IBRE: An Incremental Bootstrapping Approach for Chinese Appointment and Dismissal Relation Extraction

Lu Mao, Meiling Wang, Changliang Li*, Junfeng Fan, Changyu Hou

AI Lab, KingSoft Corp, Beijing, China

{maolu, wangmeiling1, lichangliang, fanjunfeng, houchangyu}@kingsoft.com

Abstract. In the field of government affairs, Appointment and Dismissal Relation Extraction (ADRE) of officials from personnel news is crucial for updating government knowledge. However, ADRE faces great challenges, including extremely high accuracy demand, lack of data, tuple sparsity and conflict, and requiring incremental update. To address these challenges, we propose an Incremental Bootstrapping approach for Relation Extraction (IBRE) and apply it to real-time updating of personnel knowledge in government-affair Knowledge Graphs. IBRE starts with few seeds and trains with pattern generation, pattern evaluation, tuple prediction and seed augmentation in an iterative and incremental manner. First, we design a new strategy for defining seeds as document tuple pairs to eliminate the effects of tuple sparsity and conflict. Then, a new definition of patterns with both word and part-of-speech is proposed to get high accuracy. Finally, we augment seeds with corrected tuples and apply incremental learning to continually improve performance with least training cost. We build a dataset called ADNP (Appointment and Dismissal News from People.cn) and compare our approach with baselines. The results show that our approach performs the best. Moreover, experimental results demonstrate that incremental learning continuously improves the performance of IBRE.

Keywords: Relation Extraction · Bootstrapping · Knowledge Graphs · Incremental learning.

Motivation

C1: 任命姜万荣为住房和城乡建设部副部长。
C2: 免去张汉晖的外交部副部长职务。
E1: Appoint Jiang Wanrong as Vice Minister of Housing and Urban-Rural Development.
E2: Dismiss Zhang Hanhui from the post of Vice Minister of Foreign Affairs.

Appointment Tuple: (姜万荣, 住房和城乡建设部副部长)
Dismissal Tuple: (张汉晖, 外交部副部长)

Fig. 1 Examples of appointment and dismissal relation extraction.

Great challenges in Chinese ADRE:

- **Extremely high accuracy demand.** Government-affair area requires extremely high even 100 percent accuracy. Thus, human intervention is indispensable.
- **Lack of data.** The official websites such as People.cn only display about 300 news items. As a result, the training data are limited.
- **Tuple sparsity and conflict.** Most tuples appear only once through all the personnel news documents. Besides, an appointment tuple in one document can be a dismissal tuple in another document, thus causing conflict.
- **Incremental update.** The official websites periodically publish personnel news, thus training data can be gradually accumulated.
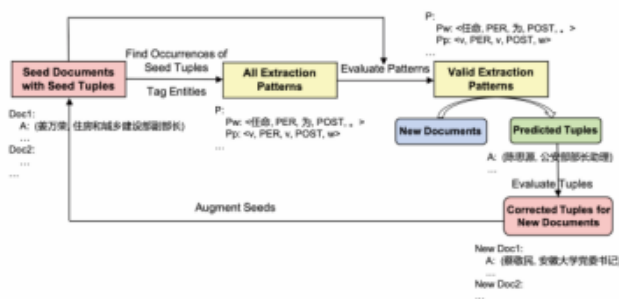
Proposed Approach

Fig. 2 Framework of IBRE.

For each relation type, IBRE primarily runs the following five steps:

- **Seed selection.** Choose seed documents with seed tuples from labeled data. This strategy can eliminate effects of tuple sparsity and conflict.
- **Pattern generation.** Automatically generate word patterns and POS patterns, which contributes to pattern matching with more comprehensive information.
- **Pattern evaluation.** Evaluate patterns by pattern matching with seed documents and keep only ones as valid that are regarded as being sufficiently reliable.
- **Tuple prediction.** For new documents, predict tuples by pattern matching with the valid patterns chosen before.
- **Seed augmentation.** Evaluate the predicted tuples, and then add the corrected tuples and their corresponding documents into seeds, which can gradually increase recall of RE.

**Table. 1** Examples of patterns generating from Snowball and IBRE.

Type	Appointment	Dismissal
Sentence	任命姜万荣为住房和城乡建设部副部长。 Appoint Jiang Wangrong as Vice Minister of Housing and Urban-Rural Development.	免去张汉晖的外交部副部长职务。 Dismiss Zhang Hanhui from the post of Vice Minister of Foreign Affairs.
Relation Tuple	(姜万荣, 住房和城乡建设部副部长)	(张汉晖, 外交部副部长)
Snowball Pattern	< 任命, PER, 为, POST, 。>	< 免去, PER, 的, POST, 职务 >
IBRE Pattern	Word:< 任命, PER, 为, POST, 。> POS:<v, PER, v, POST, w>	Word:< 免去, PER, 的, POST, 职务 > POS:<v, PER, u, POST, n>

Experimental Results

Dataset:

- We build a new dataset called ADNP. Our ANDP dataset contains 281 documents published between August and November, 2019, and we choose the first 200 as the training set and the rest as the test set.

Approach Comparison:

- Rule-based matching method. Choose tuples co-occurring with keywords.
- Snowball. A bootstrapping method starts with a few seed tuples to learn extracting relations and then use the extracted tuples for training in an iterative manner. Note that Snowball relies on multi-occurrence and consistent tuples.
- IBRE I. An IBRE method which applies word vectors with similarity greater than 0.6 as word similarity matching method.
- IBRE II. An IBRE method which applies word exact matching ignoring all function words and punctuation marks as word similarity matching method.

Table. 2 ADRE performance of IBRE and baseline in ADNP dataset.

Type	Appointment			Dismissal		
Approach	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
Rule-based	55.63	60.34	57.89	62.50	43.60	51.37
Snowball	27.86	53.06	36.53	14.06	41.28	20.97
IBRE I	87.70	74.92	80.80	86.78	61.05	71.67
IBRE II	**89.96**	**75.93**	**82.35**	**86.99**	**62.21**	**72.54**

Ablation Study:

Table. 3 ADRE results of ablation study.

Type	Appointment			Dismissal		
Approach	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
IBRE	89.96	75.93	**82.35**	86.99	62.21	**72.54**
IBRE w/o pattern evaluation	76.45	75.93	76.19	70.86	62.21	66.25
IBRE w/o POS pattern	**92.69**	68.81	78.99	**87.93**	59.30	70.83
IBRE w/o word pattern	82.31	**77.29**	79.72	38.99	**62.79**	48.11

Effects of Incremental Learning:
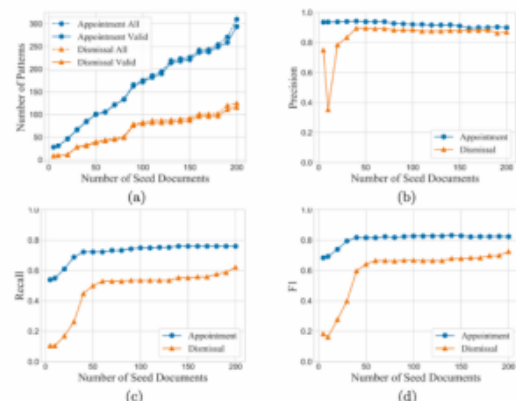
Conclusion

- We present a new bootstrapping approach IBRE to address the great challenges faced in ADRE.
- We build ADNP dataset to compare IBRE with baselines, and experimental results show that IBRE performs the best. IBRE has been applied to real-time updating of personnel knowledge in government-affair KGs.