Easy First Relation Extraction with Information Redundancy

Shuai Ma[1,2]   **Gang Wang**[1,2]   **Yansong Feng**[3]   **Jinpeng Huai**[1,2]

[1]SKLSDE Lab, Beihang University, China

[2] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing, China

[3]Peking University, China

{mashuai, iegwang, huaijp}@buaa.edu.cn   fengyansong@pku.edu.cn

Abstract

Many existing relation extraction (RE) models make decisions globally using integer linear programming (ILP). However, it is nontrivial to make use of integer linear programming as a blackbox solver for RE. Its cost of time and memory may become unacceptable with the increase of data scale, and redundant information needs to be encoded cautiously for ILP. In this paper, we propose an easy first approach for relation extraction with information redundancies, embedded in the results produced by local sentence level extractors, during which conflict decisions are resolved with domain and uniqueness constraints. Information redundancies are leveraged to support both easy first collective inference for easy decisions in the first stage and ILP for hard decisions in a subsequent stage. Experimental study shows that our approach improves the efficiency and accuracy of RE, and outperforms both ILP and neural network-based methods.

1 Introduction

Relation extraction (RE) has been extensively studied due to its crucial role for many knowledge based applications (Fan et al., 2012; Zhang et al., 2012; Chen et al., 2014a), such as question answering and knowledge graph. There are two types of relation extractors: local and global. The former identify relationships between entity pairs individually according to the local features of sentences, *e.g.,* lexical and syntactic features (Bunescu and Mooney, 2007; Mintz et al., 2009; Surdeanu et al., 2010; Hoffmann et al., 2011; Surdeanu et al., 2012; Zeng et al., 2015; Lin et al., 2016). The latter make decisions by further considering joint features of the entire corpus (Yao et al., 2010; Li et al., 2011, 2013; de Lacalle and Lapata, 2013; Chen et al., 2014a). Global relation extractors are able to resolve conflict decisions, and to utilize the dependencies among extracted facts to improve the performance (Li et al., 2011; Chen et al., 2014a), commonly by formalizing RE as a constrained optimization problem and solving RE with integer linear programming (ILP) (Roth and Yih, 2004; Choi et al., 2006; Roth and Yih, 2007; Li et al., 2011; Ji et al., 2013; Chen et al., 2014a,b). However, global optimization remains dominated by the ILP solvers, suffering from heavy time expenses.

Using integer linear programming for RE as a blackbox solver is a challenging task. First, with the increase of entity pairs and candidate relations, the variables and constraints encoded for ILP increase dramatically, which in return may consume too much computing time and memory. Second, redundant information needs to be encoded cautiously. Consider <United States, *capital : 0.6*, New York>, <United States, *capital : 0.7*, New York> and <United States, *capital : 0.96*, Washington D.C.> in Table 1. Simple statistical methods, such as confidence summation, may easily lead to a wrong decision in this case.

To address the above issues, we propose to make easy (most confident) decisions first, and then to make hard decisions with ILP. The rationale lies in that easy decisions should be made early, and eliminating conflicts with constraints aids hard decision making. We leverage information redundancies, embedded in the results produced by sentence level extractors, to compute the confidences of candidate relations for entity pairs. Redundancy commonly exists in various corpora, *e.g.,* subject "United States" and object "New York" appear multiple times in Table 1. Even if there exists a specific relation between a subject and an object, the relation may not be reflected in certain mentions due to the lack of embedded evidence in single mentions, and *information redundancies are essentially to alleviate such insufficiency for decision making*.

3851

ID	Entity Pair	Top-3 Candidate Relations		
1	Australia, Canberra	*capital : 0.7*	*locationCity : 0.2*	*nationality : 0.1*
2	United States, New York	*capital : 0.6*	*birthPlace : 0.35*	*nationality : 0.05*
3	United States, New York	*capital : 0.7*	*country : 0.23*	*deathPlace : 0.07*
4	United States, Washington D.C.	*capital : 0.96*	*country : 0.03*	*birthPlace : 0.01*
5	Australia, Sydney	*locationCountry : 0.6*	*capital : 0.2*	*largestCity : 0.2*
6	University of Sydney, Sydney	*locationCity : 0.7*	*locationCountry : 0.2*	*city : 0.1*
7	Geoff Ogilvy, Australia	*nationality : 0.4*	*locationCountry : 0.3*	*city : 0.3*

Table 1: Example entity pairs and their candidate relations.

When making easy decisions, we keep the consistency among candidate relations by making use of constraints (*i.e.,* domain and uniqueness constraints) to eliminate conflicts directly, instead of implicitly encoding constraints in ILP (Yao et al., 2010; de Lacalle and Lapata, 2013; Chen et al., 2014a; Koch et al., 2014; Chen et al., 2014b). As a result, the number of variables and constraints to be encoded in ILP is significantly reduced, which speeds up the entire decision making process.

In short, our approach employs an easy-first strategy with information redundancies to make most confident decisions first during which conflict candidate relations are resolved directly by domain and uniqueness constraints, and it only makes the remaining hard decisions with ILP. Our approach is an important improvement of ILP-models, and it is not a new RE model, but an efficient and effective approach to fully exploiting the results of (any) local RE extractors. As shown in the experiments, our approach improves the performance, on average (4.58%, 17.90%) more accurate and (21, 37) times faster than existing ILP and neural network-based methods, respectively.

In the rest, we first introduce constraints and redundancies in Section 2, then present our detailed approach in Section 3, followed by experimental study in Section 4, related work in Section 5, and conclusions in Section 6.

