

Discovering Implicit Knowledge with Unary Relations

Michael Glass

IBM Research AI
Knowledge Induction and Reasoning
mrglass@us.ibm.com

Alfio Gliozzo

IBM Research AI
Knowledge Induction and Reasoning
gliozzo@us.ibm.com

Abstract

State-of-the-art relation extraction approaches are only able to recognize relationships between mentions of entity arguments stated explicitly in the text and typically localized to the same sentence. However, the vast majority of relations are either implicit or not sententially localized. This is a major problem for Knowledge Base Population, severely limiting recall. In this paper we propose a new methodology to identify relations between two entities, consisting of detecting a very large number of unary relations, and using them to infer missing entities. We describe a deep learning architecture able to learn thousands of such relations very efficiently by using a common deep learning based representation. Our approach largely outperforms state of the art relation extraction technology on a newly introduced web scale knowledge base population benchmark, that we release to the research community.

1 Introduction

Knowledge Base Population (KBP) from text is the problem of extracting relations between entities with respect to a given schema, usually defined by a set of types and relations. The facts added to the KB are triples, consisting of two entities connected by a relation. Although providing explicit provenance for the triples is often a subgoal in KBP, we focus on the case where correct triples are gathered from text without necessarily annotating any particular text with a relation. Humans are able to perform very well on the task of understanding relations in text. For example, if the target relation is *presidentOf*, anyone will be able

to detect an occurrence of this relation between the entities TRUMP and UNITED_STATES from both the sentences “Trump issued a presidential memorandum for the United States” and “The Houston Astros will visit President Donald Trump and the White House on Monday”. However, the first example expresses an explicit relation between the two entities, while the second states the same relation implicitly and requires some background knowledge and inference to identify it properly. In fact, the entity UNITED_STATES is not even mentioned explicitly in the text, and it is up to the reader to recall that US presidents live in the White House, and therefore people visiting it are visiting the US president.

Very often, relations expressed in text are implicit. This reflects in the low recall of the current KBP relation extraction methods, that are mostly based on recognizing lexical-syntactic connections between two entities within the same sentence. The state-of-the-art systems are affected by very low performance, close to 16.6% F1, as shown in the latest TAC-KBP evaluation campaigns and in the open KBP evaluation benchmark¹. Existing approaches to dealing with implicit information such as textual entailment depend on unsolved problems like inducing entailment rules from text.

In this paper, we address the problem of identifying implicit relations in text using a radically different approach, consisting of reducing the problem of identifying binary relations into a much larger set of simpler unary relations. For example, to build a Knowledge Base (KB) about presidents in the G8 countries, the *presidentOf* relation can be expanded to *presidentOf*:UNITED_STATES, *presidentOf*:GERMANY, *presidentOf*:JAPAN, and so

¹<https://kbpo.stanford.edu>

on. For all these unary relations, we train a multi-class (and in other cases, multi-label) classifier from all the available training data. This classifier takes textual evidence where only one entity is identified (e.g. ANGELA_MERKEL) and predicts a confidence score for each unary relation. In this way, ANGELA_MERKEL will be assigned to the unary relation *presidentOf*:GERMANY, which in turn generates the triple $\langle \text{ANGELA_MERKEL } \textit{presidentOf} \text{ GERMANY} \rangle$.

To implement the idea above, we explore the use of knowledge-level supervision, sometimes called distant supervision, to train a deep learning based approach. The training data in this approach is a knowledge base and an unannotated corpus. A pre-existing Entity Detection and Linking system first identifies and links mentions of entities in the corpus. For each entity, the system gathers its *context set*, the contexts (e.g. sentences or token windows) where it is mentioned. The context set forms the textual evidence for a multi-class, multi-label deep network. The final layer of the network is vector of unary relation predictions and the intermediate layers are shared. This architecture allows us to efficiently train thousands of unary relations, while reusing the feature representations in the intermediate layers across relations as a form of transfer learning. The predictions of this network represent the probability for the input entity to belong to each unary relation.

To demonstrate the effectiveness of our approach we developed a new KBP benchmark, consisting of extracting unseen DBpedia triples from the text of a web crawl, using a portion of DBpedia to train the model. As part of the contributions for this paper, we release the benchmark to the research community providing the software needed to generate it from Common Crawl and DBpedia as an open source project².

