Learning Commonsense Knowledge Models for Semantic Analytics

Conference Paper · February 2016

DOI: 10.1109/ICSC.2016.12

2 authors:

Rajaraman Kanagasabai
Agency for Science, Technology and Research (A*STAR)
92 PUBLICATIONS **960** CITATIONS
SEE PROFILE

Shangfeng Hu
Institute for Infocomm Research
3 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Learning Commonsense Knowledge Models for Semantic Analytics

Hu Shangfeng, and Rajaraman Kanagasabai
Data Analytics Department, Institute for Infocomm Research, A*STAR, Singapore
{hus, kanagasa}@i2r.a-star.edu.sg

***Abstract*— Commonsense knowledge is considered the basis for machine understanding of natural language semantics. In this paper, we propose a new method for fine-grained commonsense knowledge extraction using dependency graphs built through an unsupervised method. We implement the method on a large-scale text collection, and show how our knowledge model can be applied to develop a generic pronoun resolution algorithm. In particular, we demonstrate that our method leads to significant improvements in identifying false co-referred mention pairs, and hence could boost the precision of even the state of the art methods. Our method is unsupervised, and thus the performance can be further improved with bigger collections of raw texts.**

Keywords- Commonsens knowledge; Semantic modeling; Unsupervised learning; Knowledge representation;

I. INTRODUCTION

Commonsense knowledge refers to the basic facts and understanding possessed by most people, and is often regarded as the basis for human beings to understand natural language. The lack of commonsense knowledge has long been a bottleneck for understanding the semantics of natural language. Traditionally commonsense knowledge has been collected manually e.g. WordNet and Cyc, but more recently there have been efforts into automated data-driven methods for commonsense knowledge acquisition. ConceptNet [7] is a semantic network of commonsense knowledge built from English sentences of the Open Mind Common Sense (OMCS) corpus, through an automatic process. Similar efforts include Gordon et al.,[4], Speer and Havasi, [11] and Akbik and Michael [1].

Most of these methods focus on extraction of facts from text by analyzing relationship between words within a context window. Such an approach is useful to build a fact-based knowledge base, but it is not clear how it can benefit advanced semantic analytics tasks such as co-reference resolution. In this paper, we aim to construct a more fine-grained commonsense knowledge model with advanced semantic analytics in mind. Key to our approach is the extraction of dependency graphs involving terms and entities potentially spanning sentence and document boundaries. Our method is largely unsupervised and has the lifelong-learning capability as more and more texts are input. We implement our method on the Gutenberg [5] English dataset collection (about 10GB size) and Wikipedia English dataset (about 10GB size) to build a large scale RDF knowledgebase of about 30 billion triples. We show how our knowledge model can be applied to develop a generic pronoun resolution algorithm. In particular, we demonstrate that our method leads to significant improvements in identifying false co-referred mention pairs, and could boost the precision of even the state of the art methods.

II. COMMONSENSE KNOWLEDGE EXTRACTION METHOD

Below we describe the extraction of commonsense knowledge from raw texts. First we define our knowledge representation model.

2.1 Knowledge Representation Model

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

2.2 Commonsense Knowledge Extraction

First we parse a document to extract the sentences and extract dependency graphs using a dependency parser, e.g. Stanford Parser. As defined above, the graph includes word nodes and dependency relation. For each word node $n_i \in N$, $\mathrm{lemma}(n_i)$ is defined as the lemma of the word, and $\mathrm{pos}(n_i)$ is defined as the part of speech (POS) of the word. This is repeated for each document until all documents are exhausted. This process generates a set of document dependency graphs, called the Commonsense Knowledge Network (CKN).

2.3 Mutual Probability

Let G_s be the dependency graph of a given statement (for example: [*] *is scared*, where [*] can be any noun). G_c is a dependency graph of a given concept (for example: *cat*). Let G_t be the dependency graph when both statement s and concept c are satisfied together (for example: *cat is scared*). Let $\boldsymbol{D}_s$ be the set of document dependency graphs D_i^G such that $G_s \in D_i^G$. (The latter is performed through graph matching illustrated later in this section.) Similarly, let $\boldsymbol{D}_c$ be the set of document dependency graphs D_i^G such that $G_c \in D_i^G$. We use $\boldsymbol{D}_{s+c}$ to denote the document dependency graphs D_i^G that $G_s \in D_i^G$ and $G_c \in D_i^G$ Also, $\boldsymbol{D}_t$ is the document dependency graph set that includes all D_i^G such that $G_t \in D_i^G$. Figure 1 provides an illustration.