2 Constraints and Redundancies

Consider a set M of mentions and a set R of predefined relations $\{r_1, \ldots, r_k\}$. For each mention m, a sentence level extractor produces an entity pair (subject s and object o) and a confidence score c_i for each relation r_i ($i \in [1, k]$), which represents the probability that s and o have relation r_i (Berger et al., 1996; Hoffmann et al., 2011). Here an NA (unknown) relation is typically included in R.

A mention m preprocessed by a sentence level extractor is essentially a $k + 2$ tuple $(s, o, r_1 : c_1, \ldots, r_k : c_k)$, as illustrated in Table 1, where

$1 \geq c_i \geq c_j \geq 0$ for any $i > j \in [1, k]$ and $\sum_{i=1}^{k} c_i = 1$. Observe that an entity pair (subject s and object o) with a relation $r \in R$ can be treated as an SRO triple (s, r, o), and a mention m contains k SRO triples. For convenience, given a mention m and a relation $r \in R$, we also denote the score of relation r in m as $m[r].c$.

Given a set M of mentions preprocessed by a sentence level extractor, our task for relation extraction is to determine the set of correct relations for each entity pair in M, and our approach adopts an easy-first strategy to make fast and accurate decisions, by leveraging constraints and redundancies that are to be introduced below.

In the following, a mention is referred to the one preprocessed by a sentence level extractor.

2.1 Constraints

We consider two types of constraints: domain and uniqueness constraints, commonly used to enforce agreements on decisions for RE (Yao et al., 2010; de Lacalle and Lapata, 2013; Chen et al., 2014a; Koch et al., 2014; Chen et al., 2014b).
Domain constraints. This type of constrains enforces constraints among the subject and object domains of relations. Given relations r_i and r_j ($1 \leq i, j \leq k$), (1) an *S-S domain constraint* ensures that r_i and r_j share no common entities between their subjects, *i.e.,* $r_i.subject \cap r_j.subject = \emptyset$, (2) an *O-O domain constraint* ensures that r_i and r_j share no common entities between their objects, *i.e.,* $r_i.object \cap r_j.object = \emptyset$, and (3) an *S-O domain constraint* ensures that r_i and r_j share no common entities between the subject of r_i and the object of r_j, *i.e.,* $r_i.subject \cap r_j.object = \emptyset$.

For example, (1) relations *largestCity* and *locationCity* have their subjects as countries (*e.g.,* "Australia") and organizations (*e.g.,* "University of Sydney"), respectively. They hold an *S-S* domain constraint; (2) relations *locationCity* and *locationCountry* have their objects as cities and countries, respectively. They hold an *O-O* domain constraint; (3) *locationCity* has its subjects as or-

# of mentions	# of entity pairs	# of subjects	# of objects	# / ratio of entity pairs mentioned multiple times		# / ratio of subjects mentioned multiple times		# / ratio of objects mentioned multiple times		Top 3 relation precision (%)
				≥ 3	≥ 5	≥ 3	≥ 5	≥ 3	≥ 5	
53162	30864	11360	9709	2754/ 8.92%	1225/ 3.97%	2914/ 25.65%	1591/ 14.01%	2442/ 25.15%	1351/ 13.91%	DB_me: 77.3/ 11.8/ 5.0 DB_nn: 76.5/ 10.6/ 5.0

Table 2: Redundancy statistics of DBpedia - (see Section 4 for details about the datasets).

ganizations and *nationality* has its objects as countries. It holds an *S-O* domain constraint.

Uniqueness constraints. This type of constrains enforces the uniqueness of the subjects and objects of relations. Given a relation r_i ($1 \leq i \leq k$), (1) an *S uniqueness constraint* ensures that there exist no subject s and objects $o_h \neq o_l$ such that both (s, r_i, o_h) and (s, r_i, o_l) hold, and (2) an *O uniqueness constraint* ensures that there exist no object o and subjects $s_h \neq s_l$ such that both (s_h, r_i, o) and (s_l, r_i, o) hold, respectively.

For example, relation *capital* holds both *S* and *O* uniqueness constraints, since a country as the subject of *capital* has only one capital, and a city as the object of *capital* can only be the capital of one country.

We refer to the total set of relations with *S-S* (resp. *O-O* and *S-O*) domain constraints as DC_{ss} (resp. DC_{oo} and DC_{so}). We also refer to the total set of relations with *S* (resp. *O*) uniqueness constraints as UC_s (resp. UC_o).

2.2 Information Redundancies

Redundancies are used to pick up hidden information (Downey et al., 2005; Banko et al., 2007; Li et al., 2011), and are very common in the corpus, as revealed by the statistics in Table 2. They are essentially the statistical characteristics (knowledge) of the results produced by local sentence level extractors, and are very important for global predictions. In this study, we introduce and leverage four classes of information redundancies: S-O, S-R, O-R and R redundancies.

S-O redundancies are introduced to aid the decision making of the top-one relations of mentions with the same subjects and objects. For a mention $m = (s, o, r_1 : c_1, \ldots, r_k : c_k)$ preprocessed by a sentence level extractor, its certainty degree $\mathsf{ent}(\mathsf{m})$ is defined as follows.

$$\mathsf{ent}(\mathsf{m}) = -\sum_{i=1}^{k} c_i \ln c_i. \qquad (1)$$

The redundancy score $\mathsf{RC}(\mathsf{s}, \mathsf{r_1}, \mathsf{o})$ of the top-one relation r_1 for subject s and object o in m, based

on its S-O redundancies, is defined as

$$\mathsf{RC}(\mathsf{s}, \mathsf{r_1}, \mathsf{o}) = \sum_{m' \in M_{s,o,r_1}} \alpha^{\mathsf{ent}(\mathsf{m'})}, \qquad (2)$$

where M_{s,o,r_1} is the set of mentions in M whose subjects are s, objects are o, and top-one relations are r_1, and α is a small positive number in $(0, 1)$ (set to 0.05 by default) to enforce that $\alpha^{\mathsf{ent}(\mathsf{m'})}$ falls into $(0, 1)$. Informally, $\mathsf{RC}(\mathsf{s}, \mathsf{r_1}, \mathsf{o})$ is a collective score based on its S-O redundancies, which makes use of *information entropy* to judge the confidence, and considers both the relative confidence scores and repeated times.