As a baseline, we adapt a state of the art deep learning based approach for relation extraction (Lin et al., 2016). Our experiments clearly show that using unary relations to generate new triples greatly complements traditional binary approaches. An analysis of the data shows that our approach is able to capture implicit information from textual mentions and to highlight the reasons why the assignments have been made.

The paper is structured as follows. In section 2 we describe the state of the art in distantly super-

vised KBP methodologies, with a focus on knowledge induction applications. Section 3 introduces the use of Unary Relations for KBP and section 4 outlines the process for producing and training them. Section 5 describes a deep learning architecture able to recognize unary relations from textual evidence. In section 6 we describe the benchmark for evaluation. Section 7 provides an extensive evaluation of unary relations, and a saliency map exploration of what the deep learning model has learned. Section 8 concludes the paper highlighting research directions for future work.

2 Related Work

Binary relation extraction using distant supervision has a long history (Surdeanu et al., 2012; Mintz et al., 2009). Mentions of entities from the knowledge base are located in text. When two entities are mentioned in the same sentence that sentence becomes part of the evidence for the relation (if any) between those entities. The set of sentences mentioning an entity pair is used in a machine learning model to predict how the entities are related, if at all.

Deep learning has been applied to binary relation extraction. CNN-based (Zeng et al., 2014), LSTM-based (Xu et al., 2015), attention based (Wang et al., 2016) and compositional embedding based (Gormley et al., 2015) models have been trained successfully using a sentence as the unit of context. Recently, cross sentence approaches have been explored by building paths connecting the two identified arguments through related entities (Peng et al., 2017; Zeng et al., 2016). These approaches are limited by requiring both entities to be mentioned in a textual context. The context aggregation approaches of state-of-the-art neural models, max-pooling (Zeng et al., 2015) and attention (Lin et al., 2016), do not consider that different contexts may contribute to the prediction in different ways. Instead, the context pooling only determines the degree of a sentence’s contribution to the relation prediction.

TAC-KBP is a long running challenge for knowledge base population. Effective systems in these competitions combine many approaches such as rule-based relation extraction, directly supervised linear and neural network extractors, distantly supervised neural network models (Zhang et al., 2016) and tensor factorization approaches to relation prediction. Compositional Universal

²<https://github.com/IBM/cc-dbp>

Schema is an approach based on combining the matrix factorization approach of universal schema (Riedel et al., 2013), with representations of textual relations produced by an LSTM (Chang et al., 2016). The rows of the universal schema matrix are entity pairs, and will only be supported by a textual relation if they occur in a sentence together.

Other approaches to relational knowledge induction have used distributed representations for words or entities and used a model to predict the relation between two terms based on their semantic vectors (Drozd et al., 2016). This enables the discovery of relations between terms that do not co-occur in the same sentence. However, the distributed representation of the entities is developed from the corpus without any ability to focus on the relations of interest. One example of such work is LexNET, which developed a model using the distributional word vectors of two terms to predict lexical relations between them (DS_h). The term vectors are concatenated and used as input to a single hidden layer neural network. Unlike our approach to unary relations the term vectors are produced by a standard relation-independent model of the term’s contexts such as word2vec (Mikolov et al., 2013).

Unary relations can be considered to be similar to types. Work on ontology population has considered the general distribution of a term in text to predict its type (Cimiano and Völker, 2005). Like the method of DS_h , this does not customize the representation of an entity to a set of target relations.

3 Unary vs Binary Relations

The basic idea presented in this paper is that in many cases relation extraction problems can be reduced to sets of simpler and inter-related unary relation extraction problems. This is possible by providing a specific value to one of the two arguments, transforming the relations into a set of categories. For example, the *livesIn* relation between persons and countries can be decomposed into 195 relations (one relation for each country), including *livesIn:UNITED_STATES*, *livesIn:CANADA*, and so on. The argument that is combined with the binary relation to produce the unary relation is called the *fixed argument* while the other argument is the *filler argument*. The *KB extension* of a unary relation is the set of all filler arguments in the KB, and the *corpus extension* is

the subset of the KB extension that occurs in the corpus.

