Word Sense Disambiguation Using Graph-based Semi-supervised Learning

Rie Yatabe and Minoru Sasaki

19nm732r@vc.ibaraki.ac.jp
minoru.sasaki.01@vc.ibaraki.ac.jp

Introduction

Word sense disambiguation (WSD)

- The task of deciding the appropriate meaning of a target ambiguous word in its context

Supervised learning approach

- It has been the most successful
- However, problem of this approach is the lack of sufficient labelled training examples of specific words due to costly annotation work
- Moreover, most supervised WSD methods suffer from small differences of examples

Aim

Semi-supervised classification method with graph convolutional neural network

- This method can jointly train the embedding of an example to predict the sense label of the example and the neighbours in the graph.
- It is possible to incorporate information obtained from unlabelled examples without assigning a sense label to unlabelled examples.
- It is not clear what kind of features are effective

We employ a graph convolutional neural network for semi-supervised WSD system to incorporate information obtained from unlabelled examples.

We show that it is possible to distinguish between two similar examples with different sense labels using the proposed method.

Related Works

WSD using neural network

- (Kågebäck and Salomonsson, 2016)
  - A Bidirectional Long Short-Term Memory (Bi-LSTM) to encode information of both preceding and succeeding words within the context of a target word.
- (Yuan et al., 2016)
  - An LSTM language model to obtain a context representation from a context layer for the whole sentence containing a target word.
- (Raganato et al., 2017)
  - WSD as a neural sequence labelling task and constructed a sequence learning model for all-words WSD.

Related Works

WSD using semi-supervised learning

- (Yarowsky, 1995)
  - A bootstrapping model that only has a small set of sense-labelled examples that gradually assigns appropriate senses to unlabeled examples.
- (Taghipour and Ng, 2015) and (Yuan et al., 2016)
  - A semi-supervised WSD method to use word embeddings of surrounding words of the target word and showed that the performance of WSD could be increased by taking advantage of word embeddings.
- (Fujita et al. 2011)
  - A semi-supervised WSD method that automatically obtains reliable sense labelled examples using example sentences from the iwanami japanese dictionary to expand the labelled training data.
  - Then, this method employs a maximum entropy model to construct a WSD classifier for each target word using common morphological features (surrounding words and POS tags) and topic features.
  - Finally, the classifier for each target word predicts the sense of the test examples. They showed that this method is effective for the SemEval-2010 Japanese WSD task.

Related Works

WSD based on graph-based approaches

- (Niu et al., 2005)
  - A label propagation-based semi-supervised learning algorithm for WSD, which combines labelled and unlabelled examples in the learning process.
- (Yuan et al., 2016)
  - A label propagation (LP) for semi-supervised classification and LSTM language model.