S-R redundancies are introduced to aid the decision making whether a subject meets the domain requirement of a relation. The likelihood score $\mathsf{LC}(\mathsf{s}, \mathsf{r})$ for subject s and relation r, based on its S-R redundancies, is

$$\mathsf{LC}(\mathsf{s}, \mathsf{r}) = \frac{\sum_{m \in M_s} m[r].c}{\sum_{m \in M_s} \sum_{r' \in R \setminus \{NA\}} m[r'].c}, \qquad (3)$$

where M_s is the set of mentions with the same subject s. Informally, $\mathsf{LC}(\mathsf{s}, \mathsf{r})$ measures the probability that relation r has a subject s among all relations except NA.

O-R redundancies are introduced to aid the decision making whether an object meets the domain requirement of a relation. The likelihood score $\mathsf{LC}(\mathsf{o}, \mathsf{r})$ for object o and relation r, based on its O-R redundancies, is

$$\mathsf{LC}(\mathsf{o}, \mathsf{r}) = \frac{\sum_{m \in M_o} m[r].c}{\sum_{m \in M_o} \sum_{r' \in R \setminus \{NA\}} m[r'].c}, \qquad (4)$$

where M_o is the set of mentions with the same object o. Informally, $\mathsf{LC}(\mathsf{o}, \mathsf{r})$ measures the probability that relation r has an object o among all relations except NA.

R redundancies are introduced to aid the decision making whether a subject and an object have a non-NA relation. The likelihood score $\mathsf{LC}(\mathsf{s}, \mathsf{o})$ for subject s and object o, based on its R redundancies, is

$$\mathsf{LC}(\mathsf{s}, \mathsf{o}) = \max_{m \in M_{s,o}} \sum_{r \in R \setminus \{NA\}} m[r].c, \qquad (5)$$
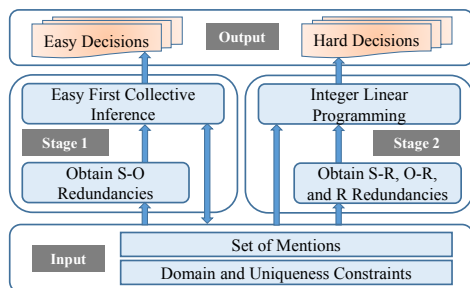
Figure 1: Framework of our approach eFIRE.

Figure 2: Algorithm for making easy decisions.

where $M_{s,o}$ is the set of mentions with the same subject s and object o. Informally, $\mathsf{LC(s,o)}$ selects prominent information from local decisions to measure the probability of having at least one non-NA relation between s and o.

As will be seen in the next section, redundancy scores $\mathsf{RC(s,r_1,o)}$ are used for making easy decisions, and likelihood scores $\mathsf{LC(s,r)}$, $\mathsf{LC(o,r)}$ and $\mathsf{LC(s,o)}$ are used for hard decisions, respectively.

3 Easy First Relation Extraction

We propose a novel easy FIrst approach for Relation Extraction with information redundancies, referred to as eFIRE. As shown by the framework in Figure 1, our approach obtains S-O redundancies, and makes easy decisions with easy first collective inference in the first stage. Then it derives S-R, O-R and R redundancies, and makes hard decisions with ILP in the second stage. We next introduce our approach in detail.

3.1 Easy First Collective Inference

In the first stage, decisions must be highly accurate to avoid error propagation. As revealed by Table 2, the decisions produced by local extractors are only reliable for top-one relations. Hence, eFIRE first makes decisions for entity pairs using their top-one relations only. The confidences are the redundancy scores obtained with the S-O redundancies (Section 2.2). Once a decision is made, disagreements are resolved immediately with domain and uniqueness constraints directly (Section 2.1).

Confidence computing and ordering. It computes the redundancy scores $\mathsf{RC(s,r_1,o)}$ of all the entity pairs (s,o) with their top-one relations r_1 in mentions M, using S-O redundancies. The higher the redundancy scores $\mathsf{RC(s,r_1,o)}$ are, the more confident subject s and object o have a relation r_1. Hence, entity pairs with their top-one relations are sorted in a descending order of their redundancy scores. As updating operations happen very often during the process of decision making, we introduce a max-heap MH to maintain all the entity pairs with their top-one relations and redundancy scores for the sake of efficiency.

Decision making and conflict resolving. It first makes a decision for the entity pair (s,o) by choosing its top-one relation r_1 such that the redundancy score $\mathsf{RC(s,r_1,o)}$ is the highest in MH.

Then it resolves conflicts accordingly. (1) For any relation $r \in R$ having an S-S domain constraint with r_1, all entity pairs with subject s and top-one relation r are deleted from MH, and for entity pairs with subject s in M, their scores of relation r are set to zeroes. It is similar for O-O and S-O domain constraints. (2) If relation r_1 has an S uniqueness constraint, all entity pairs with subject s and top-one relation r_1 are deleted from MH, and for entity pairs with subject s and object $o' \neq o$ in M, their scores of relation r_1 are set to zeroes. It is similar for O uniqueness constraints.

The above process is repeated until the highest redundancy score in the max-heap MH is less than a predefined threshold ϵ, which is to ensure the correctness of decisions. For the benefit of efficiency, we also index mentions by subjects, by objects and by entity pairs, separately.