A requisite for a unary relation is that in the training KB there should exist many triples that share a relation and a particular entity as one argument, thus providing enough training for each unary classifier. Therefore, in the example above, we will not likely be able to generate predicates for all the 195 countries, because some of them will either not occur at all in the training data or they will be very infrequent. However, even in cases where arguments tend to follow a long tail distribution, it makes sense to generate unary predicates for the most frequent ones.

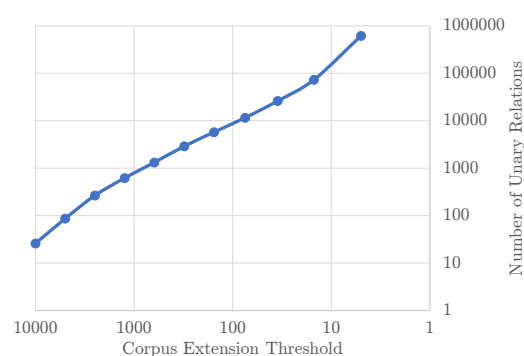


Figure 1: Minimum Corpus Extension to Number of Unary Relations

Figure 1 shows the relationship between the threshold for the size of the corpus extension of a unary relation and the number of different unary relations that can be found in our dataset. The relationship is approximately linear on a log-log scale. There are 26 unary relations with a corpus extension of at least 10,000. These relations include:

- *hasLocation:UNITED_STATES*
- *background:GROUP_OR_BAND*
- *kingdom:ANIMAL*
- *language:ENGLISH_LANGUAGE*

Lowering the threshold to 100 we have 8711 unary relations and we get close to 1M unary relations with more than 10 entities.

In a traditional binary KBP task a triple has a *relevant context set* if the two entities occur at least once together in the corpus - where the notion of ‘together’ is typically intra-sentential (within a single sentence). In KBP based on unary relations, a triple $\langle \text{FILLER } rel \text{ FIXED} \rangle$ has a relevant context

set if the unary relation $rel:FIXED$ has the filler argument in its corpus extension, i.e. the filler occurs in the corpus.

Both approaches are limited in different important respects. KBP with unary relations can only produce triples when fixing a relation and argument provides a relatively large corpus extension. Triples such as $\langle \text{BARACK_OBAMA } spouse \text{ MICHELLE_OBAMA} \rangle$ cannot be extracted in this way, since neither Barack nor Michelle Obama have a large set of spouses. The limitation of binary relation extraction is that the arguments must occur together. But for many triples, such as those relating to a person’s occupation, a film’s genre or a company’s product type, the second argument is often not given explicitly.

In both cases, a relevant context set is a necessary but not sufficient condition for extracting the triple from text, since the context set may not express (even implicitly) the relation. Figure 2 shows the number of triples in our dataset that have a relevant context set with unary relations exclusively, binary relations exclusively and both unary and binary. The corpus extension threshold for the unary relations is 100.

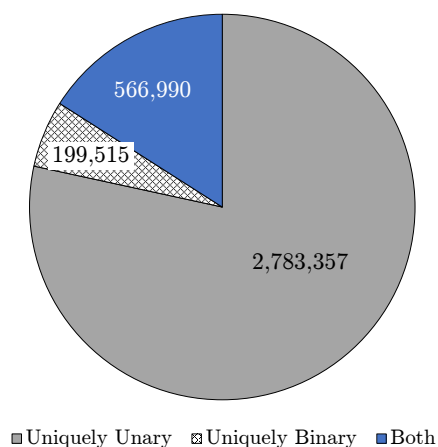


Figure 2: Triples with Relevant Context Sets Per Relation Style

Although unary relations could also be viewed as types, we argue that it is preferable to consider them as relations. For example, if the unary relation $lives_in:UNITED_STATES$ is represented as the type US-PERSON, it has no structured relationship to the type US-COMPANY ($based_in:UNITED_STATES$). So the inference rule that companies tend to employ people who live in the countries they

are based in ($\langle company \text{ employs } person \rangle \wedge \langle company \text{ based_in } country \rangle \Rightarrow \langle person \text{ lives_in } country \rangle$) is not representable.

4 Training and Using Unary Relation Classifiers

A unary relation extraction system is a multi-class, multi-label classifier that takes an entity as input and returns its probability as a slot filler for each relation. In this paper, we represent each entity by the set of contexts (sentences in our experiments) where their mentions have been located; we call them context sets.