WSD method using graph-based semi-supervised learning

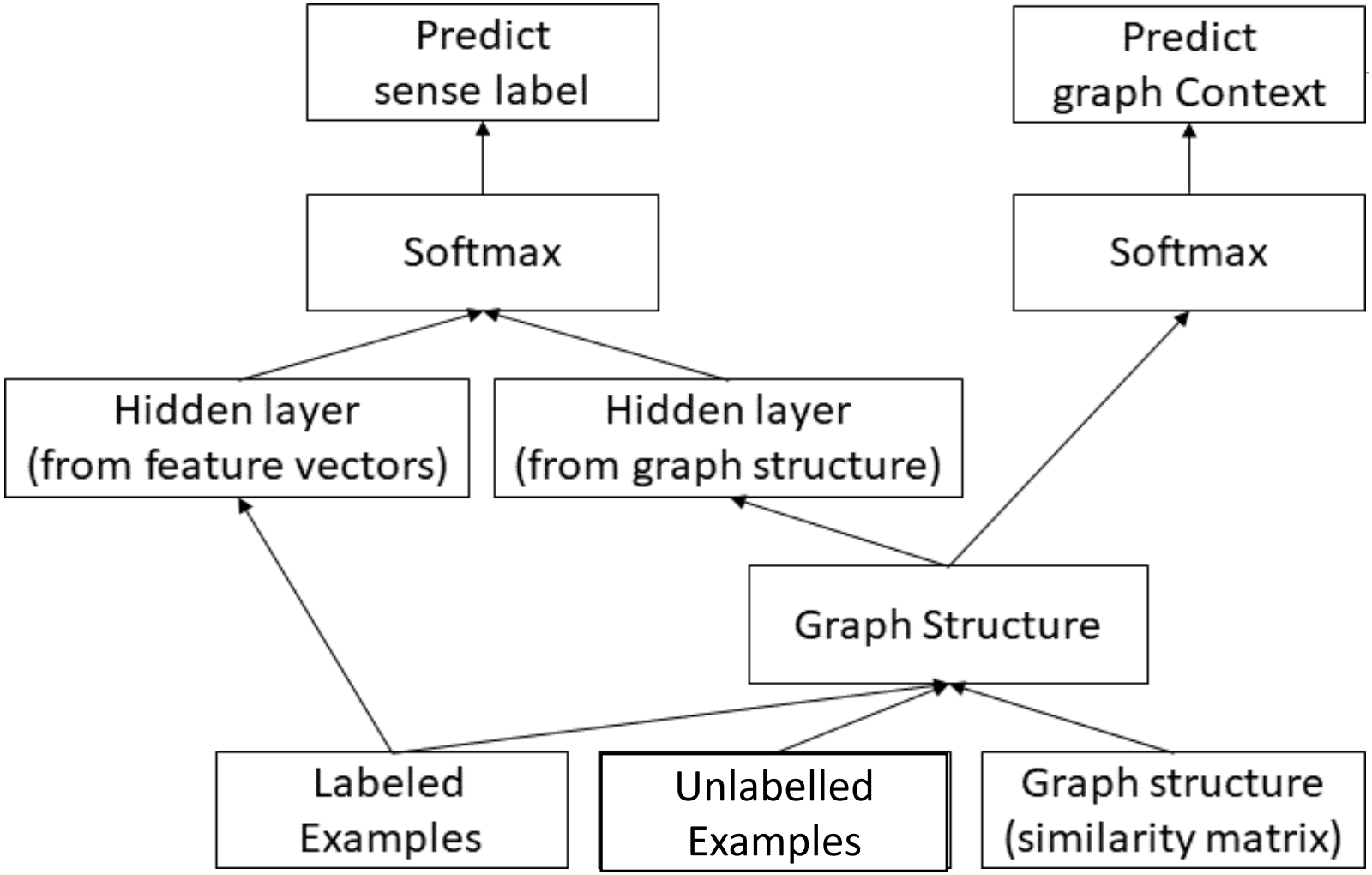
Graph convolutional neural network（Planetoid）

- Semi-supervised WSD method

Given a graph structure and feature vectors

- We learn an embedding space to jointly predict the sense label and the context of the graph.

