Combining Open IE and Distant Supervision for KBP Slot Filling

Stephen Soderland, Natalie Hawkins, John Gilmer, and Daniel S. Weld
Turing Center, Department of Computer Science and Engineering
Box 352350
University of Washington
Seattle, WA 98195, USA
{soderlan, nhawkins, jgilme1, weld}@cs.washington.edu

Abstract

The University of Washington participated in English Slot Filling for TAC-KBP 2014 with a system that combines its 2013 OPENIE-KBP system (Soderland et al., 2013) with the MULTIR extractor (Hoffmann et al., 2011), which uses distant supervision from Freebase. Since OPENIE-KBP is identical to our KBP 2013 system, this lets us quantify how this year's queries are more difficult than last year's.

OPENIE-KBP gives high precision at moderate recall – we analyze the limitations of recall to an extractor based on Open IE. Our MULTIR-KBP extractor added to the recall in a combined system, but lowered the precision. We analyze the reasons for low precision with distant supervision.

1 Overview

This year, the University of Washington participated in the English Slot Filling evaluation with a combination of two systems: OPENIE-KBP is based on Open Information Extraction (Soderland et al., 2013; Mausam et al., 2012) and MULTIR-KBP is based on MULTIR (Hoffmann et al., 2011), which uses distant supervision. Combining the two gave good synergy that increased recall. Since OPENIE-KBP is identical to our system from 2013, we are able to quantify the increased difficulty of this year's queries over 2013.

We find that applying relation-specific rules to Open IE tuples produces a high precision extractor, with precision comparable to the hand generated extractions by LDC annotators. However, there are fundamental limitations to recall. On 2013 queries, OPENIE-KBP had recall about the median of all participants – this year the same system had recall considerably less than half of the median.

In Section 2.1 we discuss some reasons for the limited recall. Open IE can only identify relations where there is an explicit relation phrase in the sentence. For example, "Jean DuPuis is a journalist at Le Monde" has a relation phrase "is a journalist at". In contrast "French journalist Jean DuPuis reported that ..." has an implicit relation *hasJobTitle* but no explicit relation phrase. Open IE will only find tuples with the relation phrase "reported", which is not useful in identifying per:title relations. A large proportion of the correct extractions by KBP participants are from sentences where there is no explicit relation phrase, and thus beyond the reach of Open IE.

Our MULTIR-KBP system extends the recall of OPENIE-KBP, but has much lower precision. MULTIR is trained on distant supervision from FreeBase, which gives it extremely noisy training for most relations. Section 3.1 discusses how this impacts the system precision.

2 Mapping Open IE to a Target Ontology

OPENIE-KBP begins by running an Open Information Extractor over the TAC-KBP corpus, which produces tuples of the form *(arg1, rel, arg2)* where *rel* is a phrase from the input sentence that expresses an arbitrary relation between *arg1* and *arg2*.

Our first Open IE system was TextRunner (Etzioni et al., 2006; Banko et al., 2007; Banko and Etzioni, 2008), followed by ReVerb (Fader et al., 2011; Etzioni et al., 2011) and OLLIE (Mausam et

Open IE tuples	KBP relations
(Steve Jobs, **died of**, cancer) (Steve Jobs, **succumbed to**, cancer) (Steve Jobs, **lost his battle** to, cancer)	per:cause_of_death
(Nasrallah, **is leader of**, Hezbollah) (Hezbollah, **headed by**, Nasrallah) (Nasrallah, **is Secretary-General of**, Hezbollah)	org:top_members _employees

Figure 1: Open IE finds textual relations with no tuning required for a domain or set of target relations. The challenge is to map these extractions to relations in an ontology.

Input sentence:

"Steve Jobs, the co-founder of Apple, died of cancer in his Palo Alto home."

Open IE tuples:

1. (Steve Jobs, died of, cancer)
2. (Steve Jobs, died in, his Palo Alto home)
3. (Steve Jobs, is co-founder of, Apple)

Figure 2: Open IE tuples from a sample sentence. OPENIE 4.0 is more robust in identifying verb-based relations, but also handles noun-based relations such as "(is) co-founder".

al., 2012). The most recent Open IE v4.0[1] handles both verb-mediated relations (e.g. "died at","lost his battle to") and noun-mediated relations (e.g. "is co-founder of", "is leader of"). However, these extractions express relations textually as shown in Figure 1.