The process of training and applying a KBP system using unary relations is outlined step-by-step below.

- Build a set of unary relations that have a corpus extension above some threshold.
- Locate the entities from the knowledge graph in text.
- Create a context set for each entity from all the sentences that mention the entity.
- Label the context set with the unary relations (if any) for the entity. The negatives for each unary relation will be all the entities where that unary relation is not true.
- Train a model to determine the unary relations for any given entity from its context set.
- Apply the model to all the entities in the corpus, including those that do not exist in the knowledge graph.
- Convert the extracted unary relations back to binary relations and add to the knowledge graph as new edges. Any new entities are added to the knowledge graph as new nodes.

A closer look to the generated training data can provide insight in the value of unary relations for distant supervision.

Below are example binary contexts relating an organization to a country. The two arguments are shown in bold. Some contexts where two entities occur together (relevant contexts) will imply a relation between them, while others will not. In the first context, **Philippines** and **Eagle Cement** are not textually related. While in the second context, **Dyna Management Services** is explicitly stated to be located in **Bermuda**.

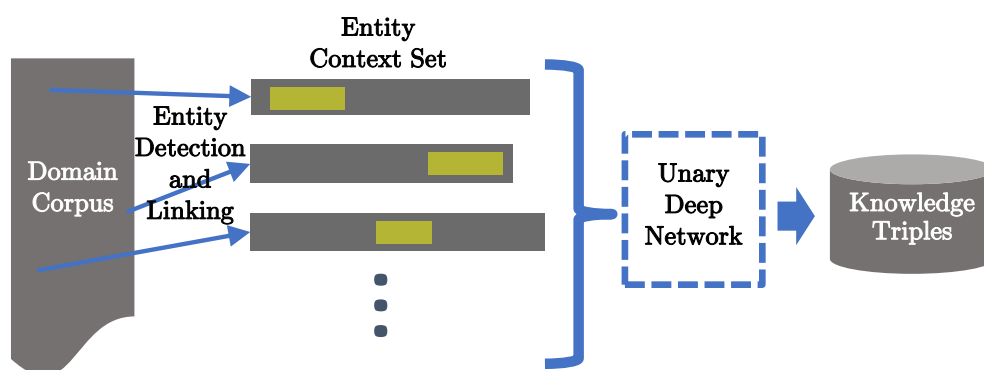


Figure 3: Unary Relational Knowledge Induction Architecture Overview

The company competes with Holcim **Philippines**, the local unit of Swiss company LafargeHolcim, and **Eagle Cement**, a company backed by diversified local conglomerate San Miguel which is aggressively expanding into infrastructure.

... said Richmond, who is vice president of **Dyna Management Services**, a **Bermuda**-based insurance management company.

On the other hand, there are many triples that have no relevant context using binary extraction, but can be supported with unary extraction. **JB Hi-Fi** is a company located in **Australia**, (unary relation *hasLocation:AUSTRALIA*). Although “**JB Hi-Fi**” never occurs together with “**Australia**” in our corpus, we can gather implicit textual evidence for this relation from its unary relation context sets. Furthermore, even cases where there is a relevant binary context set, the contexts may not provide enough or any textual support for the relation, while the unary context sets might.

Woolworths, Coles owner Wesfarmers, **JB Hi-Fi** and Harvey Norman were also trading higher.

JB Hi-Fi in talks to buy The Good Guys

In equities news, protective glove and condom maker Ansell and **JB Hi-Fi** are slated to post half year results, while Bitcoin group is expected to list on ASX.

The key indicators are: “ASX”, which is an Australian stock exchange, and the other Australian businesses mentioned, such as Woolworths,

Wesfarmers, Harvey Norman, The Good Guys, Ansell and Bitcoin group. There is no strict logical entailment, indicating **JB Hi-Fi** is located in Australia, instead there is textual evidence that makes it probable.

5 Architecture for Unary Relations

Figure 3 illustrates the overall architecture. First an Entity Detection and Linking system identifies occurrences in text of entities that are or should be in the knowledge base. Second, the contexts (here we use a sentence as the unit of context) for each entity are then gathered into an entity context set. This context set provides all the sentences that contain a mention of a particular entity and is the textual evidence for what triples are true for the entity. Third, the context set is then fed into a deep neural network, given in Figure 4. The output of the network is a set of predicted triples that can be added to the knowledge base.

