A Research about Named Entity Extraction based on Semantic Role Labeling

Haiyang Sun, Hao Zhang, Xuehan Chen, Yanqi Yao, Yirong Wang

April 2019

1 Introduction

This paper mainly discussed about an important research field, semantic role labeling in natural language processing. In this paper, our research group teased the origin and development of semantic role labeling, introduced some of mature semantic role labeling systems. The main goal we want is to use the semantic role labeling technique to do named entity recognition and extraction. In the middle of the paper, we chose some example of semantic role labeling technique that inspired us to discuss. In the end of this paper, we did research about the state-of-art approaches. We also implemented our own model using graph convolutional network.

2 Semantic Role Labeling

2.1 Introduction

Semantic role labeling, an automatic task to identify the different semantic roles in each arguments in each sentence of text. Semantic role is a shallow semantic representation level, which represents the abstract role that arguments of a predicate can take in the event.

Several decades ago, the semantic role labeling generated based on supervised machine learning (sometimes shortened as SML). With the development of the computational linguistics studies, there are still thousands of relative resources and development of statistical machine learning methods have heightened the amount of effort in this field.

Semantic role labeling usually use the resources from PropBank, VerbNet and FrameNet to determine the predicate counting, define the roles of task, and set training and test sets. In the above databases, such as PropBank and FrameNet, semantic role not only represents general semantic properties of the arguments, but also represents the relationship to the syntactic role of the argument in the sentence. This kind of relationships were codified in the databases.[1]

2.2 Semantic Roles

Semantic roles, also known as thematic roles, are one of the oldest classes of constructs in linguistic theory. Semantic roles are used to indicate the role played by each entity in a sentence and are ranging from very specific to very general.

Example (1) and (2) list some thematic roles that have been used in various computational papers, together with rough definitions and examples.

(1) Sasha broke the window.

$$\exists \text{ e,x,y Breaking(e)} \land \text{Breaker(e, Sasha)} \land \text{BrokenThing(e, y)} \land \text{Windows(y)}$$

(2) Pat opened the door.

$$\exists \text{ e,x,y Opening(e)} \land \text{Opener(e, Pat)} \land \text{OpenedThing(e, y)} \land \text{Door(y)}$$

In example (1) and (2), the semantic roles of the subjects of the verbs *break* and *open* are *Breaker* and *Opener* respectively. For each event, their deep roles are specific; *Breaking* events have *Breaker*, *Opening* events have *Openers*. *Breakers* and *Openers* are both volitional actors and have direct causal responsibility for their events.

Thematic roles are a way to capture this semantic commonality between *Breakers* and *Eaters*. The subject of both *Breakers* and *Eaters* are **AGENT**, which is the thematic role that represents the volitional causer of an event.

Similarly, the semantic role of the direct objects of both *BrokenThing* and *OpenedThing* is **THEME**, which is most directly affected by an event.

Figure 1 and figure 2 lists some thematic roles that have been used in various computational papers, together with some definitions and examples. Most thematic role sets have a dozen roles, but we will use he general term semantic roles for all sets of roles, whether it has small or large number of roles.[1]

Thematic Role	**Definition**
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

Figure 1: Some commonly used thematic roles with their definitions.

Thematic Role	**Example**
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice...*
RESULT	The city built a *regulation-size baseball diamond...*
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He poached catfish, stunning them *with a shocking device...*
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss...*
SOURCE	I flew in *from Boston.*
GOAL	I drove *to Portland.*

Figure 2: Some commonly used thematic roles with their examples.

2.2.1 Information Extraction

For NLP, the pipelines involved many processes, take the semantic level as an example, we extract and analyze each word to analyze the meaningful sentence representation. Therefore, the basic goal of NLP is to process the unstructured text and to produce a representation of its meaning.[2]

Information Extraction (sometimes shortened as IE) is a technologies that could be used to analyze text and extract limited kinds of semantic content from the free text. IE is a way to transfer the unstructured information to structured data in the free text.

