

# FINDING OUT NOISY PATTERNS FOR RELATION EXTRACTION OF BANGLA SENTENCES

Rukaiya Habib and Md. Musfique Anwar

<sup>1</sup>Department of Computer Science and Engineering, Jahangirnagar University, Savar,  
Bangladesh

[rukaiya.habib45@gmail.com](mailto:rukaiya.habib45@gmail.com) [manwar@juniv.com](mailto:manwar@juniv.com)

## **ABSTRACT**

*Relation extraction is one of the most important parts of natural language processing. It is the process of extracting relationships from a text. Extracted relationships actually occur between two or more entities of a certain type and these relations may have different patterns. The goal of the paper is to find out the noisy patterns for relation extraction of Bangla sentences. For the research work, seed tuples were needed containing two entities and the relation between them. We can get seed tuples from Freebase. Freebase is a large collaborative knowledge base and database of general, structured information for public use. But for Bangla language, there is no available Freebase. So we made Bangla Freebase which was the real challenge and it can be used for any other NLP based works. Then we tried to find out the noisy patterns for relation extraction by measuring conflict score.*

## **KEYWORDS**

*Natural Language Processing, Relation Extraction, Bangla, Conflict Score, Noisy Pattern*

## **1. INTRODUCTION**

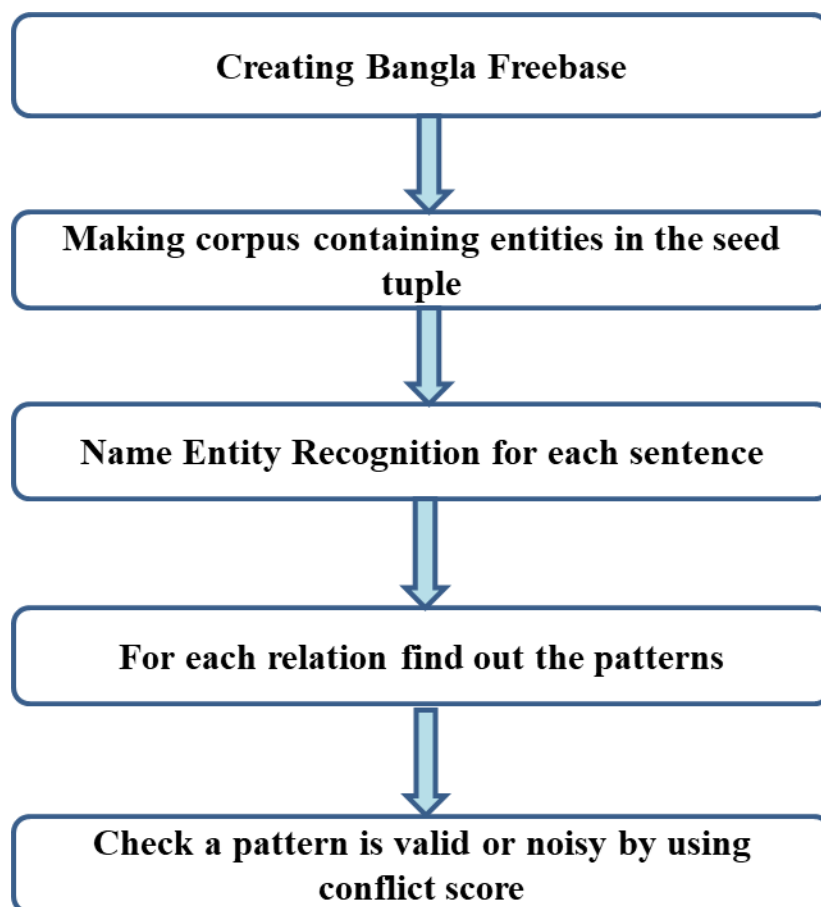
Natural language processing (NLP) is a branch of artificial intelligence which describes interaction of human and computer by manipulating human language. Its goal is to fill up the gap between human communication and computer understanding. Here, relation extraction (RE) is a fundamental topic of NLP. It is actually the task of finding semantic relationships between pairs of entities. Relation extraction is essential for many well-known tasks such as knowledge base completion, question answering, medical science and ontology construction [1]. There are so many unstructured electronic text data available on the web like newspaper, articles, journals, blogs, government and private documents etc. But unstructured text can be turned into structured by annotating semantic information.