WSD model

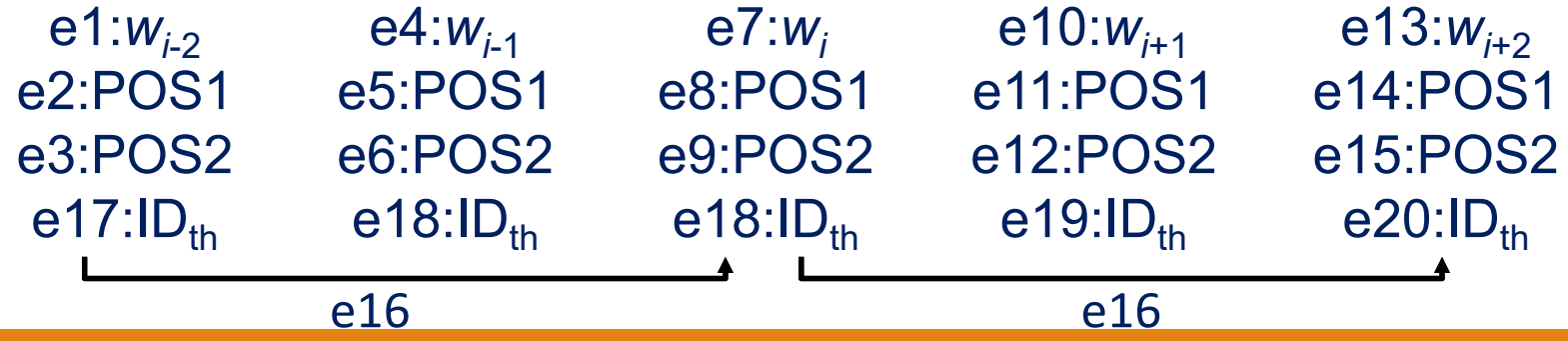

Predict
sense label
Predict
graph Context
Softmax
Softmax
Hidden layer
(from feature vectors)
Hidden layer
(from graph structure)
Graph Structure
Labeled
Examples
Unlabelled
Examples
Graph structure
(similarity matrix)

Input data (Lexical Features)

Twenty features (BF) for the target word w_i

- Features extracted from a context around the target word.
  - POS1: Part-of-Speech of the word
  - POS2: Subcategory of POS1
  - ID_{th}: Thesaurus ID number of the word

Example: …, w_{i-2} , w_{i-1} , w_i , w_{i+1} , w_{i+2} , …



Input data (Local Collocations)

Additional local collocation (LC) features

- we use bi-gram, tri-gram, and skip-bigram patterns in the three words on either side of the target word like “It Makes Sense”.
  - Skip-bigram is any pair of words in an example order with arbitrary gaps.

A context of word w_i is represented as a vector of these features, where the value of each feature indicates the number of times the feature occurs.

Input data (Graph Structure)

The relation between the training and the unlabelled data

Each node is an example and an edge is the similarity between nodes.

To construct a graph for all examples,

- Two nodes are connected if they are nearest neighbour and if their similarity is not less than the threshold 0.9.

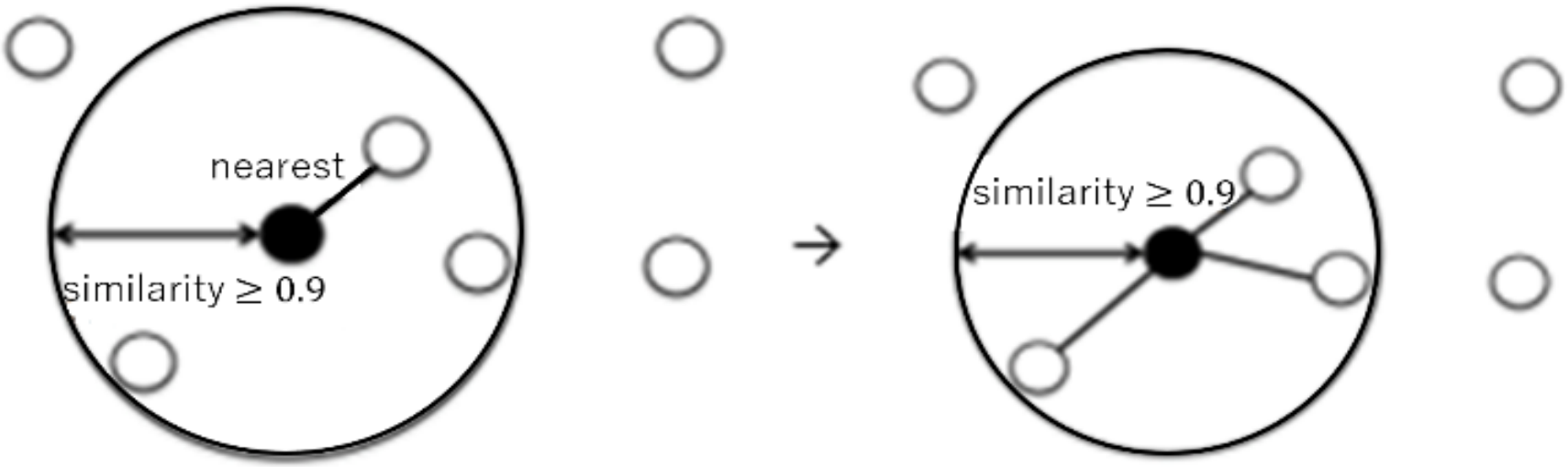
The basic idea behind this is that two nodes tend to have a high similarity if the corresponding contexts of the target word are similar.