Then we can define the following probabilities:

$$P(s) = \frac{|(\boldsymbol{D}_s)|}{|(CKN)|} \quad P(c) = \frac{|(\boldsymbol{D}_c)|}{|(CKN)|}$$
$$P(s,c) = P(t) = \frac{|(\boldsymbol{D}_t)|}{|(CKN)|}$$
$$P(s+c) = \frac{|(\boldsymbol{D}_{s+c})|}{|(CKN)|}$$

where the |.| operator stands for set cardinality.

We can see that, $P(s+c)$ referes to the joint probability of statement s and the concept c occurring together. If the probabilities of the statement s and the concept c are independent, then $(s+c) = P(s) * P(c)$. The *kernel formula* $\frac{P(s,c)}{P(s+c)}$ can then be simplified as $\frac{P(s,c)}{P(s)*P(c)}$, and computed easily. This is the approach used by Church and Hanks [2]; Gordon et al., [4]; Akbik and Micheal, [1] to compute the key measure PMI. However, the independence assumption may be too limiting in real life natural language. In our work, we relax this assumption and compute $\frac{P(s,c)}{P(s+c)}$ directly. We argue that this helps to capture the semantics of natural language more precisely.

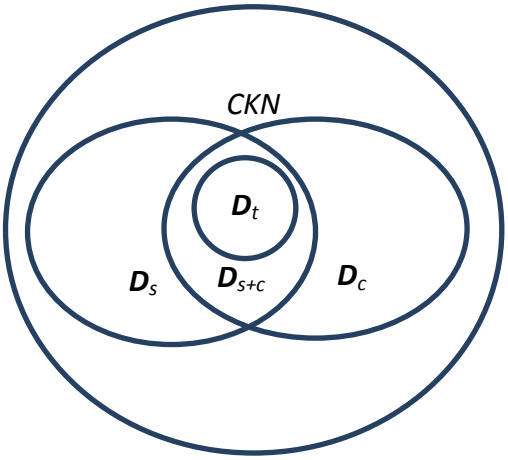

Figure 1. CKN and Mutual Probability

In our model, a commonsense knowledge term includes a statement (for example: *cat runs*), the possibility of the statement being true, and its corresponding confidence evaluation value.

Given a statement s (for example: *cat runs*), a noun n (for example: *cat*) is a word such that $pos(n)$ is *NN* or *NNS*. The dependency graph of s is defined as $G = \{ n_0, n_1, \ldots n_n \mid r_0, r_1, \ldots r_m \}$, where n_i is a node and r_j is a dependency relation. A graph pattern G' is generated for a node $n_0 \in G$, (for example, n_0 is *cat* in the example statement), by replacing n_0 of G with a variable node v_0 that $pos(v_0) = pos(n_0)$ and $lemma(v_0) = [*]$ which is a wildcard. Another graph pattern G'' is generated from G' by replacing v_0 with the given word n.

The pattern G' is represented as a graph in Figure 2, and the pattern G'' is represented as a graph in Figure 3.

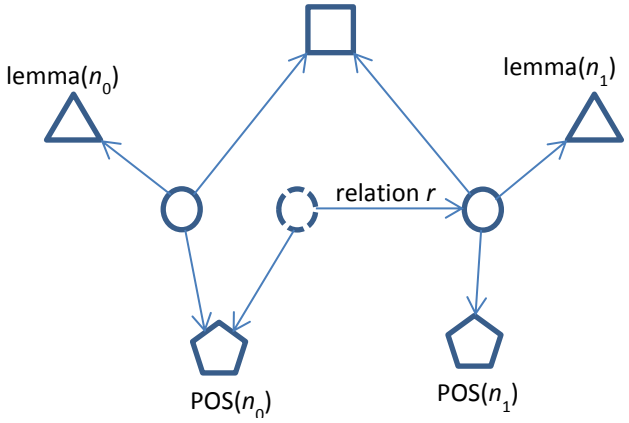

Figure 2. The simplified graph pattern G'