Here, entities can be like person, organization, locations. We have to identify the entity types of a sentence. A relation is defined in the form of a tuple  $t = (e_1, e_2, \dots, e_n)$  where the  $e_i$  are entities in a predefined relation  $r$  within document  $D$ . Most relation extraction systems focus on binary relations. Examples of binary relations include born-in(Ruma, Dhaka), father-of(John David, Eric David) [8]. For relation extraction, we have different methods like supervised method, distant supervised method and unsupervised method.

In supervised method, sentences in a corpus are first hand-labeled for the presence of entities and relations between them. Lexical, syntactic and semantic features have to be extracted by the automatic content systems (ACE) to build supervised classifiers to label the relation between a given pair of entities in a test set sentence. Labeled training data is expensive to produce and thus

limited in quantity. Another one is distant supervised method for relation extraction which aligns texts to the given KB and use the alignment to learn a relation extractor [3]. They use the large amounts of structured data sources (such as Freebase) as the distant supervision information. As these methods do not need a hand-labeled dataset and KBs grow fast recently, they are more efficient.

In our work, we tried to find out noisy patterns for relation extraction of Bangla sentences for distantly supervised method. For this method, we need seed tuples which we can get from knowledge base like (Freebase). But there is no available Freebase in Bangla. So we built Bangla Freebase which contains a large amount of relations.



**Fig 1:** Noisy patterns identification steps for RE of Bangla Sentences

## **2. PREVIOUS STUDY**

Distant supervision can be introduced as an efficient method to scale relation extraction to very large corpora which contains a lot of relations. The authors proposed a sentence-level attention model to select the valid instances, which makes full use of the supervision information from knowledge bases [2]. And entity descriptions from Freebase and Wikipedia pages to supplement background knowledge have been extracted for their task. The background knowledge not only provides more information for predicting relations, but also brings better entity representations for the attention module. Three experiments have been conducted on a widely used dataset and the experimental results showed that their approach outperforms all the baseline systems

significantly [2].

Modern models of relation extraction for tasks like ACE are based on supervised learning of relations from small hand-labeled corpora. The authors used a paradigm that does not require labeled corpora [3]. This paradigm can avoid the domain dependence of ACE style algorithms, and allow the use of corpora of any size. This experiment used Freebase which is a large semantic database of several thousand relations. The Freebase provides distant supervision. For each pair of entities that appears in some Freebase relation, all sentences containing those entities in a large unlabeled corpus have been selected and extracted textual features to train a relation classifier. Their algorithm combines the advantages of supervised IE and unsupervised IE [3].

There have been so many works of relation extraction entities in English. In this work, we have worked on relation extraction of Bangla sentences on which not so much research work has been done. So it will be very much beneficial for this language. This is the nobility of our work.

### **3. CREATING BANGLA FREEBASE**

Freebase is a large collaborative knowledge base and database of general, structured information for public use. Its structured data had been harvested from many sources, including individual, user-submitted wiki contributions. Its aim is to create a global resource so that people (and machines) can access common information more effectively [9]. It is available in English. Actually in Freebase, the triple format is like (e1, r, e2) where e1 and e2 are the two entities and r defines the relation. So relation can be found in a known KG and can generate large amount of data [4]. Since we mentioned before that we created our own Bangla Freebase which contains a large number of relation with the help of Wikidata query service and SPARQL query language. It is a large collection of knowledge base database. Today the number of Bangla articles in the internet is growing day by day. So it has become a necessary to have a structured data store in Bangla. It consists of different types of concepts (topics) and relationships between those topics. These include different types of areas like popular culture (e.g. films, music, books, sports) location information (restaurants, locations, businesses), scholarly information (linguistics, biology, astronomy), birth place of (poets, politicians, actor, actress) and general knowledge (Wikipedia). Here we collected more than 100 relations according to our need. By using SPARQL query, anyone can find out their required relation. So this knowledge base is very much helpful. It will be much more helpful for relation extraction or any kind of NLP (Natural Language Processing) works on Bangla language.

#### **3.1. Wikidata Query Service**

Wikidata is a website that belongs to the Wikimedia family of websites. Data from Wikidata is available in RDF dumps. Actually RDF stands for Resource Description Framework which is a general method for describing data by defining relationships between data objects and it allows data integration from multiple sources. RDF has triple format which is a set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions [7]. Wikidata is a place to store structured data in many languages. The basic entity in Wikidata is an item. An item can be a thing, a place, a person, an idea or anything else. Wikidata has identifier numbers for entities and properties.