The intention of threshold ϵ is to distinguish easy decisions from hard ones based on S-O redundancies. This indeed can be reflected from the redundancy scores. Consider an extreme case when there is only one mention m in M with subject s, object o and top-one relation r_1 whose score is 1.0, i.e., there are no S-O redundancies for mention m. In this case, we have $\mathsf{RC(s,r_1,o)} = 1.0$. However, for cases when multiple mentions with the same subject s, object o and top-one relation r_1, it is likely that the redundancy score is less than 1.0. Hence, threshold ϵ is typically set to a value a

little less than 1.0.

Our approach makes a better use of information redundancies in the corpus to aid the decision making process. Recall the example on determining whether the capital of "United States" is "New York" or "Washington D.C." in Section 1. With the S-O redundancies, "Washington D.C." is chosen as the capital of "United States", as RC(United States, *capital*, New York) = 0.18 and RC(United States, *capital*, Washington D.C.) = 0.57, which justifies the rationale of introducing the certainty degree ent(m) for mentions m.

Complete algorithm for easy first inference. The complete algorithm is presented in Figure 2.

It first initializes the set E of easy decisions empty (line 1). It then computes the redundancy scores of all the entity pairs with their top-one relations in mentions M, using S-O redundancies, and these entity pairs with their top-one relations are sorted in a descending order of their redundancy scores with a max heap MH (line 2). It repeatedly deals with entity pairs and their top-one relations in MH one by one until the highest redundancy score in MH is no more than ϵ (lines 3-6). Once a decision is made (lines 4, 5), conflicts are resolved immediately by updating MH and M (line 6). Finally, the modified mentions M and a set E of easy decisions are returned (line 7).

Time and space complexity analyses. The algorithm runs in $O(|M|^2(|R| + \log |M|))$ time, where $|M|$ and $|R|$ are the numbers of mentions and predefined relations, respectively. Observe the following. For a mention, (1) it takes $O(|R| + \log |M|)$ time to compute the redundancy scores for all entity pairs with their top-one relations, (2) maintaining MH can be done in $O(\log |M|)$ time, and (3) decision making and conflict resolving take $O(|M|(|R| + \log |M|))$ time. As there are in total $|M|$ mentions, we have the conclusion.

It is easy to see that the algorithm takes a space in the linear size of the set M of mentions.

3.2 Integer Linear Programming

In the second stage, our goal is to find an optimal configuration for the remaining mentions, making use of S-R, O-R and R redundancies, solving the disagreements by domain and uniqueness constraints, and maximizing the overall confidence of the made decisions. This is an NP-hard optimization problem, and many optimization models can be used to obtain approximate solutions. The tricky part is the design of the objective functions. Here, we adopt the ILP tool "IBM ILOG CPLEX".

More specifically, for each mention m and each of its candidate relations r in the set M of remaining mentions returned in the first stage, we define a binary decision variable v_m^r indicating whether relation r is chosen for the entity pair (s, o) in m by the solver. For each mention m in M, we choose its top three relations with scores no less than 0.1 as the candidate relations. As revealed by Table 2, candidates beyond top-3 are very unreliable.

Our objective is to maximize the total confidence of all the selected candidates based on the S-R, O-R and R redundancies, and the objective function can be written as:

$$\max \sum_{m \in M, r \in R^m} \bigl(\mathsf{LC}(\mathsf{s}, \mathsf{r}) + \mathsf{LC}(\mathsf{o}, \mathsf{r}) + \mathsf{LC}(\mathsf{s}, \mathsf{o}) \\ + \sum_{m' \in M_m} m'[r].c\bigr) v_m^r, \tag{6}$$

where s and o are the subject and object in m, R^m is the set of candidate relations for s and o in m, and M_m is the set of mentions in M having the same subject and object as m. The first component is the sum of S-R, O-R, and R redundancies of the selected candidates, and the second one is the sum of the original confidence scores of the selected candidates. The former is designed to encourage the model to select candidates meeting the domain requirements of relations, and the latter is designed to give consideration to decisions produced by sentence level extractors. That is, although the sentence level extractors may make wrong decisions, the global statistics of their decisions are reliable, and should be preserved.

The domain and uniqueness constraints are encoded to avoid conflict decisions as follows.

$$v_{m_i}^{r^{m_i}} + v_{m_j}^{r^{m_j}} \le 1, \tag{7}$$

where m_i and m_j have the same subject, $r^{m_i} \in R^{m_i}$, $r^{m_j} \in R^{m_j}$, and r^{m_i} and r^{m_j} have an S-S domain constraint in DC_{ss}.

$$v_{m_i}^{r^{m_i}} + v_{m_j}^{r^{m_j}} \le 1, \tag{8}$$

where m_i and m_j have the same object, $r^{m_i} \in R^{m_i}$, $r^{m_j} \in R^{m_j}$, and r^{m_i} and r^{m_j} have an O-O domain constraint in DC_{oo}.

$$v_{m_i}^{r^{m_i}} + v_{m_j}^{r^{m_j}} \le 1, \tag{9}$$

Datasets	Methods	Running time(s)	# of ILP variables	# of ILP constraints
DB_me	baseline	52.70	12353	293361
	eFIRE	**1.88**	**8185**	**34931**
	eFIRE-1S	1.90	8185	34931
	eFIRE-2S	52.35	12353	293361
	CNN+ATT	69.49	-	-
	PCNN+ATT	34.55	-	-
DB_nn	baseline	15.82	11770	94498
	eFIRE	**1.11**	**9678**	**24314**
	eFIRE-1S	1.20	9678	24314
	eFIRE-2S	15.81	11770	94498
	CNN+ATT	69.49	-	-
	PCNN+ATT	34.55	-	-

Table 3: Efficiency of eFIRE and comparison methods.

where the subject of m_i is the object of m_j, $r^{m_i} \in R^{m_i}$, $r^{m_j} \in R^{m_j}$, and r^{m_i} and r^{m_j} have an S-O domain constraint in DC_{so}.

$$\sum_{m \in M_{r,s}} v_m^r \leq 1, \qquad (10)$$

where $M_{r,s}$ is the set of mentions with candidate relation r, subject s and pairwise distinct objects, and r has an S uniqueness constraint in UC_s.

$$\sum_{m \in M_{r,o}} v_m^r \leq 1, \qquad (11)$$

where $M_{r,o}$ is the set of mentions with candidate relation r, object o and pairwise distinct subjects, and r has an O uniqueness constraint in UC_o.