Similarity calculation method

Jaccard similarity and Cosine similarity

- Jaccard similarity *J*(*A*, *B*) is the ratio of the number of words in common between the two sets *A* and *B*.

$$J(A,B) = |A \cap B| / |A \cup B| \text{ , } (0 \leq J(A,B) \leq 1)$$



Similarity calculation method

Mutual k-nearest neighbour graph

- Edge between two nodes is connected if each of the nodes belongs to the k-nearest neighbours of the other.

The edges with the highest similarity between nodes are also added to the graph structure obtained by the mutual k-nearest neighbour graph.

Experiments(Data Set)

Semeval-2010 Japanese WSD task data set

- 50 target words
  - Comprising 22 nouns, 23 verbs, and 5 adjectives
- 50 training
- 50 test instances

Unlabelled example data

- Balanced Corpus of Contemporary Written Japanese (BCCWJ)

Experiments (five types of features)

ipadicBF
- Word segmentation using dictionary "ipadic" for extracting BF features

UniDicBF
- Word segmentation using dictionary "unidic" for extracting BF features

UniDicBF+IWA
- UniDicBF and additional examples from Iwanami's dictionary

UniDicBF+LC :
- UniDicBF and additional local collocation features

UniDicBF+LC+IWA
- UniDicBF, additional local collocation features and additional examples from Iwanami's dictionary

Results

Features	Proposed Method	SVM	ME
ipadicBF	77.24	77.28	-
UniDicBF	**77.76**	76.8	76.56
UniDicBF+IWA	76.68	77.84	76.76
UniDicBF+LC	75.88	75.72	74.92
UniDicBF+LC+IWA	76.28	77.36	76.52