##### **3.1.1. Entity Identifier**

As Wikidata treats all languages in the same way, items don't have names, but generic identifiers. Each identifier is the letter Q that is followed by a number. For example, the item about the capital of Japan is called neither "Tokyo" nor "anything" but Q1490. But to give it a human-

readable name, each item has a list of labels in each language associated with it. So we'll see that the English (en) label at Q1490 is "Tokyo", also has corresponding word for the Japanese (ja) label, the Bangla (bn) label and so on.

### 3.1.2. Property Identifier

Every item has a list of statements associated with it. Each statement has a "property" and a "value". There is a long list of possible properties. Like items, properties have generic identifiers, but they begin with the letter P and not Q. For example, the property to indicate the country is P17, and it has the label "country" in English. The value of P17 (country) for Q1490 (Tokyo) is Q17 (Japan, etc.).

### 3.2. SPARQL Query Process

It is necessary to extract information from complaints, either scraped from the Web or received directly from the client for many companies nowadays. The aim is to find inside them some actionable knowledge. There is a query language, SPARQL to extract information from natural language documents, pre-annotated with NLP information. SPARQL stands for SPARQL Protocol and RDF Query Language. It is an RDF query language and able to retrieve and manipulate data stored in Resource Description Framework (RDF) format. SPARQL allows query to consist of triple patterns. It was made a standard by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium. SPARQL allows users to write queries against what can loosely be called "key-value" data [6]. We have to follow these steps:

- We can retrieve data according to our need by making query and for this, we have an online query service engine known as Wikidata query service. The URL of online query service is <https://query.wikidata.org/>.
- SPARQL is a standard query language technology which is endorsed by the World Wide Web consortium for querying any linked data information source. For making a query, we have to understand the SPARQL query. Here, one query has been added through which we will retrieve all the poets who are the citizen of Bangladesh.

```
SELECT ?item ?itemLabel ?occupationLabel ?citizenshipLabel WHERE
{
?item wdt:P31 wd:Q5.
?item wdt:P106 ?occupation.
?item wdt:P27 ?citizenship. FILTER (?citizenship=wd:Q902). FILTER
(?occupation=wd:Q49757).
SERVICE wikibase:label { bd:serviceParam wikibase:language "bn".}}
```

**Table 1.** A part of Bangla Freebase

item	itemLabel	occupationLabel	citizenshipLabel
<a href="http://www.wikidata.org/entity/Q4665322">http://www.wikidata.org/entity/Q4665322</a>	আবদুল গাফফার চৌধুরী (Abdul Gaffar Chowdhury)	কবি (Poet)	বাংলাদেশ (Bangladesh)
<a href="http://www.wikidata.org/entity/Q4665454">http://www.wikidata.org/entity/Q4665454</a>	আবদুল কাদির (Abdul Quadir)	কবি (Poet)	বাংলাদেশ (Bangladesh)
<a href="http://www.wikidata.org/entity/Q4667573">http://www.wikidata.org/entity/Q4667573</a>	আবিদ আজাদ (Abid Azad)	কবি (Poet)	বাংলাদেশ (Bangladesh)
<a href="http://www.wikidata.org/entity/Q4670213">http://www.wikidata.org/entity/Q4670213</a>	আবু হেনা মোস্তফা কামাল (Abu Hena Mustafa Kamal)	কবি (Poet)	বাংলাদেশ (Bangladesh)

## 4. METHODOLOGY

Relation extraction is very much significant in NLP based work. There are lot of methods for RE. In our work, we use **distant supervision** for relation extraction. In distant supervision, an already existing database, such as Freebase (knowledge-database) is prepared. Then we gather examples for the relations we want to extract. Thus our training data will be prepared. For example, Freebase contains the fact that Paris is the capital of France. We then label each pair of "France" and "Paris" that appear in the same sentence as a positive example for "capital\_of\_the\_country" relation. A large amount of (possibly noisy) training data can be generated. In the research work we needed seed tuples which are collected from Bangla Freebase made by us. Distant supervised method has been used which is very much efficient. In each seed tuple, there are two entities and their relation. There may be different types of entities like person, organization, location, films etc. We then extracted features from the sentences containing those entities in a large corpus. So we can say, our goal is to extract relation between two entities from sentences in a triple format and map the triple elements existing in a knowledge base [5]. After that we made decision that these extracted features are valid or not for each relation by measuring conflict score.