Moreover, the destination of information extraction is to extract salient facts about pre-specified types of events, entities, or relationships, in order to build more meaningful, rich representations of their semantic content, which can be used to populate databases that provide more structured input.[3]

2.2.2 Named Entity Recognition (NER)

Named Entity Recognition (sometimes shortened as NER) , is a system to recognize the Named Entities that are included in free texts. More specifically, the task is to find Person (PER), Organization (ORG), Location (LOC) and Geo-Political Entities (GPE), which means who, where in a sentence.

Take this sentence as an example, in the statement "Michael Jordan lives in United States", NER system extracts Michael Jordan which refers to name of the person and United States which refers to name of the country. NER serves as the basis for various crucial areas in Information Management, such as Semantic Annotation, Question Answering, Ontology Population and Opinion Mining.[1]

3 The Proposition Bank

3.1 Introduction

The Proposition Bank, generally referred to as PropBank, is a resource of sentences annotated with semantic roles[4].The PropBank is based on the Penn TreeBank and labels all sentences in Penn TreeBank.

As it is hard to define a fine system of thematic roles, PropBank defines all the semantic roles with respect to an individual verb sense. Basically, a set of roles are classified a specific sense for each verb. These senses of verb has their own specific set of roles, and are named with Arg0, Arg1, Arg2, and so on. Generally, Arg0 means the PROTO-AGENT, Arg1 means the PROTO-PATIENT. Different verb has their own set of senses while Arg0 and Arg1 are almost consistent. Also Arg2 is most likely benefactive, instrument, attribute, and Arg4 the end point[4].

3.2 Examples

This section will show some simple PropBank examples with respect of senses for verbs display and swim. ProbBank names a set of sense for a word as frame files. The definitions here in the frame file are specifically prepared for human reading, but not formal definitions dedicated for machine process.

3.2.1 display

Verb 'display' has four args in PropBank, which are showed below.

Roles:
Arg0-PAG: displayer, agent
Arg1-PPT: entity displayed
Arg2-LOC: location
Arg3-GOL: displayed to whom? Seer of thing displayed

Ex1: She-1 had gone so far as [*-1] to display the questions on an overhead projector and underline the answers.
Arg0: [*-1]
Rel: display
Arg1: the questions
Arg2: on an overhead projector

Ex2: His moral character displayed itself in the sincerity of his friendships, his love of justice and of truth .
Arg0: His moral character
Rel: displayed
Arg1: itself
Arg2: in the sincerity of his friendships, his love of justice and of truth

As you can see in this example, we can devide each part of a sentence to their corresponding 'arg'.

3.2.2 swim

Verb 'swim' has four args in PropBank, which are showed below.

Roles:
Arg0-PPT: swimmer
Arg1-PPT: course (e.g. channel)
Arg2-GOL: location
Arg3-DIR: displayed to whom? Seer of thing displayed

Ex1: Pictures of rusted oil drums swim into focus, and the female voice purrs, "That hazardous waste on his (Mr. Courter's) property – the neighbors are suing for consumer fraud."

Arg0: Pictures of rusted oil drums
Rel: swim
Arg2: into focus

Ex2: The river swam with fish, so many you could almost walk from bank to bank and never touch the water.

Arg1: the river
Rel: swam
Arg0: with fish
Argm-prd: so many you could almost walk from bank to bank and never touch the water

The PropBank semantic role labeling enables us to find some commonalities in some different sentences. Here is another example for word 'increase' from Speech and language processing[4].

increase.01 "go up incrementally"
Arg0: causer of increase
Arg1: thing increasing
Arg2: amount increased by, EXT, or MNR
Arg3: start point
Arg4: end point

[$_{\text{Arg0}}$ Big Fruit Co.] increased [$_{\text{Arg1}}$ the price of bananas].
[$_{\text{Arg1}}$ The price of bananas] was increased again [$_{\text{Arg0}}$ by Big Fruit Co.]
[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].

You can see here in each case *Big Fruit Co.* is the AGENT and *the price of bananas* is the THEME, although these sentences are groups with different format.

4 FrameNet

4.1 Introduction

As we talked in the last section, PropBank is pretty useful. Like the example 'increase' in PropBank, it generates the semantic commonalities among a lot of sentences. Let's think a little bit further. Can we generates such semantic commonalities in more situations, among different verbs, even between nouns and verbs. Let's see how could we get the similarities between these sentences below:

[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].
[$_{\text{Arg1}}$ The price of bananas] rose [$_{\text{Arg2}}$ 5%].
There has been a [$_{\text{Arg2}}$ 5%] rise [$_{\text{Arg1}}$ in the price of bananas].