By adopting ILP, eFIRE combines the scores refined in the first stage and the constraints to make hard decisions. After the optimization problem is solved, together with the easy decisions obtained in the first stage, eFIRE finally produces a list of selected candidate relations for each entity pair.

4 Experimental Study

In this section, we present an extensive experimental study of our easy first approach eFIRE.

4.1 Experimental Settings

We first present our experimental settings.

Datasets. The two datasets, DB_me and DB_nn, stem from DBpedia (Bizer et al., 2009), mapping the triples in DBpedia to sentences in the New York Time corpus. We map 51 different relations to the dataset. We use both a maximum entropy model MaxEnt (Chen et al., 2014a) and neural network model NN (Lin et al., 2016) as the sentence level extractors to output confidence scores, denoted as DB_me and DB_nn, respectively. The features of these two datasets are reported in Table 2. There are 53162 mentions in

each dataset, including 38654 mentions with NA relations. We learn domain and uniqueness constraints from Freebase as knowledge base (KB).

Algorithms for comparison. To evaluate our approach, we compare with three methods: the ILP based global method for RE in (Chen et al., 2014a) as baseline that use the global clues to help resolve the disagreements, and CNN+ATT and PCNN+ATT in (Lin et al., 2016) that are neural network-based methods with attention mechanism to use all informative sentences.

Implementation. All algorithms were implemented with C++. All experiments were run on a PC with 2 Intel(R) Xeon(R) E5Í C2640 2.6GHz CPUs and 64 GB of memory. When running time is measured, the test was repeated over 5 times and the average is reported.

4.2 Experimental Results

We next present our findings of the effectiveness and efficiency of our easy first approach eFIRE. Following previous work, we also use the precision in the low recall portion of the P-R curve as the effectiveness criterion.

Exp-1: Overall performance. In the first set of tests, we evaluated the effectiveness and efficiency, and the results are reported in Figures 3(a), 3(b) and Table 3, respectively.

Our approach eFIRE outperforms the methods for comparison. eFIRE is on average (4.80%, 4.36%), (17.99%, 28.10%) and (7.69%, 17.82%) more accurate than baseline, CNN+ATT and PCNN+ATT on (DB_me, DB_nn) in the low-recall portion $[0, 0.25]$ of the P-R curves, respectively. eFIRE consistently outperforms CNN+ATT and PCNN+ATT over the entire range of recall. While baseline tends to have results with a higher recall, it has a weakness of low precision, which is alleviated by eFIRE that is able to obtain more correct decisions. It is difficult to guarantee high precision and recall at the same time. For most KB population applications, only the high precision part is considered for the effectiveness evaluation. It is worth pointing out that we only compare the testing time of CNN+ATT and PCNN+ATT here.

Our method eFIRE also reduces the running time on all datasets. eFIRE is on average (28, 14), (37, 63) and (18, 31) times faster than baseline, CNN+ATT and PCNN+ATT on (DB_me, DB_nn), respectively. This is because the easy-first strategy of eFIRE significantly re-

(a) on DB_me dataset

(b) on DB_nn dataset

(c) on DB_me dataset

(d) on DB_nn dataset
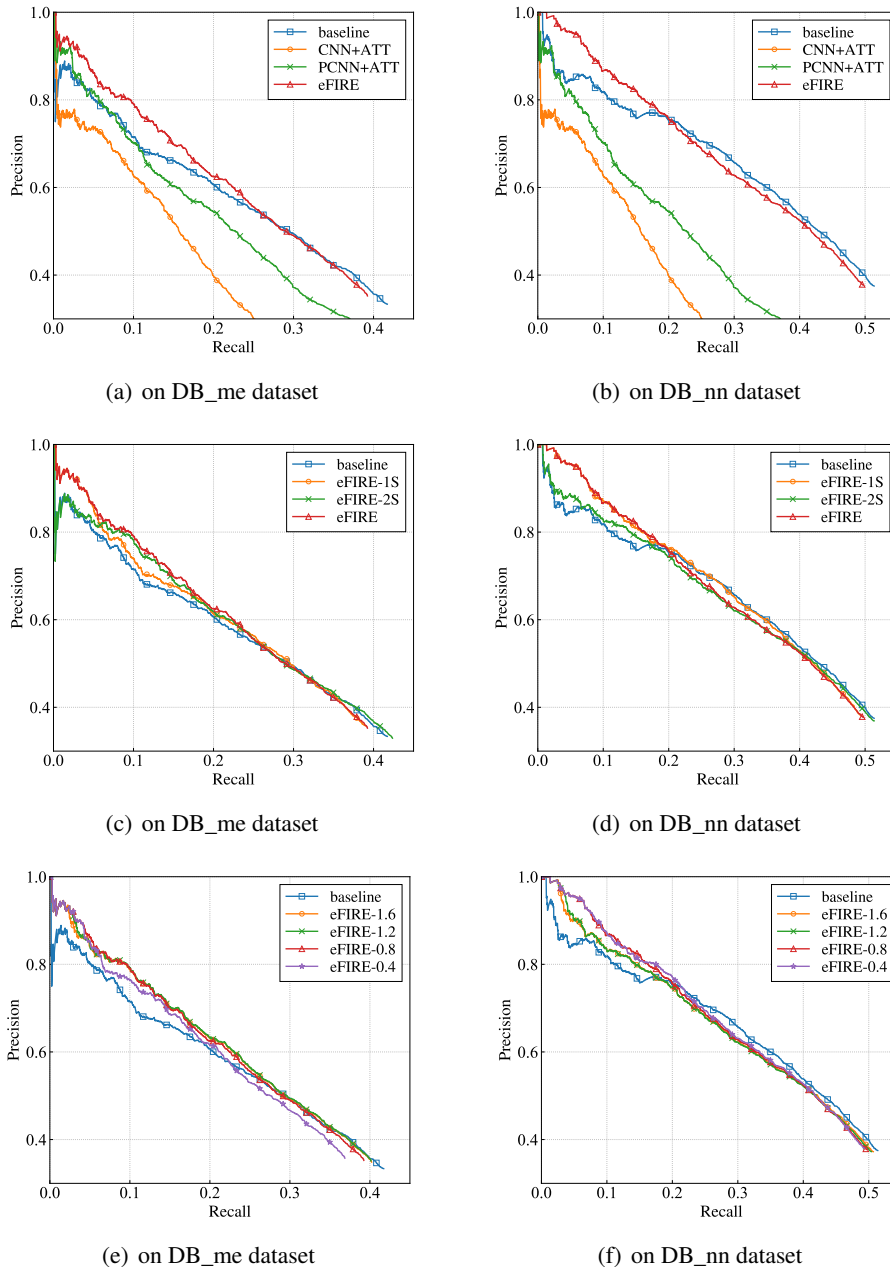
(e) on DB_me dataset