### 4.1. Name Entity Recognition

For our work, we had to identify the entity for each sentence. For entity identification, we used word level features (e.g., token, prefix, suffix), list lookup features (e.g., gazetteers). Gazetteers include names of countries, major cities, common people name, organization name etc.

### 4.2. Preparation of Corpus

A corpus is a collection of real world text in linguistics research. The intuition of our distant supervision approach is to use Freebase to give us a training set of relations and entity pairs that participate in those relations. In Freebase there are hundreds of relations. Freebase works as seed tuple. In our seed tuples, the Bangla synonym for the relations had been used. The seed tuples have different relations like 'birth-place', 'working-place', 'actor', 'film-director', 'film-producer' which Bangla synonyms are respectively জন্মস্থান, কর্মস্থল, বাসস্থান, ছবি\_পরিচালক, ছবি\_প্রযোজক. Our seed tuples look like the following.

**Table 2.** A seed tuple for relation extraction for Bangla sentences

Entity1(name)	Entity2 (place)	Relation
হুমায়ূন আহমেদ (Humayun Ahmed)	নেত্রকোনা (Netrokona)	জন্মস্থান (place_of_birth)
জসিমউদ্দিন (Jasimuddin)	ফরিদপুর (Faridpur)	জন্মস্থান (place_of_birth)
রহিম (Rahim)	রাজশাহী (Rajshahi)	কর্মস্থান (place_of_work)
আয়নাবাজি (Aynabaji)	অমিতাভ রেজা (Amitabh Reza)	ছবি_পরিচালক (film_director)
মনপুরা (Monpura)	গিয়াস উদ্দিন সেলিম (Giasuddin Selim)	ছবি_পরিচালক (film_director)
করিম (Karim)	সিলেট (Sylhet)	কর্মস্থল (place_of_work)
নাবিলা (Nabila)	মনপুরা (Monpura)	অভিনেত্রী (actress)

For each pair of entities in the seed tuple that appears in some Freebase relation, we used Wikipedia because it is relatively up to-date. We found out all sentences containing those entities in Wikipedia or large unlabeled corpus and collected them. Then we worked on them and extracted textual features. A part of our corpus looks like below:

**Table 3.** A part of our Bangla Corpus

No.	INPUT SENTENCES
1.	হুমায়ূন আহমেদ জন্মগ্রহণ করেছেন নেত্রকোনা জেলায়। ( Humayun Ahmed was born in Netrokona district.)
2.	টেলিভিশন ছবিতে তিশা অভিনয় করেছেন। (Tisha acted in Television film.)
3.	রবীন্দ্রনাথ জন্মগ্রহণ করেছিলেন কলকাতায়। (Rabindranath was born in Kolkata.)
4.	অমিতাভ রেজা আয়নাবাজি ছবিটি পরিচালনা করেছেন। (Amitabh Reza directed the Aynabaji film.)
5.	গত পাঁচ বছর ধরে রুমা ঢাকায় বাস করছে। ( Ruma has been living in Dhaka for five years.)
6.	আমার ছোট বোন শানু কাজ করে ঢাকায়। (My younger sister, Shanu works at Dhaka.)
7.	মাটির ময়না ছবিটি প্রযোজনা করেন ক্যাথেরিন মাসুদ। (Catherine Masud produced the film 'Earthen Moina'.)
8.	রহিম কর্মরত ছিলেন রাজশাহীতে। (Rahim was working in Rajshahi.)

### 4.3. Preparing Data And Pattern Identification

Before identifying patterns, we performed some tasks. We had to prepare the data. The following steps are performed:

- Tokenizing: We tokenized the data by inserting spaces between words and punctuations.
- Cleaning: Then we cleaned the data by removing empty lines, extra spaces and some lines that were too short or too long.
- Entity identification: In a given sentence, we identified the entity types and extracted the

relation patterns between them.