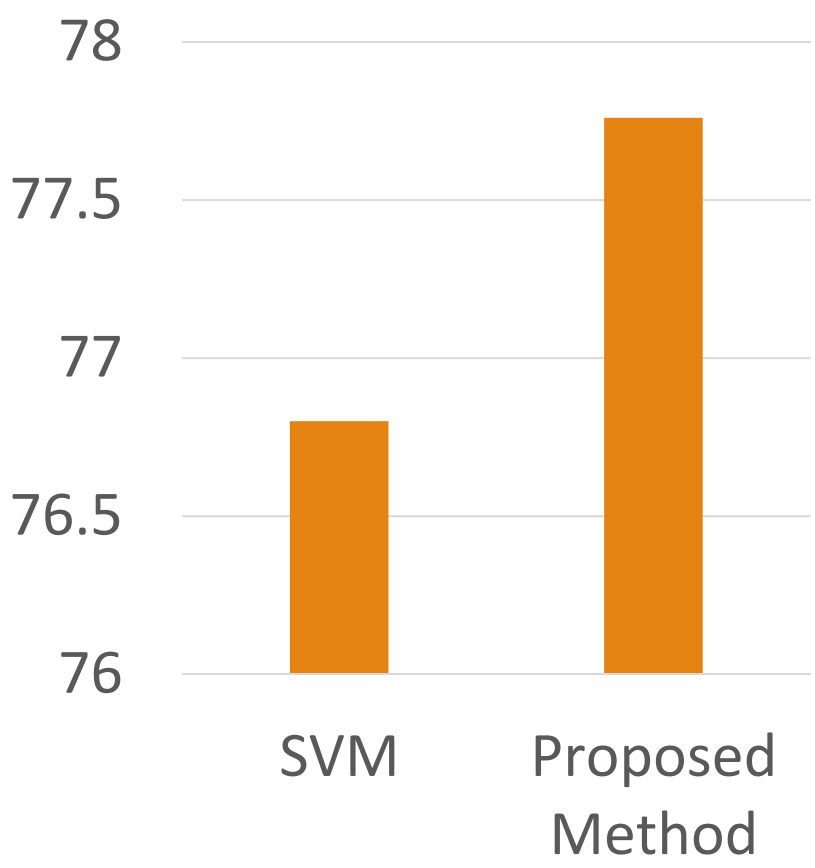
Discussions

1) Comparison with SVM
2) Comparison by different dictionaries (ipadic, UniDic)
3) Comparison of similarity
4) Additional examples from Iwanami's dictionary
5) Additional local collocation(LC) features
6) Comparison of the previous semi-supervised method

(1) Comparison with SVM

The proposed method is higher than the SVM classifier

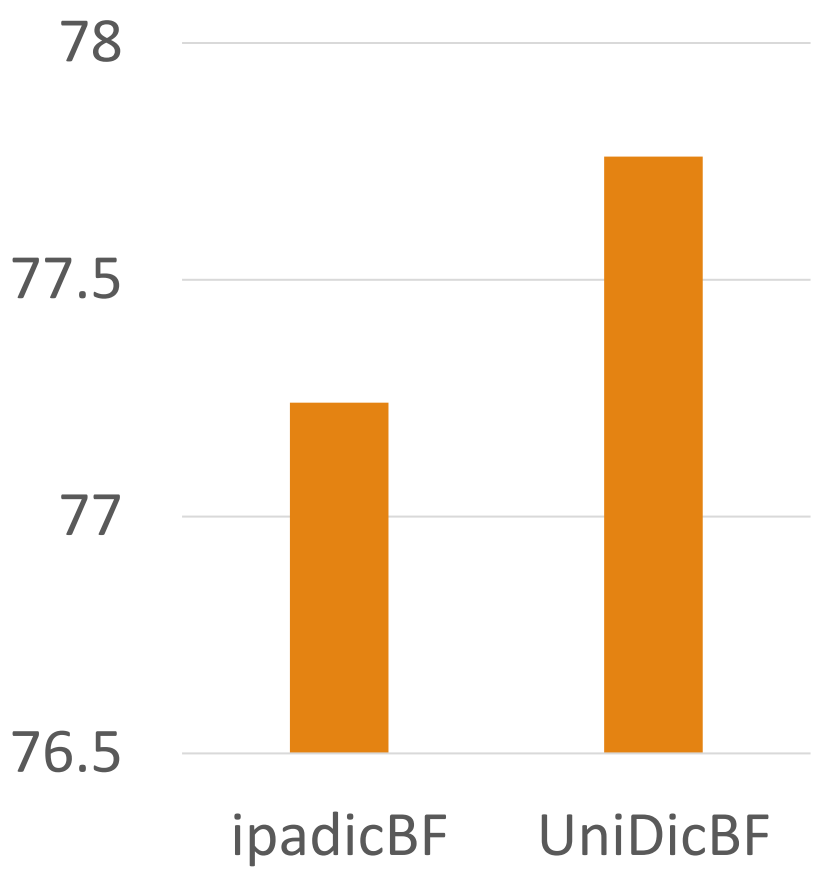
- It can cope with the lack of labelled data for WSD.



(2) Comparison by different dictionaries

The features of UniDicBF are more effective than the features of ipadicBF

By using UniDic, it is possible to obtain more consistent word segmentation for Japanese sentences of many genres than using ipadic.