(f) on DB_nn dataset

Figure 3: Overall performances: eFIRE vs. its variants vs. baseline vs. CNN+ATT vs. PCNN+ATT.

duces the number of variables and constraints encoded in the ILP solver, as shown in Table 3.

Note that the running time has no obvious linear relationships with the number of variables and constraints among different datasets. In addition to the number of variables and constraints, objective functions have a impact on the running time of ILP too. Further, CPLEX is used as a black box, which makes it hard to have a precise analysis.

These results tell us that the easy-first strategy for RE, by making use of the redundancy information embedded in the local results of sentence level extractors, is an effective and efficient complement for RE using ILP solvers.

Exp-2: Performance of easy first collective inference with S-O redundancies. In the second set of tests, in order to evaluate the impacts of S-O redundancies, we implemented a variant of our approach, referred to as eFIRE-1S, that makes easy decisions with the easy first collective inference, and then adopts the same ILP method in baseline for making the rest decisions. The results are reported in Figures 3(c), 3(d), and Table 3.

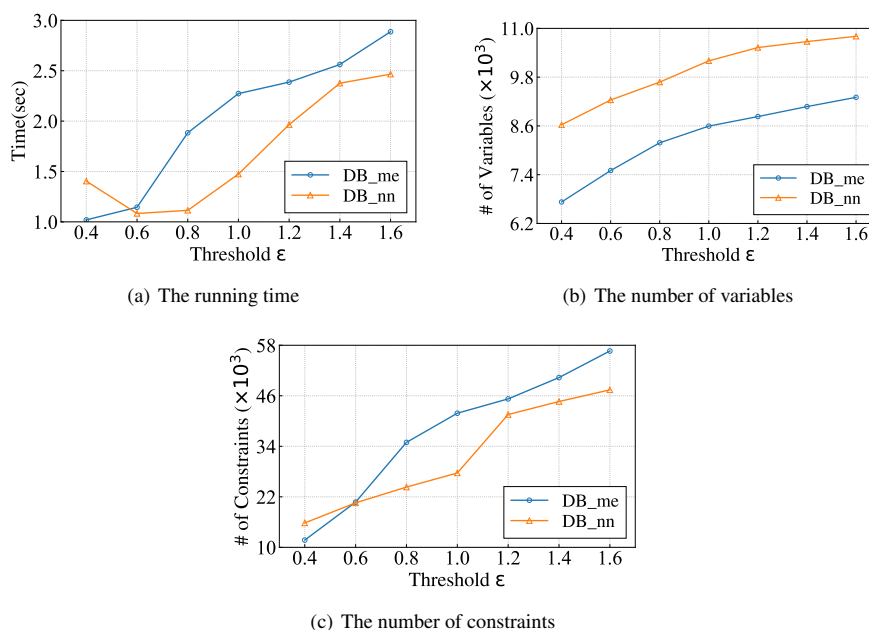Method eFIRE-1S outperforms baseline in the

(a) The running time

(b) The number of variables

(c) The number of constraints

Figure 4: Impacts of threshold ϵ on time, variables and constraints.

low-recall portions of the P-R curves on all two datasets. These results tell us that the easy first collective inference using S-O redundancies can not only improve the effectiveness of decision making for RE, but also improve the efficiency, as it significantly reduces the constraints and variables encoded in the ILP solver.

Exp-3: Performance of ILP with S-R, O-R and R redundancies. In the third set of tests, to evaluate the impacts of S-R, O-R and R redundancies, we implemented a variant of eFIRE, referred to as eFIRE-2S, that only consists of the second stage of eFIRE. That is, all decisions of eFIRE-2S are made by the ILP solver. The results are reported in Figures 3(c), 3(d) and Table 3.