- **Chunking:** In preprocessing, consecutive words with the same named entity tag are ‘chunked’, like Ruma/PERSON Rahman/PERSON. So in a sentence if Ruma and Rahman appear together, these will be chunked together like [Ruma Rahman]/PERSON.
- **Lexical features:** Then to find out the pattern, we worked on lexical features of each sentence. Like:-
  - i) the sequence of words between the two entities is very much important.
  - ii) A flag indicating which entity came first in the sentence
  - iii) A window of n words to the left of Entity 1
  - iv) A window of n words to the right of Entity 2

These lexical features help us to identify the patterns of each relations. For example,

<Sentence> হুমায়ূন আহমেদ জন্মগ্রহণ করেছেন নেত্রকোণায়। Here, we get two entities,

person entity- “হুমায়ূন আহমেদ” and its position is 1

location entity- “নেত্রকোণা” and its position is 4

At first we got a ‘হুমায়ূন’ entity which is a person entity and then got ‘আহমেদ’ entity which is also a person entity. As they are same type of entity, they are chunked together. These two entities were available in the seed tuple list. We then identified the position of the person entity and location entity so that we can get the words between two entities which is actually the pattern for the relation of two entities. We then filtered out the noisy patterns by using conflict score. So the pattern is ‘জন্মগ্রহণ করেছেন’

#### 4.4. Conflict Score Formula

The main goal of our work is to identify the valid patterns for relations and find out noisy patterns existing there. The task can be done by identifying the conflict score for each pattern. Bangla Freebase is considered as seed tuples. Each seed tuple contains two entities and their relation. Our corpus has sentences containing these entities to get distant supervision. Our research work takes 5 relations. They are place\_of\_birth, place\_of\_work, living\_place, film\_director, film\_producer and for the relations, we used the Bangla synonym of the words which are respectively জন্মস্থান, কর্মস্থল, বাসস্থান, ছবি\_পরিচালক, ছবি\_প্রযোজক. Here the sentences which contain only one entity of the seed tuple create conflict for a relation. So identifying the conflict score, we can make a decision which patterns are valid for a relation. Thus we can find out noisy patterns which are invalid. The formula for conflict score is

$$\text{Conflict Score} = \frac{\text{Number of patterns with conflict}}{\text{Number of valid patterns}}$$

A threshold value is fixed which is 0.3. If the conflict score is less than or equal to threshold value then it is a valid pattern. Otherwise the pattern is invalid or noisy for a relation. For person and organization entities we take three relations. We know it is not necessarily for a person that his working place and birth place will be same. It helps us to find out the conflict patterns.

## 5. RESULT AND EVALUATION ON CONFLICT SCORE

We showed our results on five relations for Bangla sentences. The threshold value is 0.3. If the conflict is less than or equal to 0.3, it is considered as the valid pattern for that relation. Otherwise it is the noisy pattern. In future work we will remove these noisy patterns to train our relation classifier. The conflict scores have been mentioned in the following:

- **Relation 1:** Here, 'place\_of\_birth' relation is জন্মস্থান in Bangla. So the entities are person and location. For জন্মস্থান relation the conflict scores of different patterns are given below:

**Table 4.** Valid pattern identification for জন্মস্থান (place\_of\_birth) relation

No.	Patterns we get	The number of patterns with conflict	The number of valid patterns	Conflict Score	Valid or Invalid
1.	জন্মগ্রহণ করেছেন (was born in/at)	1	20	0.05	Valid
2.	জন্মগ্রহণ করে (was born in/at)	2	17	0.18	Valid
3.	কাজ করে (works at)	9	1	9	Invalid
4.	জন্ম হয় (born in/at)	0	14	0	Valid
5.	বেড়াতে গেছেন (went to travel)	4	2	2	Invalid
6.	কর্মরত আছেন (is/are working at)	12	3	4	Invalid
7.	জন্মস্থান হলো (birth place is)	3	34	0.09	Valid
8.	মৃত্যুবরণ করেন (has died)	7	2	3.5	Invalid

So, the valid patterns are:- জন্মগ্রহণ করেছেন (was born in), জন্মগ্রহণ করে (is born in), জন্ম হয় (born at/in), জন্মস্থান হলো (the birthplace is). Other patterns are noisy patterns for this relation. So the sentences containing these patterns will be removed. Noisy patterns are কাজ করে, বেড়াতে গেছেন, কর্মরত আছেন, মৃত্যুবরণ করেন

- **Relation 2:** Here, 'place\_of\_work' relation is 'কর্মস্থল' in Bangla. So the entities are person and location. For কর্মস্থল relation the conflict scores of different patterns are given below.