Figure 4 shows the architecture of the deep learning model for unary relation based KBP. From an entity context set, each sentence is projected into a vector space using a piecewise convolutional neural network (Zeng et al., 2015). The sentence vectors are then aggregated using a Network-in-Network layer (NiN) (Lin et al., 2013).

The sentence-to-vector portion of the neural architecture begins by looking up the words in a word embedding table. The word embeddings are initialized with word2vec (Mikolov et al., 2013) and updated during training. The position of each word relative to the entity is also looked up in a position embedding table. Each word vector is concatenated with its position vector to produce each word representation vector. A piecewise max-pooled convolution (PCNN) is applied over

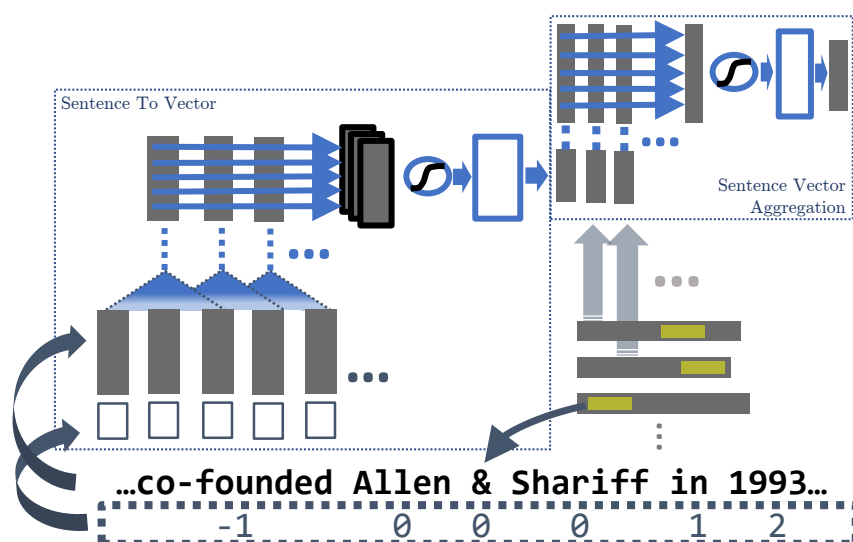


Figure 4: Deep Learning Architecture for Unary Relations

the resulting sentence matrix, with the pieces defined by the position of the entity argument: before the entity, the entity, and after the entity. A fully connected layer then produces the sentence vector representation. This is a refinement of the Neural Relation Extraction (NRE) (Lin et al., 2016) approach to sentence-to-vector mapping. The presence of only a single argument simply reduces from two position encoding vectors to one. The fully connected layer over the PCNN is an addition.

The sentence vector aggregation portion of the neural architecture uses a Network-in-Network over the sentence vectors. Network-in-Network (NiN) (Lin et al., 2013) is an approach of 1x1 CNNs to image processing. The width-1 CNN we use for mention aggregation is an adaptation to a set of sentence vectors. The result is max-pooled and put through a fully connected layer to produce the score for each unary relation. Unlike a maximum aggregation used in many previous works (Riedel et al., 2010; Zeng et al., 2015) for binary relation extraction the evidence from many contexts can be combined to produce a prediction. Unlike attention-based pooling also used previously for binary relation extraction (Lin et al., 2016), the different contexts can contribute to different aspects, not just different degrees. For example, a prediction that a city is in France might depend on the conjunction of several facets of textual evidence linking the city to the French language, the Euro, and Norman history.

In contrast, the common maximum aggregation approach is to move the final prediction layer to the sentence-to-vector modules and then aggregate by max-pooling the sentence level predictions. This aggregation strategy means that only the sentence most strongly indicating the relation contributes to its prediction. We measured the impact of the Network-in-Network sentence vector aggregation approach on the validation set. Relative to Network-in-Network aggregation and using the same hyperparameters, a maximum aggregation strategy gets two percent lower precision at one thousand: 66.55% compared to 68.49%.