An advantage of Open IE over previous information extraction systems is that it works out of the box, requiring no training or tuning for a new domain. The relations it extracts are represented as text strings rather than as relations in an ontology. This is not a problem if the tuples are for human use, for example searching a database of Open IE tuples extracted from a text corpus.

However, some applications require the relations to be mapped to the relations in a particular ontology. Figure 1 shows just a few of the textual relations that correspond to *per:cause_of_death* or *org:top_members_employees*. In general, there are a few high frequency surface forms used to express a relation such as "died of" or "died from", and a long tail of other surface forms with diminishing frequency.

It is this Zipfian distribution of surface forms that gives us the possibility to create a mapping from target relations in an ontology to Open IE tuples with minimal knowledge engineering effort. A simple rule language built on Open IE is sufficient to identify the most common surface forms with high precision.

Figure 1 illustrates several Open IE extractions. The first tuple (Steve Jobs, died of, cancer) is one of the extractions from "Steve Jobs, the co-founder of Apple, died of cancer in his Palo Alto home." Other tuples from this sentence are shown in Figure 2.

We evaluated the results on the 2014 queries and found that they achieved high extraction precision, as shown in Table 2. The number of extractions was only 83, compared to 239 for the same system run on 2013 queries. This seems to be primarily from a qualitative difference in queries between the two years.

2.1 Limits to Open IE recall

We analyzed the correct extractions from from our OPENIE-KBP and from all runs submitted by KBP participants. As Table 1 shows, most of the correct OPENIE-KBP extractions were from noun-based constructions, either appositives or slot fills that were noun modifiers to the entity.

Correct slot fills in responses from all KBP participants shows a similar trend. A large proportion are found in noun phrases, often with no explicit relation phrase to create an Open IE tuple. This was particularly true of per:origin and per:title relations, two of the most common slot fills. We examined all correct per:origin and a sample of 100 of the per:title slot fills. Only 9% of the per:title slot fills were

Table 1: Only a small percentage of the correct KBP extractions from OPENIE-KBP were from verb-based relation phrases. The great majority were from noun-based patterns.

Syntactic structure	percent
apppositive	0.38
noun modifier	0.26
verb phrase	0.26
other	0.09

[1] Available at github.com/knowitall/openie

in a context that had a verb predictive of the relation (*e.g.* "worked as" or "served as"); 29% were in a light verb construction (*e.g.* "was" or "became"); and 62% had the slot fill in the same NP as the entity. For per:origin, none were in a context with a verb that indicated nationality; 9% were found in light verb constructions; and 91% were in the same NP as the entity.

Another limit of Open IE is that it forms tuples only for *binary* relations, where there is both an Arg1 and Arg2 for the relation phrase. Consider the example given earlier, "Jean DuPuis is a journalist at Le Monde" and a noun-based variant "Jean DuPuis, a journalist at Le Monde, reported that ...". Each of these produces the same tuple, (Jean DuPuis, is a journalist at, Le Monde).

In many cases, however, a sentence expresses an attribute of an entity, but there is no Arg2. Take for example "French journalist Jean DuPuis reported that ...". There is no second argument for a "journalist" relation – we don't know a place, date, or newspaper name to serve as Arg2. What we would like is a tuple with an *implicit relation* such as "has job title": (Jean DuPuis, [has job title], journalist). Such implicit relations, with no relation phrase in the sentence, is beyond the scope of current Open IE systems.

2.2 Comparison of 2013 and 2014 Queries

Using an identical OPENIE-KBP system for both 2013 and 2014 lets us make a fair comparison of the increased difficulty of 2014 queries. In 2013 9% of the query entities had fewer than 10 Open IE tuples in the corpus compared to 31% for 2014. Our linker found a Freebase entity for 36% of the 2013 query entities, but only 9% of the 2014 entities. This points to a qualitative difference between the two sets of queries – 2014 had more obscure entities with a smaller presence in the KBP corpus and more difficult to link to Freebase entities.

3 MultiR System

We adapted our MULTIR distant supervision system to the KBP Slot Filling task. MULTIR's training phase begins with a set of target relations, finds entity pairs that have that relation in Freebase, and then finds sentences in the KBP corpus that contain one of the entity pairs. We run an NER tagger and partition the training by argument types, such as PER-LOC, and learn a model for each partition. At test time, MULTIR identifies candidate argument pairs in a sentence and applies one of the extraction models. In our KBP system, we find the set of documents that contain a mention of a query entity e_q and run the MULTIR extractor on those documents. We then filter the extractions to those where *arg1* has e_q or is in the coref set with e_q, using the Stanford NLP pipeline.