Method eFIRE-2S outperforms baseline on all two datasets. It essentially improves the precision without sacrificing the recall. For ILP based methods, their key differences are the objective functions. eFIRE-2S incorporates more reliable statistics of sentence level extractors, *i.e.,* S-R, O-R and R redundancies, while baseline only uses maximal scores to encourage choosing the candidates with higher individual sentence level confidence scores. So, S-R, O-R and R redundancies benefit the decision making of ILP solvers. The efficiency of eFIRE-2S and baseline are comparable, which implies that the efficiency benefit of eFIRE comes from its first stage easy first collective inference.

Exp-4: Setting verification of threshold ϵ. In the last set of tests, to verify the setting of threshold

ϵ, we varied its values to $[0.4, 1.6]$. The results of effectiveness and running time are reported in Figures 3(e), 3(f) and Figure 4, respectively.

The results show that our approach eFIRE is effective and efficient when ϵ falls into $[0.5, 0.9]$, during which eFIRE outperforms baseline in the low-recall portion. This justifies our setting for threshold ϵ, which is typically a little less than 1.0 to distinguish easy decisions from hard ones (Section 3.1). Threshold ϵ obviously has an impact on the running time, as the smaller ϵ is, the more running time eFIRE has in the first stage, and the less it has in the second stage.

Summary. From these tests, we find followings.

(1) Our approach eFIRE is both effective and efficient. It is more accurate than baseline in the low-recall portion $[0, 0.25]$ and CNN+ATT and PCNN+ATT consistently. eFIRE is on average $(4.58\%, 23.05\%, 12.76\%)$ more accurate in the low-recall portion and $(21, 50, 25)$ times faster than baseline, CNN+ATT and PCNN+ATT, respectively. (2) The use of the easy-first strategy and S-O redundancies both improves the accuracy of RE and reduces the running time, and the use of S-R, O-R and R redundancies improves the accuracy of RE. (3) The setting of threshold ϵ is flexible in a range of $[0.5, 0.9]$ for our approach eFIRE.

5 Related Work

Relation extraction has been studied extensively in recent years, and can be divided into local re-

lation extractors (Mintz et al., 2009; Surdeanu et al., 2010; Hoffmann et al., 2011; Surdeanu et al., 2012; Søgaard et al., 2015) using the lexical features, syntactic features, and other local features of sentences, and global relation extractors utilizing the corpus features and relationships among local decisions (Kate and Mooney, 2010; Yao et al., 2010; Li et al., 2011; Singh et al., 2013; Li and Ji, 2014; Nguyen et al., 2017; Su et al., 2017). Global relation extractors leverage more information to resolve conflict decisions, which typically leads to more accurate decisions than local ones.

Recently, neural network-based models (Zeng et al., 2014; Ji et al., 2017) have been proposed for RE. Lin et al. (2016) proposes an attention mechanism to calculate weights for all sentences of one entity pair and selects plausible sentences from noisy sentences. Different from these are supervised methods that need label data for training, we propose an unsupervised method in this study.

Disagreements among decisions can be resolved by constraints. Yao et al. (2010) propose a relation extraction model that captures selectional preferences and functionality constraints to integrate information across documents. Fader et al. (2011) implement syntactic and lexical constraints on binary relations expressed by verbs in Open IE systems. Koch et al. (2014) impose type (or domain) constraints to only allow relations over appropriately typed mentions for relation extraction. Similar to (Chen et al., 2014a), our approach utilizes both domain and uniqueness constraints to resolve disagreements.

Many global relation extractors use integer linear programming as a blackbox solver (Roth and Yih, 2004; Choi et al., 2006; Roth and Yih, 2007; Li et al., 2011; Ji et al., 2013; Chen et al., 2014a,b; Wang et al., 2015). The ILP solver derives decisions through a well designed objective function, and resolves conflict decisions by encoding constraints into ILP. Our easy first approach is complimentary to these methods with each other, as these methods can take the easy first collective inference of our approach for making easy decisions as a first step, and our approach can make use of any of these methods as its solution for making hard decisions in its second stage.

Redundancies in the corpus have been used to pick up hidden information. Downey et al. (2005) consider redundant extractions for judging the correctness of extractions. Li et al. (2011) take ad-

vantage of redundant information to conduct reasoning across documents based on the information network structure. We introduce four classes of information redundancies: S-O, S-R, O-R and R redundancies, and we leverage S-O redundancies for making easy decisions, and the others to aid hard decision making.

Easy-first strategy relies on the intuition that "easy decisions should be made early, while harder decisions should be left for later when more information is available (Stoyanov and Eisner, 2012)". Their method makes easy decisions first for coreference resolution, and further makes use of the information from coreference clusters in the form of features to make later decisions. In this study, we propose to make easy (most confident) decisions first for relation extraction, and then to make hard decisions with ILP, where easy decisions are distinguished from hard ones with redundance scores based on S-O redundancies.

Data dependencies have well studied for improving data quality (Ma, 2011; Ma et al., 2015; Fan and Geerts, 2012), which essentially make use of data redundancies and dependencies. Our approach is partially inspired by these studies. Indeed, the uniqueness constraints defined in this study can be treated as functional dependencies (Abiteboul et al., 1995).

6 Conclusions

In this paper, we have proposed a fast easy first approach for relation extraction by making use of information redundancies, embedded in the results produced by local sentence level extractors, under domain and uniqueness constraints. We have introduced four classes of information redundancies to aid both easy first collective inference for easy decisions in the first stage and ILP for hard decisions in the second stage. Finally, using datasets processed by sentence level extractors with different models, we have experimentally verified that our easy first approach is both effective and efficient compared with state-of-the-art both ILP and neural network-based methods.

Acknowledgments

References

Serge Abiteboul, Richard Hull, and Victor Vianu. 1995. *Foundations of Databases*. Addison-Wesley.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2670–2676.