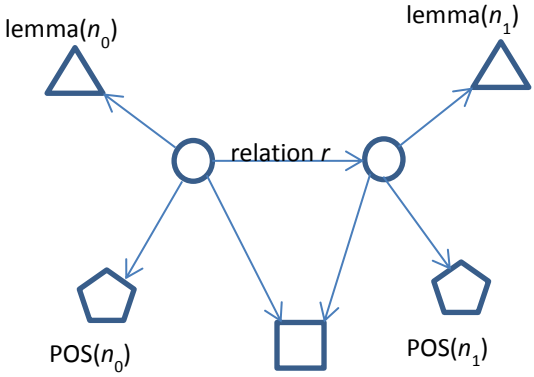

Figure 3. The simplified graph pattern G''

Given graph patterns G' and G'', let $\boldsymbol{D}'$ and $\boldsymbol{D}''$ *be graph sets* constructed by including all D' such that $G_i \in D'$ if G' is a sub-graph of G_i, $G_i \in D''$ if G'' is a sub-graph of G_i. Then, Mutual Probability (MP) $P(t)$ is defined for that n satisfies the variable node $v_0 \in G'$ as

$$\mathrm{P}(t) = \frac{|(\boldsymbol{D}'')|}{|(\boldsymbol{D}')|}$$

The probability $P(t)$ can be used to evaluate the semantic consistency of a pair of mentions in co-reference resolution as below.

For example, consider the statement *A cat is playing a ball. It is scared by the sharp noise.* Here it is not obvious which noun is co-referred to the pronoun *it*. The given word n can be *cat* or *ball*. The second sentence is parsed into a dependency graph { *nsubjpass (scare-VBN , it-PRP), auxpass (scare-VBN , is-VBZ), det (noise-NN , the-DT), amod (noise-NN , sharp-JJ), agent (scare-VBN , noise-NN)* }. We can select a small subgraph containing the pronoun to be resolved to generate the graph patterns. For example, we select G {*scare-VBN, it-PRP | nsubjpass (scare-VBN , it-PRP)*} at first. G'_1= *{cat-NN}*∪ *{ *-NN, scare-VBN| nsubjpass (scare-VBN, *-NN)*}, G'_2= *{ball-NN}*∪ *{ *-NN, scare-VBN| nsubjpass (scare-VBN, *-NN)*}, G''_1= *{cat-NN, scare-VBN | nsubjpass (scare-*

VBN, cat-NN)}, G_2'' = {*ball-NN, scare-VBN | nsubjpass (scare-VBN, ball-NN)*}. Correspondingly, we get D_1' $\boldsymbol{D}_2'$, $\boldsymbol{D}_1''$, and $\boldsymbol{D}''_2$. Then we can get two probabilities as follows.

$$\mathrm{P}(t_1) = \frac{|(\boldsymbol{D}_1'')|}{|(\boldsymbol{D}_1')|}$$

$$\mathrm{P}(t_2) = \frac{|(\boldsymbol{D}_2'')|}{|(\boldsymbol{D}_2')|}$$

Comparing above two probabilities, we find $P(t_1) >> P(t_2)$. Then *it* should be co-referred to *cat*. Table 1 shows some examples of Mutual Probability for some statements based on our experimental data.

Statement	Mutual Probability
[*] is scared	[cat] 8.33% [ball] 0%
[*] runs	[man] 14.21% [stone] 1.11%
[*] teaches	[teacher] 26.99% [student] 4.05%
cup on [*]	[table] 72.09% [chair] 10% [bed] 0%

Table 1. Mutual Probability Examples

We propose Mutual Probability to replace Mutual Information formula [1] [2][3] [4].

III. Application to Pronoun Resolution

Pronoun resolution is the task of finding the co-referred words/phrases for a given pronoun in texts. It is an important and challenging subtask of the more generic co-reference resolution which involves the identification of different types of anaphors that may occur anywhere in the text entity [8]. Resolution of co-referred pronouns is important in many NLP applications such as Discourse analysis, Text Summarization, and Information Retrieval, etc.