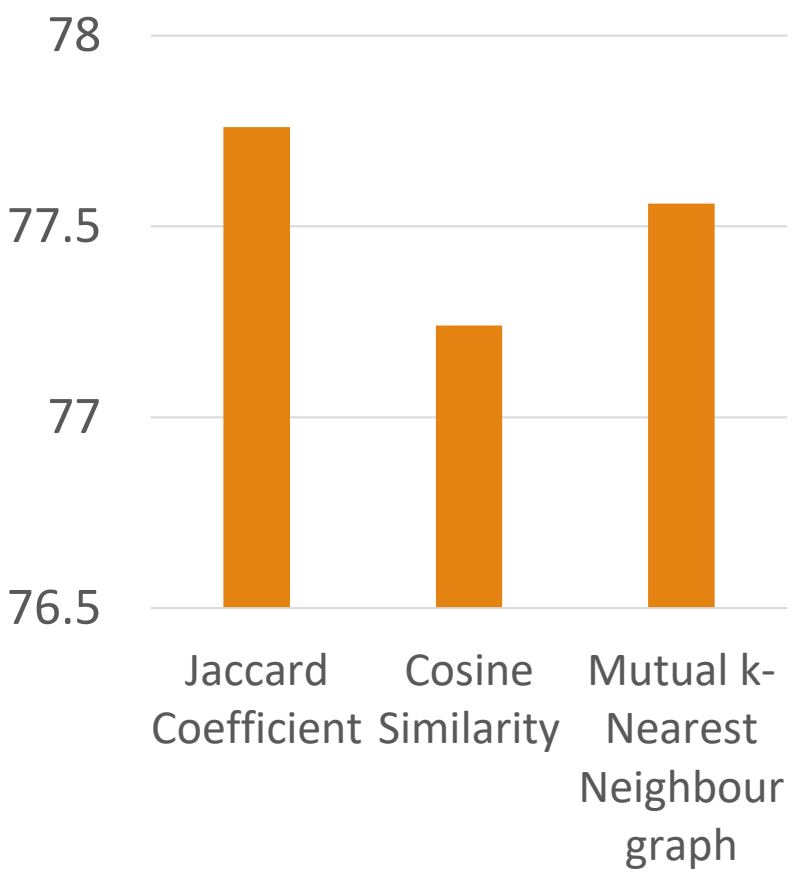


(3) Comparison of similarity

The Jaccard coefficient measure is the most effective of all similarity measures.

Thus, if available features are small and dense,

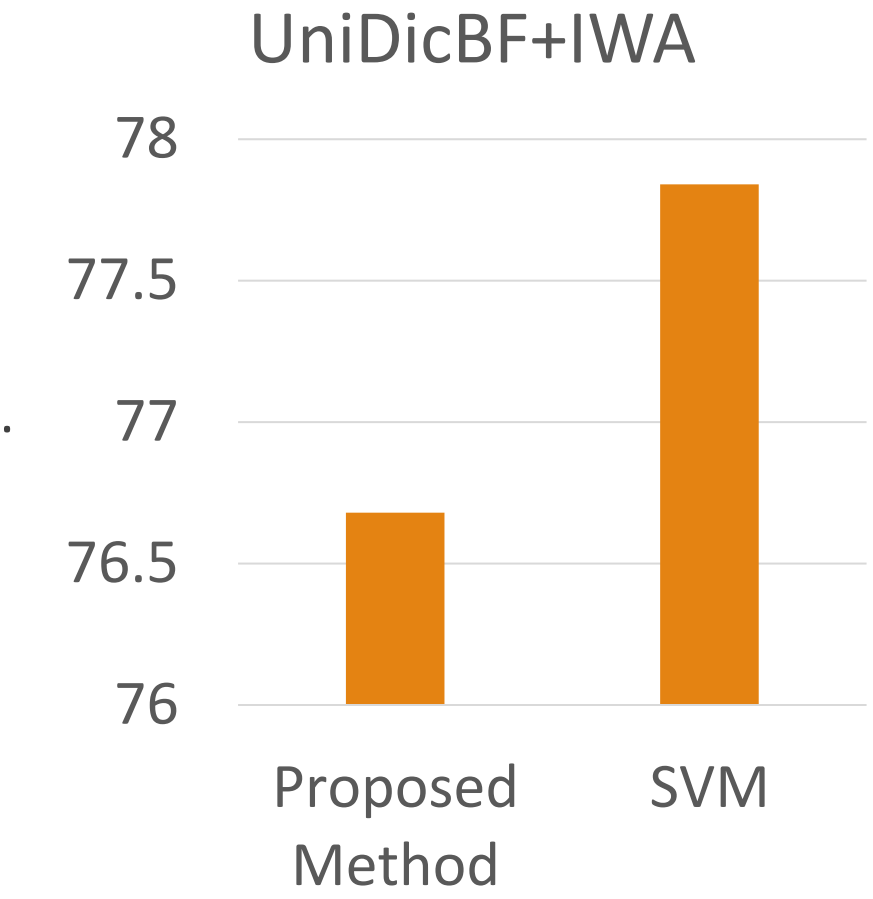
- The Jaccard coefficient is considered to be suitable for the construction of the graph structure.