Adam L. Berger, Stephen Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. Dbpedia - A crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165.

Razvan C. Bunescu and Raymond J. Mooney. 2007. Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 576–583. Association for Computational Linguistics.

Liwei Chen, Yansong Feng, Songfang Huang, Yong Qin, and Dongyan Zhao. 2014a. Encoding relation requirements for relation extraction via joint inference. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 818–827. Association for Computational Linguistics.

Liwei Chen, Yansong Feng, Jinghui Mo, Songfang Huang, and Dongyan Zhao. 2014b. Joint inference for knowledge base population. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1912–1923. Association for Computational Linguistics.

Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 431–439. Association for Computational Linguistics.

Doug Downey, Oren Etzioni, and Stephen Soderland. 2005. A probabilistic model of redundancy in information extraction. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 1034–1041. Professional Book Center.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics.

James Fan, Raphael Hoffman, Aditya Kalyanpur, Sebastian Riedel, Fabian M. Suchanek, and Partha Pratim Talukdar, editors. 2012. *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, AKBC-WEKEX@NAACL-HLT 2012, Montrèal, Canada, June 7-8, 2012*. Association for Computational Linguistics.

Wenfei Fan and Floris Geerts. 2012. *Foundations of Data Quality Management*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.

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 *Proceedings of ACL 2011, the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 541–550. Association for Computational Linguistics.

Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2017. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 3060–3066. AAAI Press.

Heng Ji, Benoît Favre, Wen-Pin Lin, Dan Gillick, Dilek Hakkani-Tür, and Ralph Grishman. 2013. Open-domain multi-document summarization via information extraction: Challenges and prospects. In *Multi-source, Multilingual Information Extraction and Summarization*, pages 177–201. Springer.

Rohit J. Kate and Raymond J. Mooney. 2010. Joint entity and relation extraction using card-pyramid parsing. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 203–212. Association for Computational Linguistics.

Mitchell Koch, John Gilmer, Stephen Soderland, and Daniel S. Weld. 2014. Type-aware distantly supervised relation extraction with linked arguments. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1891–1901. Association for Computational Linguistics.

Oier Lopez de Lacalle and Mirella Lapata. 2013. Unsupervised relation extraction with general domain knowledge. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 415–425. Association for Computational Linguistics.

Qi Li, Sam Anzaroot, Wen-Pin Lin, Xiang Li, and Heng Ji. 2011. Joint inference for cross-document information extraction. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management*, pages 2225–2228. Association for Computing Machinery.

Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 402–412. Association for Computational Linguistics.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, (Volume 1: Long Papers)*, pages 73–82. Association for Computational Linguistics.

Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, (Volume 1: Long Papers)*, pages 2124–2133. Association for Computational Linguistics.

Shuai Ma. 2011. *Extending dependencies for improving data quality*. Ph.D. thesis, University of Edinburgh, UK.

Shuai Ma, Liang Duan, Wenfei Fan, Chunming Hu, and Wenguang Chen. 2015. Extending conditional dependencies with built-in predicates. *IEEE Transactions on Knowledge and Data Engineering*, 27(12):3274–3288.

Mike Mintz, Steven Bills, Rion Snow, and Daniel 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*, pages 1003–1011. Association for Computational Linguistics.

Dat Ba Nguyen, Martin Theobald, and Gerhard Weikum. 2017. J-REED: joint relation extraction and entity disambiguation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2227–2230. Association for Computing Machinery.

Dan Roth and Wen-tau Yih. 2004. A linear programming formulation for global inference in natural language tasks. In *Proceedings of the Eighth Conference on Computational Natural Language Learning at HLT-NAACL 2004*, pages 1–8. Association for Computational Linguistics.

Dan Roth and Wen-tau Yih. 2007. Global inference for entity and relation identification via a linear programming formulation. *Introduction to statistical relational learning*, pages 553–580.

Sameer Singh, Sebastian Riedel, Brian Martin, Jiaping Zheng, and Andrew McCallum. 2013. Joint inference of entities, relations, and coreference. In *Proceedings of the 2013 workshop on Automated knowledge base construction, AKBC@CIKM*, pages 1–6. Association for Computing Machinery.

Anders Søgaard, Barbara Plank, and Héctor Martínez Alonso. 2015. Using frame semantics for knowledge extraction from twitter. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2447–2452. AAAI Press.

Veselin Stoyanov and Jason Eisner. 2012. Easy-first coreference resolution. In *Proceedings of COLING 2012, the 24th International Conference on Computational Linguistics: Technical Papers*, pages 2519–2534. The COLING 2012 Organizing Committee.

Yu Su, Honglei Liu, Semih Yavuz, Izzeddin Gur, Huan Sun, and Xifeng Yan. 2017. Global relation embedding for relation extraction. *CoRR*, abs/1704.05958.

Mihai Surdeanu, David McClosky, Julie Tibshirani, John Bauer, Angel X. Chang, Valentin I. Spitkovsky, and Christopher D. Manning. 2010. A simple distant supervision approach for the TAC-KBP slot filling task. In *Proceedings of the Third Text Analysis Conference*. National Institute of Standards and Technology.

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*, pages 455–465. Association for Computational Linguistics.

Quan Wang, Bin Wang, and Li Guo. 2015. Knowledge base completion using embeddings and rules. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 1859–1866. AAAI Press.

Limin Yao, Sebastian Riedel, and Andrew McCallum. 2010. Collective cross-document relation extraction without labelled data. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1013–1023. Association for Computational Linguistics.

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762. Association for Computational Linguistics.

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. Dublin City University and Association for Computational Linguistics.

Congle Zhang, Raphael Hoffmann, and Daniel S. Weld. 2012. Ontological smoothing for relation extraction with minimal supervision. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. AAAI Press.