You can find that for the second sentence we use a different verb *rise*. For the third sentence we use the noun version of *rise*. We want the system know what was going up and how much it went up. This is why FrameNet invented.

The FramNet has two important features: (a) a commitment to corpus evidence for semantic and syntactic generalizations, and (b) the representation of the valences of its target words (mostly nouns, adjectives, and verbs) in which the semantic portion makes use of frame semantics. [5]

4.2 Example

In FrameNet project, we call a set of background knowledge related to theme as a frame[6]. For example:

reservation, flight, travel, buy, price, cost, fare, rates, meal, plane

These individual words are clearly defined with a background of air trabel. Other scholars also defined similar frames, like a model[7] or a script[8]. These words in a frame, which are frame-specific roles, are defined as frame elements. The FrameNet dataset includes a set of frames and frame elements, the lexical units associated with each frame, and a set of labeld example sentences[4]. Here we will show the **having_commercial_agreement** frame as an example:

> Party1 and Party2 (collectively refered to as Parties) make a commercial agreement, an Obligation, which both Parties are expected to keep or fulfill.

Normally, semantic roles in the frames are defined as core roles and non-core-roles. For the **having_commercial_agreement** frame, part of the core roles are below:

OBLIGATION	An expression of the commitment which the Parties have to engage in a commercial transaction.
PARTIES	The group of individuals portrayed as equally involved in having a commercial agreement.
TOPIC	Topic is a description of the domain covered by the commercial agreement.

Part of the non-core roles are below:

MEDIUM	As with other frames in the Communication domain, the Medium of communication may be expressed.
PLACE	This FE identifies the Place where an agreement was made.
TIME	This FE identifies the Time when the agreement was made.

5 Semantic Representation

How the blocks of a semantic system, entities, concepts, relations, and predicates, should be represented is a long-standing issue in semantic role labeling. This section provides a brief survey of approaches for semantic representation (SR).

5.1 First-order logic (FOL)

First-order logic is symbolized reasoning in which each sentence, or statement, is broken down into a subject and a predicate[4]. The predicate modifies or defines the properties of the subject. In first-order logic, a predicate can only refer to a single subject. First-order logic is also known as first-order predicate calculus or first-order functional calculus[9].

A sentence in first-order logic is written in the form of P(x), where P is the predicate and x is the subject, represented as a variable[10]. Complete sentences are logically combined and manipulated according to the same rules as those used in Boolean algebra.

For example, "IsAnimal" could be a predicate that when applied to an object returns True is its is a animal.

IsAnimal(dog) # true if argument, dog, is a animal.

Besides that, FOL uses connectives *and* and *or* to combine statements

Eat(dog, meat) $\wedge$ IsAnimal(dog) # dog eats meat and dog is animal.

FOL also use the universal quantifier $\forall$ or the existential quantifier $\exists$ to assert particular properties are true for all or just some objects[11].

$$\exists \text{ X isDog(X)} \wedge \text{Eat(X, meat)}$$

FOL is good at obtain the meaning of sentences in English, especially if these are concerned with objective, well-structured text such as product inventories, flight or hotel booking. The semantics of logic in these kind of text is basically equate to the meaning in the real world[12], so FOL is suitable here. Also, FOL is well-understood and inference machinery for it's features.

5.2 Semantic Networks

A semantic network, or frame network is a knowledge base that represents semantic relations between concepts in a network. Semantic network consist of nodes, links and link labels. Nodes in network represent concepts, entities and events. Links appear as arrows to express the relationships between objects, and link labels specify particular relations.
Figure above is the semantic network of representation sentence "Jack cooks a meal for Bob.". Here Jack and Bob are entities, Cook is a event and meal is a concept.

