A Probabilistic Classification Approach for Lexical Textual Entailment

Oren Glickman, Ido Dagan and Moshe Koppel
Computer Science Department, Bar Ilan University

報告者:邱炫盛

Outline

- Introduction
- Background
- Probabilistic Textual Entailment
- A unsupervised Lexical Model
- Empirical Evaluation
- Conclusion and Future work

Introduction

- Many Natural Language Processing (NLP) applications need to recognize when the meaning of one text can be expressed by or inferred from another text.

- That is, they assess semantic relationship between text segments
  - IR, QA, IE, summarization , MT

- Textual Entailment Recognition has recently been proposed as an application independent task to capture such inferences

Introduction (cont.)

- Within the textual entailment framework:
  - A text *t* is said to entail a textual hypothesis *h* if the truth of *h* can be inferred from *t*
  - Example of QA:
    - *h*: ”Does John Speak French?”
    - t_1: “John is a fluent French Speaker” (strictly)
    - t_2:”John was born in France” (not strictly)

- It is clear that t_2 does increase substantially the likelihood that the hypothesized answer is true

Introduction (cont.)

- The uncertain nature of textual entailment calls for explicit modeling in probabilistic terms

- A general generative probabilistic setting for textual entailment is proposed

- It may be considered analogous to (though different than) the probabilistic setting defined for other phenomena, like language modeling, etc

Introduction (cont.)

- Lexical entailment
  - An important sub task of textual entailment
  - The concepts in hypothesis *h* are entailed from a given *t*
  - Relation may be not entailed
  - Necessary , but not sufficient

- Example
  - t: “Yahoo acquired Overture”
  - H concept: “Yahoo”, “acquisition”, “Overture”

Introduction (cont.)

- The general probabilistic generative model recasts the lexical entailment problem as a variant of text categorization
  - Class = content words (hypothesis concepts)
- Recognizing whether a hypothesis concept is entailed from the text t is carried out by classifying t to the corresponding class

Background

- A wide variety of semantic inference techniques were proposed
- Interpret text into a formal language on which inference is performed (Hobbs et al)
  - Computational cost and the lack of efficient tools
- Paraphrase and entailment rule acquisition (McKeown, et al)
  - Ignore the directionality aspect of entailment

Background (cont.)

- WordNet-based term expansion (lexical scope)
  - Most commonly used technique for enhancing the recall of NLP system
  - Weight scheme
  - Ranking
- Many approaches for modeling entailment were developed in application specific settings

Probabilistic Textual Entailment

- Common definition of entailment:
  - A text *t* entail another text *h* (hypothesis) if *h* is true in every circumstance (possible world) in which *t* is true
- Example:
  - t_1= "Marry is Fluent French Speaker"
  - t_2="Marry was born in France"
  - h_1="Marry Speaks French"
- Given the text, we expect the probability that the hypothesis is indeed true to significantly higher than its probability of being true without reading the text

Probabilistic Textual Entailment (cont.)

- Probabilistic setting
  - T: a space of possible texts
  - t: a specific text in T
  - H: the set of all possible hypotheses
  - h: a propositional statement which can be assign a truth value in H
  - w: H $\rightarrow$ {0,1} a semantic state of affairs captured by a mapping from H to {0,1} ,called possible world
  - W: the set of all possible worlds

Probabilistic Textual Entailment (cont.)

- Assume a probabilistic generative model for texts and possible worlds
  - Whenever the source generates a text *t*, it generates also corresponding hidden truth assignments that constitute a possible world

- Assume the distribution of truth assignments is not bound to reflect the state of affairs in any "real" worlds

Probabilistic Textual Entailment (cont.)

- The probability for generating a true hypothesis h that is not related at all to corresponding text is determined by some prior probability P(h)

- The notion of textual entailment is relevant only for hypotheses for which P(h) < 1, as otherwise there is no need to consider text that would support h's truth.

Probabilistic Textual Entailment (cont.)

- Two types of events over probability space for TxW:
  - $Tr_{h:}$ the random variable whose value is truth value assigned to h in the world of the generated text (Tr_h=1$\rightarrow$ h is true)
  - *t*: the event that generated text is *t*

- T probabilistically entails h if t increase the likelihood of h being true, i.e. $P(Tr_h=1 |t) > P(Tr_h=1)$
  - equivalently if the pointwise mutual information $I(Tr_h=1,t) >1$

An Unsupervised Lexical Model

- Modeling the full extent of the text entailment problem is a long term research goal
  - focus on sub-task of lexical entailment-identifying when the lexical elements of a textual hypothesis h are inferred from a given text *t*
- Assume that the meanings of the individual words in hypothesis h $=\{u_1,\ldots,u_m\}$ can be assigned truth values
  - For a given t, $Tr_{acquired}$ =1 if it can be inferred in t's state of affairs that acquisition even exists

An Unsupervised Lexical Model (cont.)