There are 790 unary relations with at least one thousand positives in our benchmark. To speed the training, we divided these into eight sets of approximately 100 relations each and trained the models for them in parallel. Unary relations based on the same binary relation were grouped together to share useful learned representations. The resulting split also put similar numbers of positive examples in the training set for each model.

Training continued until no improvement was found on the validation set. This occurred at between five and nine epochs. All eight models were trained with the hyperparameters in Table 1. Dropout was applied on the penultimate layer, the max-pooled NiN.

Based on validation set performance, we found that when larger numbers of relations are trained together the NiN filters and sentence vector dimension must be increased. Of all the hyperparameters, the training time is most sensitive to the

Hyperparameter	Value
word embedding	50
position embedding	5
PCNN filters	1000
PCNN filter width	3
sentence vector	400
NiN filters	400
dropout	0.5
learnRate	0.003
decay multiplier	0.95
batch size	16
optimizer	SGD

Table 1: Hyperparameters used

number of PCNN filters, since these are applied to every sentence in a context set. We found major improvements moving from the 230 filters used for NRE to 1000 filters, but less improvement or no improvement to increases beyond that.

6 Benchmark

Large KBs and corpora are needed to train KBP systems in order to collect enough mentions for each relation. However, most of the existing Knowledge Base Population tasks are small in size (e.g. NYT-FB (Riedel et al., 2010) and TAC-KBP³) or focused on title-oriented-documents which are not available for most domains (e.g. WikiReading (Hewlett et al., 2016)). Therefore, we needed to create a new web-scale knowledge base population benchmark that we called *CC-DBP*⁴. It combines the text of Common Crawl⁵ with the triples from 298 frequent relations in DBpedia (Auer et al., 2007). Mentions of DBpedia entities are located in text by gazetteer matching of the preferred label. We use the October 2017 Common Crawl and the most recent (2016-10) version of DBpedia, in both cases limited to English.

We divided the entity pairs into training, validation and test sets with a 80%, 10%, 10% split. All triples for a given entity pair are in one of the three splits. This split increases the challenge, since many relations could be used to predict others (such as birthPlace implying nationality). The task is to generate new triples for each relation and rank them according to their probability. We show

³<https://tac.nist.gov/>

⁴<https://github.com/IBM/cc-dbp>

⁵<http://commoncrawl.org>

the precision / recall curves and focus on the relative area under the curves to evaluate the quality of different systems.

Figure 5 shows the distribution of triples with relevant unary context sets per relation type. The relations giving rise to the most triples are high level relations such as hasLocation, a super-relation comprised of the sub-relations: country, state, city, headquarter, hometown, birthPlace, deathPlace, and others. Interestingly there are 165 years with enough people born in them to produce unary relations. While these all will have at least 100 relevant context sets, typically the context sets do not have textual evidence for any birth year. Perhaps most importantly, there are a large number of diverse relations that are suitable for a unary KBP approach. This indicates the broad applicability of our method.

To test what improvement can be found by incorporating unary relations into KBP, we combine the output of a state-of-the-art binary relation extraction system with our unary relation extraction system. For binary relation extraction, we use a slightly altered version of the PCNN model from NRE (Lin et al., 2016), with the addition of a fully connected layer for each sentence representation before the max-pooled aggregation over relation predictions. We found this refinement to perform slightly better in NYT-FB (Riedel et al., 2010), a standard dataset for distantly supervised relation extraction.

The binary and unary systems are trained from their relevant context sets to predict the triples in train. The validation set is used to tune hyperparameters and choose a stopping point for training. We combine the output of the two systems by, for each triple, taking the highest confidence from each system.

7 Evaluation

Figure 6 shows the precision-recall curves for unary only, binary only and the combined system. The unary and binary systems alone achieve similar performance. But they are effective at very different triples. This is shown in the large gains from combining these complementary approaches. For example, at 0.5 precision, the combined approach has a recall of more than double (15,750 vs 7,400) compared to binary alone, which represents over 100% relative improvement.