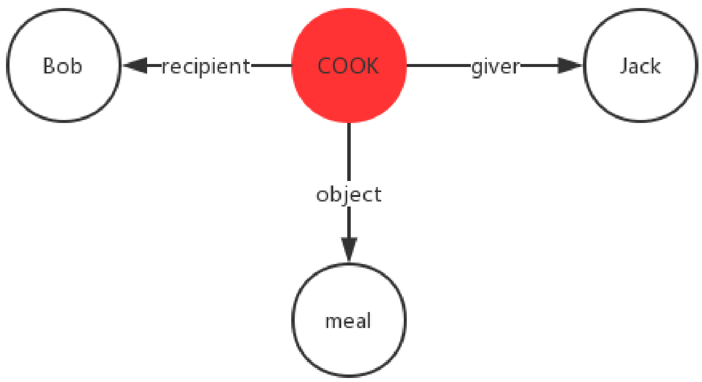

Figure 3: Sample of Semantic Network

This kind of representation avoids duplication of identical terms so that each distinct term or corresponds to a distinct node and there are labeled links from node to the node corresponding to its parts[13].

As more and more nodes and links are added to a semantic network diagram, one disadvantage of semantic network is showing: it tends to become more and more unreadable[14].

Besides non-binary relations, Inheritance is one of the main kind of reasoning done in semantic nets. The ISA (is a) relation is often used to link a class and its super class.
Figure above is the semantic network of representation, "Snoopy is a white and cute dog which is a kind of animal with teeth.". Some links (e.g. haspart in the figure) are inherited along ISA paths.

Note that a node can have any number of super classes that contain it, enabling a node to inherit properties from multiple parent nodes and their ancestors in the network. It can cause conflicting inheritance.

The advantages of semantic network are very obvious: (i) Easy to visualize.

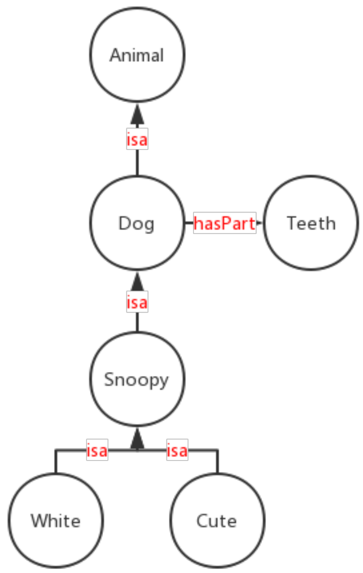

Figure 4: Sample of Semantic Network Inheritance

(ii) Related knowledge is easily clustered. (iii) Efficient in space requirements since objects represented only once and relationships handled by pointers.

6 Case Grammar

There are many approaches to do semantic representation, some of them are discussed in the last paragraph. Case Grammar is one of another solution of semantic representation. Nowadays, Case Grammar become more and more popular in Semantic Role Labeling.

6.1 Origin

Case Grammar is a system of linguistic analysis. This system was first created by American Linguist Charles J. Fillmore in 1968[15]. Basically, it focus on links between valence, objects and number of subjects. Before Fillmore's theory, Chomsky's Transformational Grammar[16] would reduce active and passive versions of the same deep structure, but doesn't go far enough to reveal why this is is possible semantically. Case Grammar, on the other hands, analyzes the surface syntactic structure of sentences by studying the combination of semantic roles. Chomsky proposed three rules in his first book, *Syntactic Structures*[17], published in 1957: phrase structure rules, conversion rules, and morpheme phoneme rules. The goal of its phrase structure rules (S $\rightarrow$ NP + VP; V + NP) is to generate all sentences. As a result, although the goal of generating all the sentences was reached, the correct sentence (John drinks wine) was generated, and the wrong sentence was also generated (Wine drinks John). This shows that there must be a semantic restriction between verbs and nouns. In order to avoid the generation of erroneous sentences, there should be some lexical and semantic restrictions on the rules. For example: the verb "drink", generally speaking, the noun in front of it is a person, at least alive, and the subsequent noun is a liquid.

For This problem in his first book, Chomsky published the second book, *The Aspects of the Theory of Syntax*[18]. Mainly add some restrictions to the rules he discribed in his first book.

A year after the publication of the second book, new problems were discovered. The first person who rise against is Chomsky's student Fillmore. He believes that syntactic structures are more convenient and sophisticated than Chomsky's conversion rules. In order to make up for the lack of conversion-generating grammar from the perspective of semantics, Fillmore published *Toward a Modern Theory of Case*[19] in 1966, *The Case for Case*[15] in 1968, *Some Problems for Case Grammar*[20] in 1971 and *The Case for Case Reopened*[21] in 1977.