Our original MULTIR had been developed with the assumption that both arguments could be identified by an NER tagger, an assumption that needs to be relaxed to handle slot fills with dates, numbers, and common nouns such as job title and cause of death. Due to staff turnover shortly before the KBP evaluation, this was not fully implemented, so our output includes only relations with PER, LOC, and ORG as slot fills.

We had hoped for higher recall than OPENIE-KBP, but MULTIR-KBP produced about the same number of slot fills, but at lower precision. This was partly due to the incomplete implementation, and partly from inherent limitations to distant supervision.

3.1 Limits to Distant Supervision Precision

The earliest system in the spirit of distant supervision was DIPRE (Brin, 1998). The DIPRE paper gives an example of learning a *date-of-birth* relation from sentences or other Web text containing the seed (Mozart, 1756). This is an ideal case, since Mozart was doing nothing of note that year except being born. This is not the case in general. Consider learning the relation *place-of-birth* from seeds such as (Nicolas Sarkozy, Paris). The French president was indeed born in Paris, but hardly any sentences from news text that mention both entities are about his birth there.

We found that distant supervision for KBP relations was overwhelmingly false positives. Only 16% of a sample of supposedly positive training instances actually expressed the target relation. The preponderance of false positive training overwhelms our classifier and results in low precision. This is despite MULTIR using bags of training instances for an entity pair and using a probabilistic graphical model to tease apart the true positive from false positive training. Our internal evaluation of MULTIR-KBP

Table 2: Results with official recall and precision and extraction precision (whether the slot fill is correct with respect to the sentence).

System	Recall	Official precision	Extraction precision
Run 1: Combined	0.077	0.51	0.64
Run 2: Open IE	0.060	0.72	0.89
Run 3: MultiR	0.025	0.30	0.40

shows extraction precision starting at 0.82 at low recall and declining to 0.40.

4 Combined Systems

We ran both the Open IE and MULTIR systems separately, then took the union of the results as our main submission. If the two systems had different output for a particular slot fill for a query, we included the Open IE output and dropped that of MULTIR.

There was surprisingly little overlap of responses by the two systems. Open IE had 64 correct responses not found by MULTIR; MULTIR had 26 correct responses not found by Open IE; and there were 8 correct responses that were found by both systems. Of the extractions that Open IE found that MULTIR missed, 38% of them were for relations with date, number, title, or cause of death where MULTIR was not fully implemented.

Our official results are shown in Table 2 along with our own informal evaluation, shown as extraction precision. We tagged an extraction as correct if it is correct with respect to the sentence. The official precision also takes into account whether the entity in the sentence is the query entity and not just another entity with the same name. Official precision also takes into account whether the slot fill is the most complete phrase found in the corpus, labeling the less complete slot fills as inexact rather than correct.

5 Conclusions

We participated in the 2014 KBP English Slot Filling evaluation with a combination of two systems, one based on Open IE and the other trained with distant supervsion. The OPENIE-KBP system had high precision, but limited recall. We discuss some fundamental limitations to recall in using Open IE for KBP Slot Filling – a large portion of the slot fills are found in the same NP as the entity, often with no explicit relation phrase. Our MULTIR-KBP system, in contrast, had low precision. This seems to be a fundamental limitation of distant supervision due to the preponderance of false positive training instances.

6 Acknowledgements

This research was supported in part by ONR grant N00014-11-1-0294, DARPA contract FA8750-13-2-0019, and ARO grant W911NF-13-1-0246, and was carried out at the University of Washington's Turing Center.

References

M. Banko and O. Etzioni. 2008. The tradeoffs between traditional and open relation extraction. In *Proceedings of ACL*.

M. Banko, M. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. 2007. Open information extraction from the Web. In *Procs. of IJCAI*.

S. Brin. 1998. Extracting Patterns and Relations from the World Wide Web. In *WebDB Workshop at 6th International Conference on Extending Database Technology, EDBT'98*, pages 172–183, Valencia, Spain.

O. Etzioni, M. Banko, and M. Cafarella. 2006. Machine Reading. In *AAAI*.

Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam. 2011. Open information extraction: the second generation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '11)*.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of EMNLP*.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke S. Zettlemoyer, and Daniel S. Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *ACL*, pages 541–550.

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of EMNLP*.

Stephen Soderland, John Gilmer, Robert Bart, Oren Etzioni, and Daniel S. Weld. 2013. Open information extraction to KBP relations in 3 hours. In *Proceedings of TAC-KBP 2013*.