(4) Additional examples from Iwanami's dictionary

The proposed method does not perform better than SVM classifier.

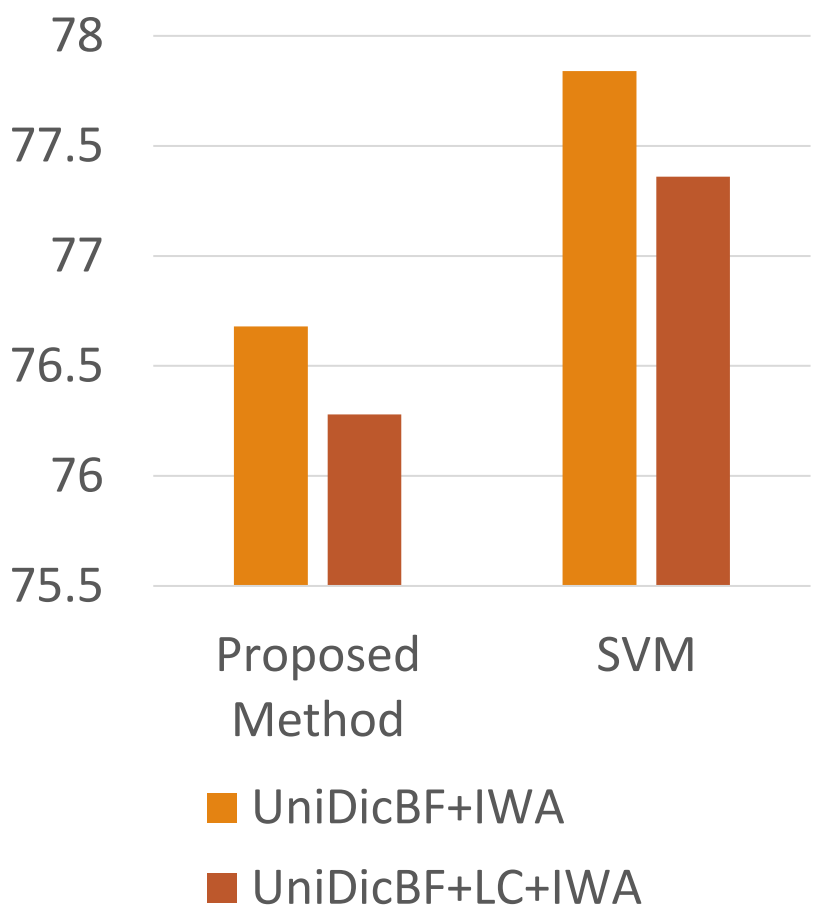
- Example sentences of the Iwanami's Japanese dictionary tend to be connected to short example sentences in the corpus.
- Therefore, examples of Iwanami's Japanese dictionary tend not to be effective in constructing a graph structure.



(5) Additional local collocation(LC) features

UniDicBF+LC+IWA does not perform better than that using UniDicBF+IWA.

Many examples of the Iwanami's Japanese dictionary are short so that the LC features are not so effective for both methods.