6.2 The Meaning of Case

In traditional grammar, "case" refers to the morphological changes of nouns and pronouns used in the language to express the grammatical relationship between words. This kind of case must have a dominant or morphological mark, that is, based on the change of the form. For example, there are four cases in German.

The "case" in traditional linguistics is the "surface case." The normal sign is a suffix change or a pronunciation change, which is an unique phenomenon in some languages. The "case" in the case grammar is a "deep case". It is the transitive relationship between the body words (nouns, pronouns) and predicates (verbs, adjectives).

6.3 Current Cases

At first, Fillmore presented six cases, which are Agentive(A), Instrumental(I), Dative(D), Factitive(F), Locative(L) and Objective(O). Later, he added some cases into the language analysis, including Benefactive(B), Source(S), Goal(G) and Cominative(C).

The center of case grammar is verb, and each verb can dominate a certain cases. Its composition includes three parts: deep structure, surface structure and transformation. The deep structure consists of two parts: proposition and modality. Propositions represent the relationship between verbs and nouns, The most important one is verbs. The relationship between verbs and nouns is fixed[22].

6.4 Language Analysis Using Case Grammar

Case grammar is widely used in natural language processing, it plays an important role in machine translation, artificial intelligence, etc. It is an important basic theory of linguistic information processing.

After the mid-1970s, the grammar entered the second stage. Fillmore called the structure that describe the case role as "underlying structure". The "underlying structure" consists of case role. In the first stage of case grammar, the "underlying structure" obtains the surface structure through transformation. But in the second stage, the "underlying structure" must also be assigned through grammatical relations before conversion, then obtains the "deep structure". Then use the "deep structure" to convert into "surface structure".

Thus, each sentence has two analysis parts, a "case role" and a grammatical relationship. These two parts associate the sentence with the events described by the sentence, explaining the semantic and syntactic phenomena of the sentence.

According to Fillmore, the sentence describes the scene. Each participant in the scene acts as a case, forming the substructure of the sentence. The "underlying structure" is selected by the perspective, some participants enter the perspective. Becoming the nucleus of the sentence.

A scene is a real world outside of language, such as objects, events, states, behaviors, changes, and people's memories, feelings, and perceptions of the real world. Every word, phrase, and sentence in the language is a description of the scene. When people say a word, phrase, sentence or a paragraph, they are all determining a scene and highlighting or emphasizing a part of that scene.[23][24]

7 State-of-the-Art Approach

Semantic Role Labeling (SRL) is a way to implement shallow semantic analysis. It focuses on the predicate of the sentence, and only analyzes the relationship between the components and the predicate in the sentence, that is, the Predicate-Argument structure. It uses the semantic role to describe these structural relations. It is an important step in tasks such as information extraction and chapter analysis.

A version of graph convolutional networks (GCNs), a recent class of neural networks operating on graphs, over syntactic dependency trees are used as sentence encoders, producing latent feature representations of words in a sentence. The stacked GCN and LSTM layers produce the best reported scored on the standard benchmark (CoNLL-2009) both for Chinese and English.[25]

7.1 Comparison of two Model

In Semantic Role Labeling, if the predicate of sentence is given, the question can be converted to "labeling a sentence", in this problem, Bi-LSTM is the state-of-the-art model in sequence labeling. A classical Bi-LSTM sequence labeling model is shown in Figure 5.

But this approach sometime seems too direct, it usually makes people confusing. It looks like the Bi-LSTM model only uses the information in the sentence. In fact, more information can be used such as traditional syntactic analysis, the result of observation syntactic analysis and results of actual