Several methods for co-reference resolution have been proposed in the literature. Besides rule-based methods, machine learning approaches have had prominence and were reported to result in better performance in various studies [10]. Most of these machine learning methods are supervised, and so require labelled training data. This is often expensive, so these methods are limited in practical applications. This has motivated unsupervised learning methods for co-reference resolution. One of the best-performing methods on CoNLL2011, StanfordCoreNLP [6] does not require labeled data and primarily uses hand-crafted rules [9].

Here we demonstrate how our commonsense knowledge extraction method can be applied to improve any existing pronoun resolution method.

Figure 4 shows the overview workflow of our method. The main idea is to make use of contextual knowledge built from large unlabeled texts and identify the co-referring mention pairs accurately. First, we build a commonsense knowledge network from a huge raw text corpus. Then we extract a lot of commonsense knowledge terms with probabilities.

For a given text, we use a baseline co-reference resolution method to find co-referred mention pairs. Finally, for a given mention pair of pronoun resolution (example: [*cat*], [*it*] is scared), we evaluate the possibility based on the commonsense knowledge terms to find mistakes.

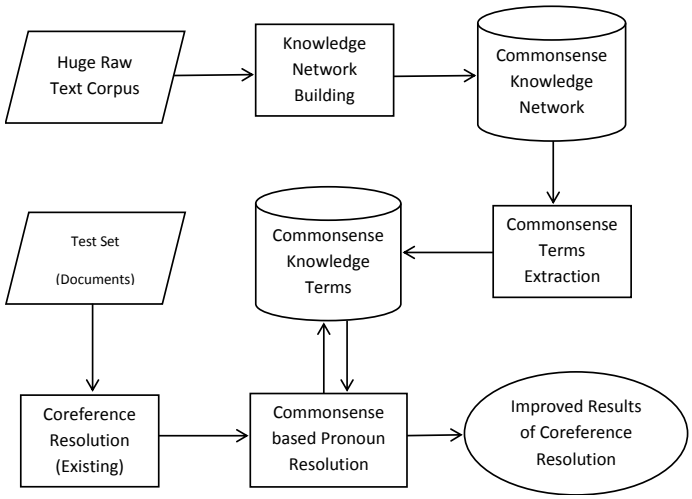

Figure 4. Overall workflow of our method

We employ the commonsense knowledge to improve pronoun resolution as follows. It can be noted that, in co-reference resolution results, a pair of mentions which refer to a same entity is co-referred to each other. The key idea is that, by virtue of the fact that the Mutual Probability evaluates the semantic reasonability of a term, if the mentions are least likely to be referring to a same entity by commonsense knowledge, then the co-referred mention pair should be a false positive, and can be removed. This is formalized as below.

For each mention pair (m, n) that m is a pronoun and n is a noun, we define a statement s, a concept c and a term t. We also define the probabilities $P(t) = P(s, c)$ and $P(s + c)$ in the same way.

We consider the mention pair (m, n) is semantically "false" if $(P(s, c) = 0\ and\ P(s + c) > h_1)$ or $(P(s, c) > 0\ and\ P(s, c)/P(s + c) < h_2)$, h_1 and h_2 are two thresholds. We define an integrated threshold as $h > 0$, $h_1 = h$, and $h_2 = 1/(h * 2)$.

IV. Experiments and Evaluation

We consider the task of finding "false" co-referred mention pairs of StanfordCoreNLP [6] pronoun resolution results for CoNLL2012 test data set. At first, we run StanfordCoreNLP on CoNLL2012 test data set and select mention pairs which are believed co-referred in the results. The selected mention pairs satisfy following two conditions:

The mention pair includes a noun mention and a pronoun mention and the noun mention is not a part of a named entity.

The noun mention is in front of the pronoun mention in the document. Upon closer investigation, we find 2098 mention pairs which satisfy the conditions from 14947 co-referred mention pairs. Among these 2098 co-referred mention pairs, there are 1203 mistakes (false positives). In other words, more than 57% of the returned mention pairs are mistakes compared to the gold standard.

Our idea is to make use of contextual knowledge built from large unlabeled texts using the method outlined in Section 3 and identify the false co-referring mention pairs. We used Wikipedia (~10GB size) and Gutenberg [5] English collection (~10GB size) to train set our commonsense knowledge model. The knowledge model learn from this collection is stored as an RDF using AllegroGraph4 triple store. Then all probability computations are performed by suitably composing SPARQL queries.