- A hypothesis is assumed to be true if and only if all its lexical components are true

- When estimating the entailment probability we assume that the truth probability of a term in a hypothesis *h* is independent of the truth of the other term in *h*

$$P(Tr_h = 1 \mid t) = \prod_{u \in h} P(Tr_u = 1 \mid t)$$

$$P(Tr_h = 1) = \prod_{u \in h} P(Tr_u = 1)$$

An Unsupervised Lexical Model (cont.)

Textual Entailments as Text Classification

- Main sleight-of-hand:
  - Estimate probabilities $P(Tr_u=1 |t)$ for text t and a lexical item *u* as text classification probabilities in which the classes are the different words *u* in the vocabulary
  - Use Naïve Bayes classifier for text classification (Nigam et al)
  - We classify texts to a binary abstract notion of lexical truth rather than to well-defined supervised classes

An Unsupervised Lexical Model (cont.)

Initial Labeling

- As an initial approximation, we assume that for any document in the corpus the truth value corresponding to a term u is determined by explicit presence or absence of u in that document

$$P(Tr_h = 1 \mid t) = 1 \ \textbf{if} \ u \in t \ \textbf{and} \ 0 \ \textbf{otherwise}$$

An Unsupervised Lexical Model (cont.)

Naïve Bayes Refinement

$$P(Tr_u = 1 \mid t) = \frac{P(t \mid Tr_u = 1)P(Tr_u = 1)}{P(t)}$$

$$= \frac{P(t \mid Tr_u = 1)P(Tr_u = 1)}{\sum_{c \in \{0,1\}} P(t \mid Tr_u = c)P(Tr_u = c)}$$

$$\approx \frac{P(Tr_u = 1)\prod_{i=1}^{|t|} P(t_i \mid Tr_u = 1)}{\sum_{c \in \{0,1\}} P(Tr_u = c)\prod_{i=1}^{|t|} P(t_i \mid Tr_u = c)}$$

An Unsupervised Lexical Model (cont.)

- Prior:

$$P(Tr_u = 1) = \frac{|\, d \in D : u \in d \,|}{|\, D \,|}$$

- Prior with Laplace smoothing:

$$P(v \mid Tr_u = 1) = \frac{1 + \sum_{d \in D: u \in d} N(v, d)}{|\, V \,| + \sum_{v' \in |V|} \sum_{d \in D: u \in d} N(v', d)}$$

Empirical Evaluation

Experimental Setup:

- Evaluate a textual entailment system in an application independent manner
  - QA, IR

- Corpus: Reuters Corpus Volume 1
  - 810,000 English News stories (economy related)

Empirical Evaluation (cont.)

- Classification results: acc: 70%(baseline:52%)

#	text	hypothesis	system	judge
1	Wall Street ended a stormy session with sharp losses Wednesday after the stock market surrendered a rally after news that the Federal Reserve was keeping interest rates unchanged.	federal reserve raise interest rates	1	0
2	The Libyan officials, accompanied by members of a Palestinian youth group, arrived at the camp in several small buses but it was not clear how they intended to transport the residents of the camp.	assemble truck	0	0
3	Earlier Tuesday CompuServe reported on its first quarter, warned of the expected second-quarter loss and said it expects improvements in the second half.	CompuServe predicted a loss	1	1
4	Fresh signs of labor market tightness emerged from the August employment report released Friday morning. The economy created 250,000 new jobs, while the unemployment rate fell to 5.1 percent, a six-and-a-half-year low.	cut jobs	1	0
5	Tokyo investors couldn't wait to get back to business on Tuesday after a long weekend, sending the key Nikkei stock average up by more than two percent, on the back of Wall Street's record climb in overnight trade.	Nikkei average rose	1	1
6	Shares in tobacco group Seita, privatised last year, slipped on Thursday in a quiet market under the impact of negative publicity concerning a second civil suit in a week by the family of smokers who died.	tobacco industry sued	0	1
7	Payne also asked whether Lloyd's was able to pay the total premium if U.S. investors did not back the recovery plan. Sandler again said he was unable to answer.	Sandler questioned	0	1
8	"We're starting to see some growth. We're seeing a recovery in fees and we can view that as a leading indicator of growth in the economy," said Susana Ornelas, banking analyst at Deutsche Morgan Grenfell in Mexico City.	Mexico’s economy is recovering	1	1

Table 1: text-hypothesis examples along with classification results and annotation

Empirical Evaluation (cont.)

Ranking results

- Cws: confidence weighted score (average precision)

$$cws = \frac{1}{N}\sum_{i=1}^{N}\frac{\#\,\textbf{correct up to rank}\ i}{i}$$

	min	rand	idf	model	max
cws	0.35	0.49	0.51	0.54	0.63

Table 2: results of confidence weighted score

Empirical Evaluation (cont.)

Analysis

- Our model make good topical distinctions

- EM made no significant improvement in further iteration

- Comparable results were achieved by our model when using 1/8 of the corpus

Conclusions and Future work

- This paper presented a generative probabilistic setting for textual entailment

- Then a concrete model at the lexical level was proposed
  - Can be practically approached within probabilistic framework

- We propose this framework can be utilized further to gradually address the additional aspects of the textual entailment phenomenon