The recall is given as a triple count rather than

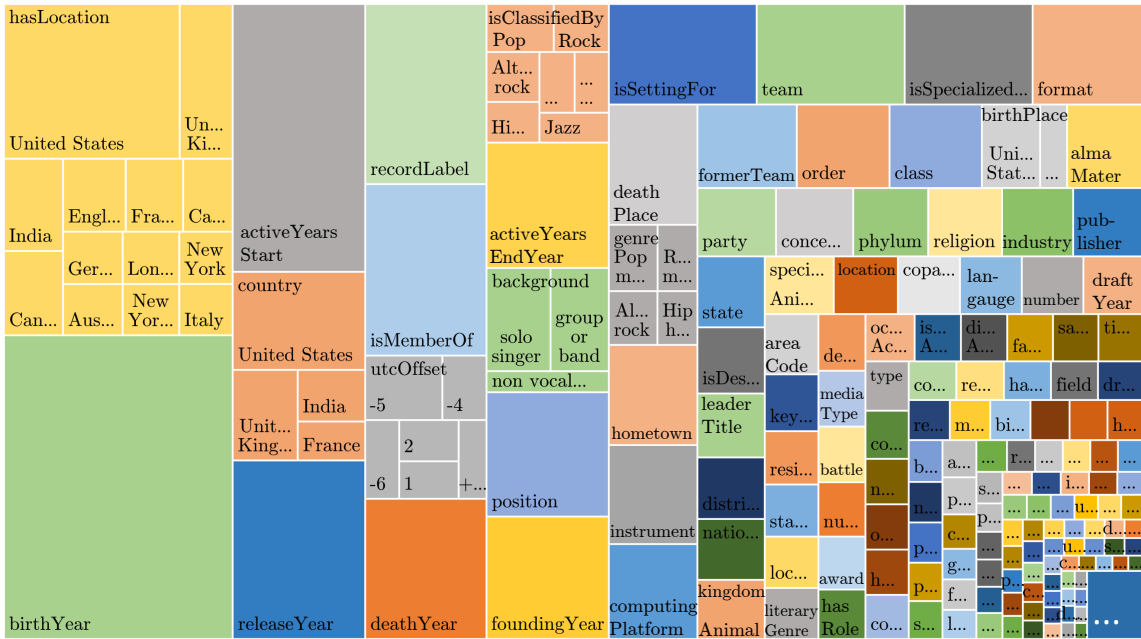


Figure 5: Distribution of Unary Relation Counts

a percentage. Traditional attempts to measure the recall of KBP systems use the set of all triples explicitly stated in text for the denominator of recall. This is unsuitable for evaluating our approach because the system is able to make probabilistic predictions based on implicit and partial textual evidence, thus producing correct triples outside the classic recall basis.

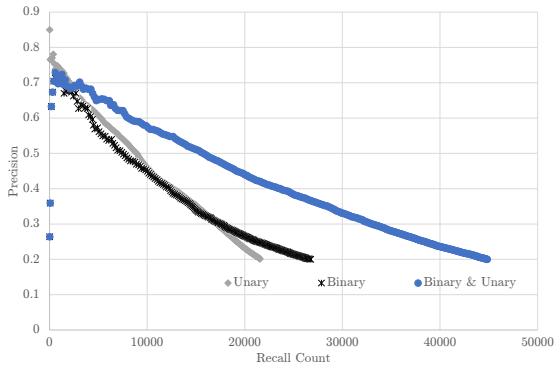


Figure 6: Precision Recall Curves for KBP

7.1 Saliency Maps

To gain some insight into how the unary KBP system is able to extract implicit knowledge we turn to saliency maps (Simonyan et al., 2014). By finding the derivative of a particular prediction with respect to the inputs, we can discover a locally linear approximation of how much each part of the input contributed (Zeiler and Fergus, 2014).

Cold Lake Provincial Park (Alberta, Canada) is mentioned in two sentences in the Common Crawl English text. The unary relational knowledge induction system predicts *hasLocation*:CANADA with the highest confidence (over 90%). Both sentences contribute to the decision. We see high weight from words including “cold”, “provincial” and “french”. A handful of countries have “provincial parks” including Argentina, Belgium, South Africa and Canada. Belgium and Canada have substantial French speaking populations and Canada has by far the coldest climate.

- located within 10 minutes of cold lake with quick access to OOV ridge ski hill , cold lake provincial park and french bay .
- welcome to cold lake provincial park on average 4.00 pages are viewed each , by the estimated 959 daily visitors .