**Table 5.** Valid pattern identification for কর্মস্থল (place\_of\_work) relation

No.	Patterns we get	The number of patterns with conflict	The number of valid patterns	Conflict Score	Valid or Invalid
1.	কাজ করেন (works/work at)	2	16	0.125	Valid
2	কর্মরত ছিলেন (had been working)	0	31	0	Valid
3.	কাজ করে (works/work at)	2	21	0.01	Valid
4.	জন্মগ্রহণ করেছেন (was/were born in)	12	1	12	Invalid
5.	ঘুরতে গেছে (has/have gone to travel)	5	2	2.5	Invalid
6.	চাকরি করেন (works/work at)	1	19	0.05	Valid
7	কাজে নিযুক্ত আছেন (has been appointed to the work)	3	18	0.16	Valid
8.	অনুষ্ঠানের আয়োজন করেন (arranged the party)	5	3	1.67	Invalid

So, the valid patterns are:- কাজ করেন (works at), কর্মরত ছিলেন (had been working), কাজ করে (works/work at), চাকরি করেন (works at), কাজে নিযুক্ত আছেন (has been appointed to the work). Other patterns ঘুরতে গেছে, অনুষ্ঠানের আয়োজন করে ছিলেন, জন্মগ্রহণ করেছেন are noisy patterns.

- **Relation 3:** Here, 'living-place' relation is 'বাসস্থান' in Bangla. So the entities are person and location. For বাসস্থান relation the conflict scores of different patterns are given below:

**Table 6.** Valid pattern identification for বাসস্থান (living-place) relation

No.	Patterns we get	The number of patterns with conflict	The number of valid patterns	Conflict Score	Valid or Invalid
1.	বাস করে (lives/live in)	4	20	0.2	Valid
2	বসবাস করে আসছেন (has been living)	2	11	0.11	Valid
3.	কাজ করে (works at)	5	3	1.67	Invalid
4.	স্থায়ী বাসিন্দা (are the permanent resident)	1	14	0.08	Valid
5.	ঘুরতে গেছে (has gone to travel)	3	1	3	Invalid
6.	চাকরি করেন (works at)	4	2	2	Invalid

7.	শিক্ষাজীবন কাটিয়েছেন (has passed the student life)	10	2	5	Invalid
----	---	----	---	---	---------

So the valid patterns are:- বাস করে (lives/live in), বসবাস করে আসছেন (has been living), স্থায়ী বাসিন্দা ((are the permanent resident), Noisy patterns are কাজ করে, ঘুরতে গেছে, শিক্ষাজীবন কাটিয়েছেন, চাকরি করেন

- **Relation 4:** Here, ‘film-director’ relation is ‘ছবি\_পরিচালক’ in Bangla. So the entities are person and film. For ছবি\_পরিচালক relation the conflict scores of different patterns are given below:

**Table 7.** Valid patterns identification for ছবি\_পরিচালক (film-director) relation

No.	Patterns we get	The number of patterns with conflict	The number of valid patterns	Conflict Score	Valid or Invalid
1.	নির্দেশনা দিয়েছেন (has directed)	4	17	0.24	Valid
2	প্রযোজনা করেন (has produced)	6	7	0.86	Invalid
3.	পরিচালনা করেছেন (directed)	1	20	0.05	Valid
4.	পরিচালনা করেন (has directed)	4	22	0.19	Valid
5.	চলচ্চিত্রটির পরিচালক হলেন (film director is)	1	9	0.11	Valid
6.	অভিনয় করেছে (has acted)	7	4	1.75	Invalid

So, the valid patterns are:- নির্দেশনা দিয়েছেন (has directed), পরিচালনা করেছেন (directed), পরিচালনা করেন (has directed), চলচ্চিত্রটির পরিচালক হলেন (the film director is).

- **Relation 5:** Here, film\_producer’ relation is ‘ছবি\_প্রযোজক’ in Bangla. So the entities are person and film. For ছবি\_প্রযোজক relation the conflict scores of different patterns are given below:

**Table 8.** Valid patterns identification for ‘ছবি\_প্রযোজক (film-producer)’ relation.

No.	Patterns we get	The number of patterns with conflict	The number of valid patterns	Conflict Score	Valid or Invalid
1.	প্রযোজনা করেন (has produced)	1	10	0.1	Valid
2	অভিনয় করেছে (has acted)	5	1	5	Invalid
3.	নির্দেশনা দিয়েছেন (directed)	10	2	5	Invalid

So, the valid patterns are:- *প্রযোজনা করেন* (has produced).