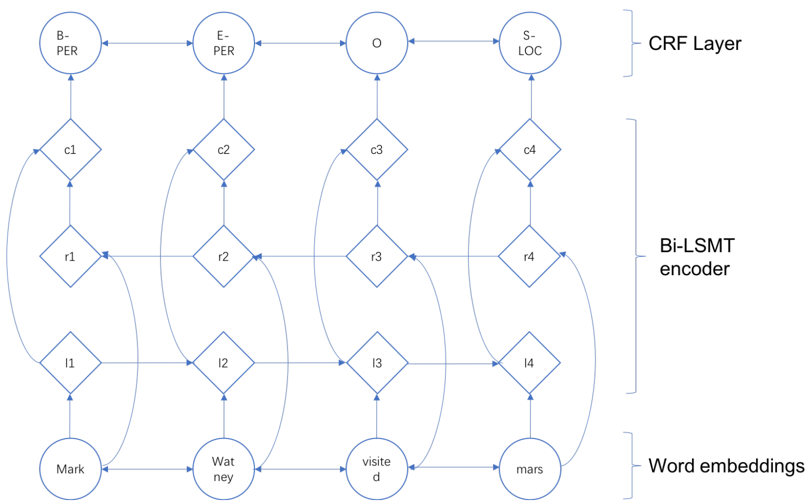

Figure 5: A classical Bi-LSTM sequence labeling model

semantic role labeling, it is hard to believe that there are no relationship between them.

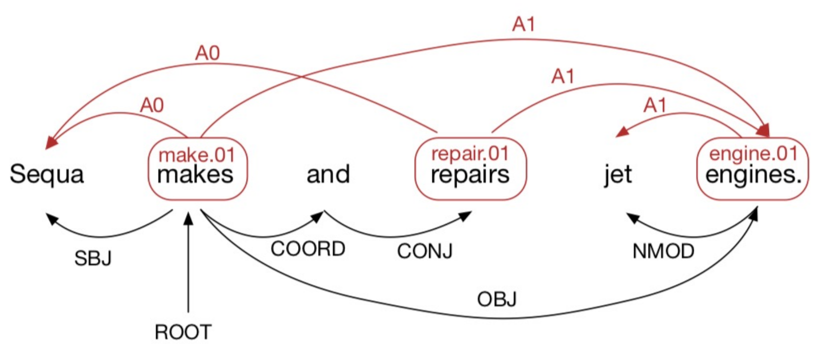

Figure 6: An example sentence annotated with semantic (top) and syntactic dependencies (bottom).

In Diego Marcheggiani and Ivan Titov's paper *Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling*[25], researchers found that the results of syntactic analysis and the results of semantic role labeling are mostly mirrored(Figure 6).

Based on the above observations, researchers proposed a graph-based convolutional network(GCN) method for semantic role labeling. Use GCN to encode a syntax-dependent tree, resulting in a potential feature representation of the word in the sentence. The article observes that the GCN layer is complementary with the LSTM layer: when the GCN layer and the LSTM layer are superimposed, their performance gets the latest state-of-the-art.

7.2 Experiment and Result

In order to test the behavior of the GCN model, we implemented a complete GCN model[26] using python and tensorflow[27]. If fact, in the open source project provided by the model, it gives a demo to user that help them to use the GCN with their own data. Follow the instructions, our research group use CoNLL 2003[28] as our dataet and try to use GCN to do Named Entity Extraction[29] here.

To feed the GCN model with our own data, we need to do preprocess on our raw dataset, CoNLL 2003 dataset contains thousands of sentences from newspaper and magazine. But GCN model need three matrices, first one is a N*N adjacency matrix that represent the relationship between each nodes, here N represent the number of nodes in a graph. The second one is a N*D matrix that represent the feature vector of each node, here D is the number of features of a node. The third matrix is an N*E matrix that reprensent the real class of each node, here E is the number of classes.

But in our project the nodes in a graph is actually words in a sentence. So we use Stanford Parser[30] to help us to get the relationship in a sentence. With the dependencies provided by Stanford Parser, the N*N adjacency matrix can be build easily.

Another tool we use in this project is word2vec[31], word2vec is a two-layer neural net that processes text. Its input is a text corpus and its output is a set of vectors: feature vectors for words in that corpus. With the help of this tool, for each word in a sentence, we can get a unique 300 dimensions vector to represent it, we use this vector as this word's feature vector and build the N*D matrix. In this project, we set the length of feature vector 300 so here the value of D is 300.

For the *N*E* matrix, the CoNLL 2003 gives the exact named entity of each word. Our research group use python to extract them and build them into matrix. After the preprocessing, we devided the dataset into 2 part, the first 50% as training set and the last 50% as testing set. Finally the program get 86.02% of accuracy on the testing set.