Rock Kills Kid is a band mentioned twice in the corpus. From this context set, the relation *background*:GROUP_OR_BAND is predicted with high confidence. The fact that “Kid” occurs in the name of the entity seems to be important in identifying it as a musical group. The first sentence also draws focus to the band’s connection to rock and pop.

While the second sentence seems to recognize the *band - song (year)* pattern as well as the comparison to Duran Duran.

- the latest stylish pop synth band is rock kills kid .
- rock kills kid - are you nervous ? (2006) who ever thought duran duran would become so influential ?

The Japanese singer-songwriter Masaki Haruna, aka Klaha is mentioned twice in the corpus. From this context set, the relation *background:SOLO_SINGER* is predicted with high confidence. The first sentence clearly establishes the connection to music while the second indicates that Klaha is a solo artist. The conjunction of these two facets, accomplished through the context vector aggregation using NiN permits the conclusion of SOLO_SINGER.

- tvk music chat interview klaha parade .
- klaha tvk music chat OOV red scarf interview the tv - k folks did after klaha went solo .

8 Conclusions

In this paper we presented a new methodology to identify relations between entities in text. Our approach, focusing on unary relations, can greatly improve the recall in automatic construction and updating of knowledge bases by making use of implicit and partial textual markers. Our method is extremely effective and complement very nicely existing binary relation extraction methods for KBP.

This is just the first step in our wider research program on KBP, whose goal is to improve recall by identifying implicit information from texts. First of all, we plan to explore the use of more advanced forms of entity detection and linking, including propagating features from the EDL system forward for both unary and binary deep models. In addition we plan to exploit unary and binary relations as source of evidence to bootstrap a probabilistic reasoning approach, with the goal of leveraging constraints from the KB schema such

as domain, range and taxonomies. We will also integrate the new triples gathered from textual evidence with new triples predicted from existing KB relationships by knowledge base completion.

References

- Sren Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *In 6th Intl Semantic Web Conference, Busan, Korea*. Springer, pages 11–15.
- H Chang, M Abdurrahman, Ao Liu, J Tian-Zheng Wei, Aaron Traylor, Ajay Nagesh, Nicholas Monath, Patrick Verga, Emma Strubell, and Andrew McCallum. 2016. Extracting multilingual relations under limited resources: Tac 2016 cold-start kb construction and slot-filling using compositional universal schema. *Proceedings of TAC* .
- Philipp Cimiano and Johanna Völker. 2005. Towards large-scale, open-domain and ontology-based named entity classification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*.
- Aleksandr Drozd, Anna Gladkova, and Satoshi Matsuo. 2016. Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*. pages 3519–3530.
- Matthew R Gormley, Mo Yu, and Mark Dredze. 2015. Improved relation extraction with feature-rich compositional embedding models. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pages 1774–1784.
- Daniel Hewlett, Alexandre Lacoste, Llion Jones, Illia Polosukhin, Andrew Fandrianto, Jay Han, Matthew Kelcey, and David Berthelot. 2016. Wikireading: A novel large-scale language understanding task over wikipedia. In *Proceedings of the Conference on Association for Computational Linguistics*.
- Min Lin, Qiang Chen, and Shuicheng Yan. 2013. Network in network. *arXiv preprint arXiv:1312.4400* .
- Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of ACL*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. *Distributed Representations of Words and Phrases and their Compositionality*. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc.,

- pages 3111–3119. <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>.
- Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '09, pages 1003–1011. <http://dl.acm.org/citation.cfm?id=1690219.1690287>.
- Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017. Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics* 5:101–115.
- Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. *Machine learning and knowledge discovery in databases* pages 148–163.
- Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pages 74–84.
- Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *ICLR Workshop*.
- Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D Manning. 2012. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. Association for Computational Linguistics, pages 455–465.
- Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention cnns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 1298–1307.
- Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pages 1785–1794.
- Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*. Springer, pages 818–833.
- Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *EMNLP*. pages 1753–1762.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. pages 2335–2344.
- Wenyuan Zeng, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2016. Incorporating relation paths in neural relation extraction. In *CoRR*.
- Yuhao Zhang, Arun Chaganty, Ashwin Paranjape, Danqi Chen, Jason Bolton, Peng Qi, and Christopher D Manning. 2016. Stanford at tac kbp 2016: Sealing pipeline leaks and understanding chinese. *Proceedings of TAC* .