Let m_0 be the "false" mention pairs returned by our method, and m_1 of them are mistakes compare to the gold standard. Let m be the "false" mention pairs returned by StanfordCoreNLP. We study the performance by defining two measures viz. $precision = m_1/m_0, recall = m_1/m.$, We list the results in Table 2, for different threshold settings h.

From Table 2, we can see that our method can identify as many as 22% of the false positives at more than 75% precision, wih 2 billion word training set. The precision can be improved further by increasing the threshold, with a corresponding tradeoff in recall. This shows that the confidence evaluation based on $P(s + c)$ is direct related to the accuracy of the commonsense knowledge.

Furthermore, as the size of training data increases, the recall (or, the number of "false" co-referred mention pairs) improves with a comparable stable precision level increases significantly. Thus, by using more raw texts as input to the system, the performance can be further improved.

Threshold *h*	Performance Measure	Training data size (in billion words)		
		0.5	1	2
5	Precision	75.28%	73.44%	75.56%
	Recall	16.71%	19.53%	22.36%
10	Precision	83.33%	82.10%	83.33%
	Recall	9.14%	11.06%	13.72%
15	Precision	86.58%	86.60%	85.60%
	Recall	5.90%	6.98%	8.89%
20	Precision	87.10%	86.96%	86.00%
	Recall	4.49%	4.99%	7.15%
25	Precision	87.23%	87.30%	90.41%
	Recall	3.41%	4.57%	5.49%

Table 2. Performance results on identifying "false" mention pairs in the pronoun resolution task using CoNLL2012 test data.

V. Conclusion and Discussion

Commonsense knowledge is often considered the basis for machine understanding of natural language semantics. In this paper, we proposed a new commonsense knowledge model and developed a method for learning it using dependency graphs built through an unsupervised algorithm. Key to our method is the notion of Mutual Probability that captures the probabilistic commonsense knowledge extracted from raw texts. We implemented the method on a large-scale text collection, and showed how our knowledge model can be applied to develop a generic pronoun resolution algorithm. In particular, we demonstrated that our method leads to significant improvements in identifying false co-referred mention pairs, and hence could boost the precision of even the state of the art methods.

The method we proposed in this paper is unsupervised. Our results show that with more data, the recall can be improved with a comparable stable precision level. We believe the performance of our method can be improved further with a bigger data set. Also it can be noted that our knowledge model is generic and can be applied for other semantic analytics tasks such as paraphrase detection, ontology learning and sentiment analysis. Currently we are investigating this research direction.

References

[1] Alan Akbik, Thilo Michael: The Weltmodell: A Data-Driven Commonsense Knowledge Base. LREC 2014

[2] Kenneth Ward Church, Patrick Hanks: Word Associa-tion Norms, Mutual Information, and Lexicogra-phy. Computational Linguistics 1990

[3] Robert M. Fano : Transmission of Information: A Statistical Theory of Communications. MIT Press, Cambridge, MA. 1961

[4] Jonathan Gordon, Benjamin Van Durme, Lenhart K. Schubert: Learning from the Web: Extracting Gen-eral World Knowledge from Noisy Text. Collabo-ratively-Built Knowledge Sources and AI 2010

[5] Project Gutenberg Literary Archive Foundation. Pro-ject Gutenberg. http://www.gutenberg.org/ (checked 2009-08-03), 2009.

[6] Heeyoung Lee, Yves Peirsman, Angel Chang, Na-thanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In Proceedings of the CoNLL 2011

[7] Hugo Liu , Push Singh: ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal, To Appear. Volume 22, forthcoming issue. Kluwer Academic Publishers. 2004

[8] Sebastian Martschat: Multigraph Clustering for Unsu-pervised Coreference Resolution. In Proceedings of ACL (Student Research Workshop) 2013

[9] Sameer Pradhan , Lance Ramshaw , Mitchell Marcus , Martha Palmer , Ralph Weischedel , Nianwen Xue,: modeling unrestricted coreference in Onto-Notes, Proceedings of CoNLL 2011

[10] Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, Yuchen Zhang: CoNLL-2012 shared task: Modeling multilingual unrestricted co-reference in OntoNotes - on EMNLP and CoNLL 2012

[11] Robert Speer and Catherine Havasi : Representing gen-eral relational knowledge in conceptnet 5. In LREC 2012