## 6. FUTURE WORK

Here in this research work, we worked on relation extraction of Bangla sentences. RE is used in information extraction. As Bengali articles are increasing in the Web, this work holds very much significance for Bangla language based research work. Bangla language is very much enriched. In future, we will build a classifier where noisy patterns for any relation will be removed.

## 7. CONCLUSIONS

Relation extraction is very fundamental topic in NLP. In this work, we made a Freebase for Bangla which is a large collection of structured data by using Wikidata query service. It will be much more helpful for further research work like in natural language processing of Bangla, where the researchers need to get seed tuples. Actually researchers in areas such as entity extraction and reconciliation, data mining, Semantic Web, information retrieval, ontology creation and analysis can use this to support their work. With the help of Freebase, we get our seed tuples. We worked on Bangla sentences for relation extraction using distant supervision. Then we found out the noisy patterns using conflict score.

## ACKNOWLEDGEMENTS

We are thankful to the *Department of Computer Science & Engineering, Jahangirnagar University*.

## REFERENCES

- [1] Liu, Liyuan & Ren, Xiang & Zhu, Qi & Zhi, Shi & Gui, Huan & Ji, Heng & Han, Jiawei. "Heterogeneous Supervision for Relation Extraction: A Representation Learning Approach." arXiv preprint arXiv:1707.00166 [cs.CL] (2017).
- [2] Mike Mintz, Steven Bills, Rion Snow, Dan Jurafsky, "Distant supervision for relation extraction without labeled data"
- [3] Guoliang Ji, Kang Liu, Shizhu He, Jun Zhao, (2017). "Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions", Semantic Scholar.
- [4] Wang, Guanying & Zhang, Wen & Wang, Ruoxu & Zhou, Yalin & Chen, Xi & Zhang, Wei & Zhu, Hai & Chen, Huajun. "Label-Free Distant Supervision for Relation Extraction via Knowledge Graph Embedding." In proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (2018).
- [5] Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, et al. "Neural Relation Extraction for Knowledge Base Enrichment". In proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (July, 2019).
- [6] F. H. Marc Weise, Steffen Lohmann, "Ld-vowl: Extracting and visualizing schema information for linked data".
- [7] D. Hernández, A. Hogan, and M. Krotzsch, (2015). "Reifying RDF: what works well with wikidata?" Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems,

- [8] Nguyen Bach, Sameer Badaskar. “A Review of Relation Extraction.”. Semantic Scholar.
- [9] K. D. Bollacker, P. Tufts, T. Pierce, and R. Cook, (2007). “A platform for scalable, collaborative, structured information integration.”
- [10] Zeng, D. & Dai, Y. & Li, F. & Sherratt, R.S. & Wang, J. “Adversarial learning for distant supervised relation extraction.” Computers, Materials and Continua (2018).
- [11] Wang, Dongsheng & Tiwari, Prayag & Garg, Sahil & Zhu, Hongyin & Bruza, Peter. (2019). “Structural block driven - enhanced convolutional neural representation for relation extraction.”. Applied Soft Computing, 86. 105913. 10.1016/j.asoc.2019.105913.

### Authors

**Rukaiya Habib** is an M.Sc. student of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. She has completed her B.Sc. also in Computer Science and Engineering, from Jahangirnagar University in 2017. She is interested in Natural Language Processing, Mobile Adhoc Network and Computer vision related research works.



**Md. Musfique Anwar** has awarded PhD degree from the Department of Computer Science and Software Engineering, Faculty of Science, Engineering and Technology of Swinburne University of Technology, Melbourne, Australia in 2018. He has received M.Sc. degree from the Department of Intelligence Science and Technology, Graduate School of Informatics of Kyoto University, Japan in 2013 and B.Sc. degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2006. Since 2008, he is a faculty member having current designation Associate Professor in the Department of Computer Science and Engineering of Jahangirnagar University, Savar, Dhaka, Bangladesh. Currently, his research focuses on Data Mining, Social Network Analysis, Natural Language Processing and Software Engineering. He achieved Best Student Paper award in 29th Australasian Database Conference (ADC) in 2018, Best Poster award in 26th Australasian Database Conference (ADC) in 2015 and Best Poster Paper award in International Workshop on Computer Vision and Intelligent Systems-2019 (IWCVIS2019) in 2019.

