r^1 + r^2\|}$$

for example, as same sentence pair above,

$$r^1 = [1, 2, 3, 4, 1, 2, y]$$

, where y is the index of x in $T1$

3.2.3 Overall Sentence Similarity

overall sentence similarity,

$$S = \delta S_s + (1 - \delta) S_r$$

,where $\delta = 0.85$ from [1]

3.3 Paragraph Similarity

The method of measuring paragraph similarity is derived by myself, so has not been verified by large amount of data to see if it really works,

$$S_p = \sum_i S\{S_1[i], S_2\left[ceil.\left(i \times \frac{\#ofSentenceInParagraph2}{\#ofSentenceInParagraph1}\right)\right]\}$$

,where(1)$S_1[i]$:ith sentence in $paragraph1$, (2)$paragraph1$ has more sentence than $paragrph2$

for this similarity, I have try three different ways to measure, the other two are: (1) Maximum, which means find the pairs that can form the maximum similarity.(2)Sum over all, which means sum up every pair. Both of above take too long time($O\left(n^2\right)$), and don't have better performance.

4 Transformation Method

This is also developed by myself, so, again, it's not verified by large data. The idea will be much more clear by the example:

$A = [a, b, c, d, e, ...]$; A: the news waiting to transfer, $a, b, c, ...$the paragraph of A in order.

$T = \{[x, y, z, ...], [j, k, m, ...]\}$; T: the training data, $[x, y, z, ...], [j, k, m, ...]$are different training news.

now if (a, x)has the max similarity in $(a, \alpha), \alpha \in T$,

when comparing the $(b, \beta), \beta \in \{T - a\}$,

$$similarity^*(b, y) = (1 + LINK) \times similarity(b, y)$$

,where $LINK$ is a parameter deciding the contribution of news structure,

$LINK = 0.2$ in this project.

5 Experimental Data & Result

5.1 Data

Manually collect 5 right-news and 5 left-news about the issue of "Government Shutdown", which is obvious that right-wings have huge difference from the left-wings.

5.2 Result

It can be seemed in the result that it still not preform really good. The system will repeatly pick up the same paragraph to transform. Also, the topic might be cut apart because the topic is included in two connecting paragraphs in training data, but they were seperate due to I choose one element per paragraph. For example:

> *Back in the the early '80s, when I served in the OMB under President Reagan, we went through several brief government shutdowns. Yes, the Washington Monument and a bunch of public parks closed. So what? Non-essential personnel got a holiday. The rest of us had to work.*
>
> *But non-essential programs were not funded during the shutdown, and their unused budgets were subsequently rescinded. Savings were significant.*

it's obvious by human that these two should group together.

6 Other Alternatives

6.1 First Order Markov Chain Model[2]

I've tried to use First Order Markov Chain Model as the method of transformation, but since Markov Chain Model works well in the situation that the "code"(the paragraph of news waits for transformation) only exists once, and so is the trraining data. Since then, the variance of transfer matrix in Markov Chain Model is too large to get a consistent model.

6.2 Segmentation by Topic

Using the simplified version of method in [3, 5], I tried to segment the small topics in one news, but since this method, Lexical Cohension, is for longer text, like news telecasts, and isn't quite suitable in short text segmentation.

6.3 Sentence Structure Similarity

For replacing Word Order Similarity by the structure of sentence, I tried a method by tagging part-of-speech[4] of each word, but the method didn't work well.

7 Conclusion&Future Works

There are still several steps of improvement to get a reliable system, like topic segmentation, similarity of topic, the relative role that "inverted pyramid" plays in news articles. Also, besides of the method and system, I believe after increasing the training data, the system will have a better performance.

I'm really imspired by doing the project. Thanks!

References

[1] Li, YH McLean, D Bandar, ZA O'Shea, JD Crockett, K, "Sentence similarity based on semantic nets and corpus statistics", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Vol.18 Issue: 8 Pages: 1138-1150, AUG 2006

[2] Diaconis P, "The Markov Chain Monte Carlo Revolution", BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY Volume: 46 Issue: 2 Pages: 179-205 Published: 2009

[3] Jacobs, A.; Ioannidis, G.T.; Christodoulakis, S.; Moumoutzis, N.; Georgoulakis, S.; Papachristoudis, Y., "Automatic, context-of-capture-based categorization, structure detection and segmentation of news telecasts", Digital Libraries: Research and Development. 1st International DELOS Conference, Revised Selected Papers. (Lecture Notes in Computer Science vol. 4877) Pages: 278-87 Published: 2007

[4] Kristina Toutanova and Christopher D. Manning. 2000. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000), pp. 63-70.

[5] Stokes N, Carthy J, Smeaton AF, "SeLeCT: a lexical cohesion based news story segmentation system", AI COMMUNICATIONS Volume: 17 Issue: 1 Pages: 3-12 Published: 2004

[6] MILLER GA, "WORDNET - A LEXICAL DATABASE FOR ENGLISH", COMMUNICATIONS OF THE ACM Volume: 38 Issue: 11 Pages: 39-41 Published: NOV 1995

[7] Pedersen, Patwardhan, and Michelizzi, "WordNet::Similarity - Measuring the Relatedness of Concepts", NAACL-04, pp. 38-41, May 3-5, 2004

[8] Dekang Lin, "An information-theoretic definition of similarity", Machine Learning. Proceedings of the Fifteenth International Conference (ICML'98) Pages: 296-304 Published: 1998