8 Conclusion

In this paper, we discussed some mature method to do semantic role labeling. In the last part, we have shown a GCN model to do named entity extraction on CoNLL dataset. We believe that with the development of new technology such as deep learning and artificial neural network, some useful research area such as semantic role labeling and named entity extraction in natural language processing, will get a lot of help with it.

9 Future scope

Although we got 86.02% of accuracy for our model,which is even higher than the state-of-art approach, we still want to think further about how to improve our model. One problem for our model is that all sentences used to train our model are from CoNLL 2003, which collected sentences from newspapers and magazines. We can try more different kinds of tagged datasets to train our model in order to increase its ability to adapt new datasets. In addition, We can also think that how the length and complexity of each sentence affect the accuracy of the model. Collecting longer and more complex sentences with correct tags might help us to train a more accurate model.

References

[1] Dan Jurafsky and James H Martin. *Speech and language processing*, volume 3. Pearson London, 2014.

[2] Sonit Singh. Natural language processing for information extraction. *arXiv preprint arXiv:1807.02383*, 2018.

[3] Janara Christensen, Stephen Soderland, Oren Etzioni, et al. Semantic role labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60. Association for Computational Linguistics, 2010.

[4] James H Martin and Daniel Jurafsky. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Pearson/Prentice Hall Upper Saddle River, 2009.

[5] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. *COLING/ACL-98*, pages 86–90, 1998.

[6] C. J Fillmore. Frames and the semantics of understanding. *Quaderni di Semantica*, page 222–254, 1985.

[7] P. N Johnson-Laird. Mental models. *Harvard University Press, Cambridge, MA.*

[8] R. C. Schank and R. P Abelson. Scripts, plans, and knowledge. *Proceedings of IJCAI-75*, page 151–157, 1975.

[9] Raymond R Smullyan. *First-order logic*, volume 43. Springer Science & Business Media, 2012.

[10] James Allen. *Natural language understanding.* Pearson, 1995.

[11] Melvin Fitting. *First-order logic and automated theorem proving.* Springer Science & Business Media, 2012.

[12] Jon Barwise. An introduction to first-order logic. In *Studies in Logic and the Foundations of Mathematics*, volume 90, pages 5–46. Elsevier, 1977.

[13] John F Sowa. Toward the expressive power of natural language. In *Principles of Semantic Networks*, pages 157–189. Elsevier, 1991.

[14] Lenhart K Schubert. Semantic representation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[15] Charles J Fillmore. The case for case. universals in linguistic theory, ed. by emmon bach and robert t. harms, 1-88. *New York. Holt, Rinehart & Winston. Syntax-Semantics Interface in Psych-verb Constructions (2341)*, 201:100363–76, 1968.

[16] Andrew Radford. *Transformational syntax: A student's guide to Chomsky's extended standard theory*, volume 29. Cambridge University Press Cambridge, 1981.

[17] Noam Chomsky. *Syntactic structures*. Walter de Gruyter, 2002.

[18] Noam Chomsky. *Aspects of the Theory of Syntax*, volume 11. MIT press, 2014.

[19] Charles J Fillmore. Toward a modern theory of case. 1966.

[20] Charles J Fillmore. Some problems for case grammar. 1971.

[21] Charles J Fillmore. The case for case reopened. *Syntax and semantics*, 8(1977):59–82, 1977.

[22] Walter Anthony Cook. *Case grammar theory*. Georgetown University Press, 1989.

[23] Charles J Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32, 1976.

[24] Barbara Landau, Lila R Gleitman, and Barbara Landau. *Language and experience: Evidence from the blind child*, volume 8. Harvard University Press, 2009.

[25] Diego Marcheggiani and Ivan Titov. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*, 2017.

[26] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[27] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[28] Erik F Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*, 2003.

[29] Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S Weld, and Alexander Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134, 2005.

[30] Marie-Catherine De Marneffe and Christopher D Manning. The stanford typed dependencies representation. In *Coling 2008: proceedings of the workshop on cross-framework and cross-domain parser evaluation*, pages 1–8. Association for Computational Linguistics, 2008.

[31] Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.