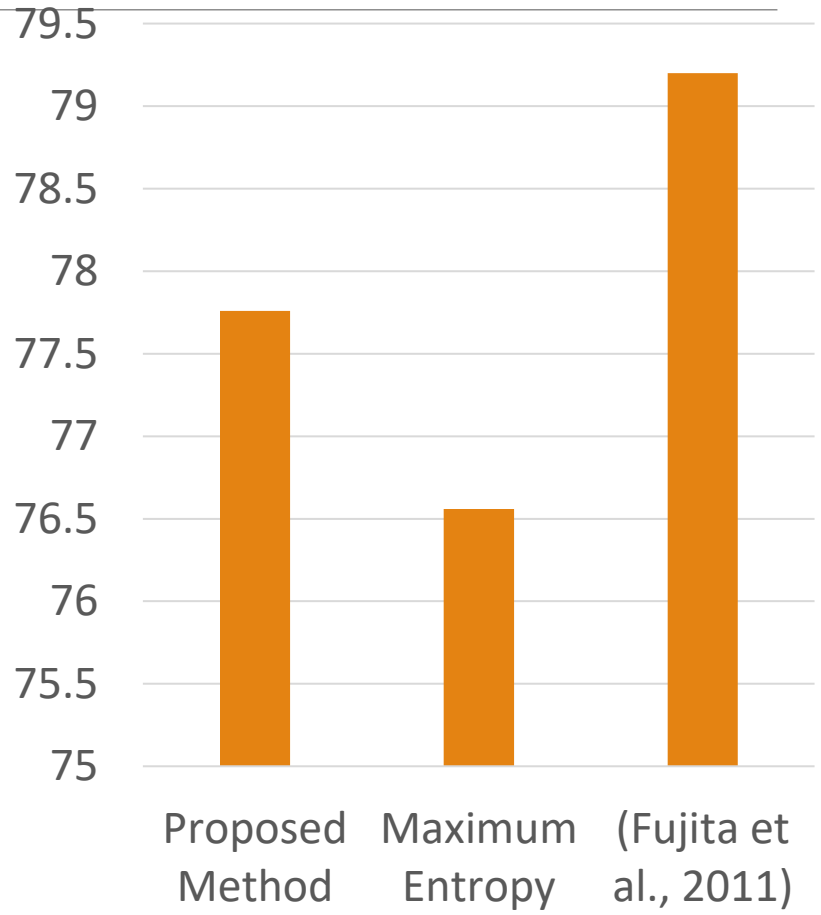


(6) Comparison of the previous semi-supervised method

The proposed method does not perform better than the previous method.

- The previous method uses the basic form (lemma) of the word and the Hinoki Sensebank in addition to the BF features without thesaurus IDs.

Therefore, using the UniDicBF features for both methods for a fair comparison, the proposed method performs better than the previous method.



For the target word "教える (oshieru),"

There exist five examples that have similar context, but they have different meanings in the test data.

Using the SVM classifier, the classifier could not classify these examples correctly.

The proposed method was able to classify one test example correctly out of the five examples.

- To construct the graph structure, the proposed method connects these five examples by the edge.

We consider that it is possible to distinguish two examples because the edge between these two examples has been deleted by repeating training with the training examples.

Conclusion

We proposed a semi-supervised method using a graph convolutional neural network for the WSD task.

- The proposed method performs better than the previous supervised method and the morphological features obtained by the UniDic short-unit dictionary is effective for the semi-supervised WSD method.
- Moreover, the Jaccard coefficient is the most effective measure among three measures to construct a graph structure.
- Moreover, for the problem with small difference such as examples that have similar context but have different meanings, the proposed method improved the performance of WSD.
  - Therefore, if we can distinguish such example sentences, we consider the performance of WSD systems improved.

Future work

we would like to explore methods

- To construct an effective graph structure by using paraphrase information.
- And the dependency analysis technique.
- The effective filtering method for unlabelled data.

In addition, we would like to develop a method to use the example sentences of the Iwanami's Japanese dictionary effectively.

Thank You!

For any question or comment, please contact

19nm732r@vc.ibaraki.ac.jp

or

minoru.sasaki.01@vc.ibaraki.ac.jp