Dialogue Act Classification in Reference Interview Using Convolutional Neural Network with Byte Pair Encoding

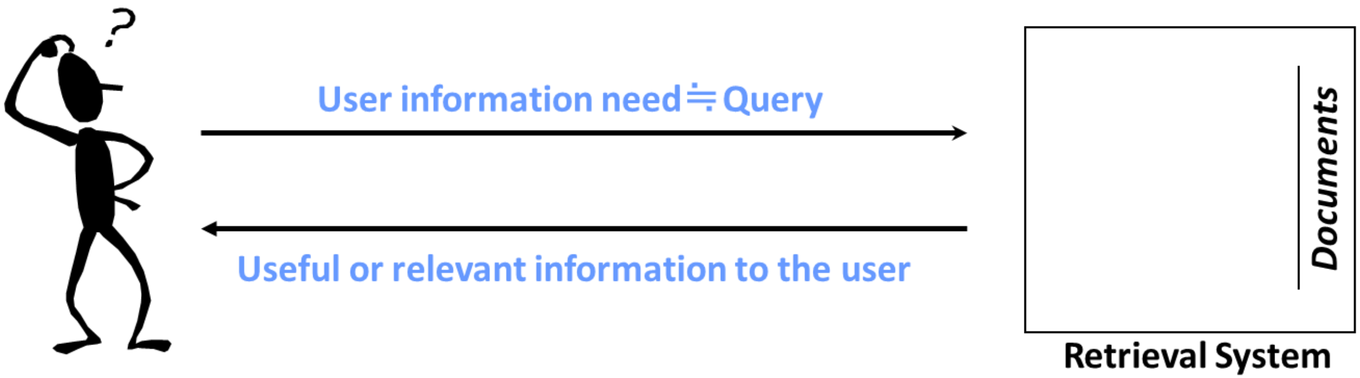
Seiya Kawano, Koichiro Yoshino, Yu Suzuki, Satoshi Nakamura
Nara Institute of Science and Technology (NAIST), Japan
http://isw3.naist.jp/~seiya-ka/index.html

Nakamura SRG
Augmented Human Communication (AHC) Laboratory

NAIST

Background

☐ Demand for conversation based seacrh



✓ **Many users cannot clarify their “information-needs”**

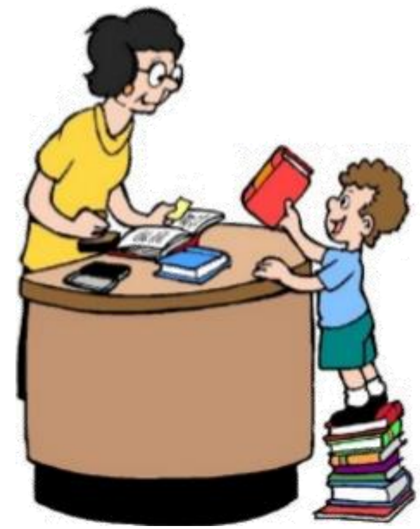
✓ **Clarification of the requirement through interactions is important**

➢ e.g. confirmation, asking for users “motivation” or “background”

☐ How to model the dialogue strategy for clarification?

➢ **Focus on the behaviour of human expert such as librarian**

Reference Interview



☐ Reference Service

- ➢ Information navigation service in library
- ➢ Library users can ask the s at the librarian for helping to find information

☐ Reference Interview

- ➢ Structured interview for clarify the information-needs
- ➢ librarian works with the user to clarify their ambiguous needs

Improves the accuracy of information navigation

[Ross et al. 2002]

Example of Reference Interview

question(ambiguous)

I'm having trouble finding the **volumes of postal mail** throughout the 20th century

Confirm: type of information

Hi, so you are trying to find some **measure of the volume of mail** sent through the US postal service

feedback: yes

Yes

confirm: search history

Ok, Let me see if I can find something
Can you tell me what you have done already?

feedback: search history

I found one site, about a week ago, but I just realized it's more recent data and the paper is for '20th century' history, so I want to try to focus on **statistics**

answer

OK, let me look a little.... please hold
Do you think this page would help?
https://www.usps.com/cpim/pub100.htm

Feedback: positive

Wow, this is perfect!!
Thanks a lot

follow-up

So, will that answer your question?

closing

Yes thanks.
Bye.

Utterances: Librarian

Library User

Toward Modeling the Reference Interview

☐ How do we model the dialogue strategy?

① Abstract of the utterance: e.g. dialogue act, dialogue state

✓ **Tracking and predicting the speaker's intentions**

② Modeling the dialogue strategy with reinforcement learning

✓ Understanding the dynamics of reference interview

✓ Imitate the librarian behaviors

③ Construct the response generation module

✓ Corresponds to each response action of librarian

➢ **Toward the dialogue management**

Focus the dialogue act classification task

Available Corpus of Reference Interview

☐ QuestionPoint Transcripts [Radford et al. 2011]

- ➢ QuestionPoint: chat based reference service
- ➢ 600 dialogue sessions, 12634 utterances (preprocessed)
- ➢ Personal information are anonymized

☐ Dialogue act tag in reference interview [Inoue 2013]

- ➢ Defined the two intent levels dialogue act
- Dialogue Act Function; DAF (5 class)
- Dialogue Act Domain; DAD (19 class)

Dialogue Act in Reference Interview

Table. 5 Class DA Categories (DAF) [Inoue 2013]

No	Dialogue Act Function	Count	Description
1	Information Provision	2858	To provide information
2	Information Request	758	To request information
3	Task Management	689	To assign or commit to tasks
4	Social Relationship Management	593	To manage socio-emotional aspects of communication
5	Communication Management	430	To manage physical aspects of communication

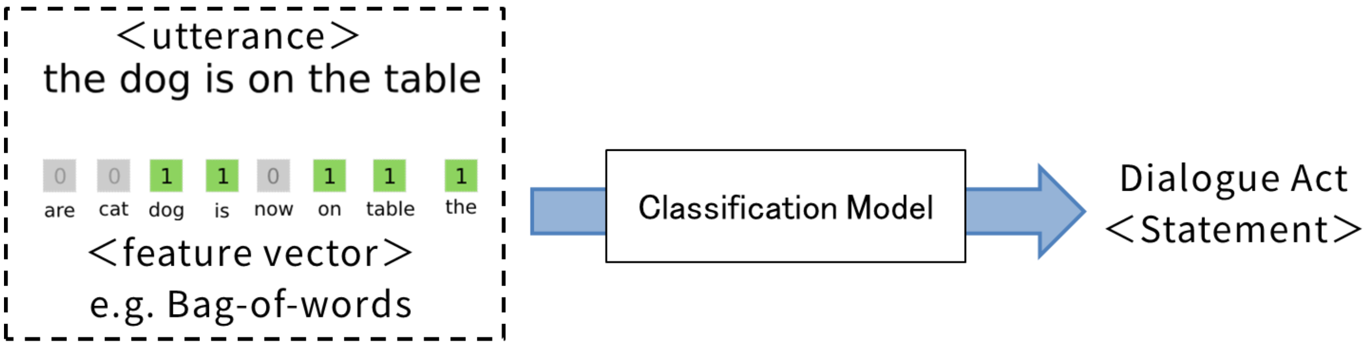
Dialogue Act in Reference Interview

Table. 19 Class DA Categories (DAD) [Inoue 2013]

No.	Function	Domain	Count
1	Information Transfer	Information Problem	1203
2		Search Process	672
3		Information Object	111
4		Feedback	111
5		Other	397
6	Task Management	Librarian’ s Task	126
7		User’ s Task	96
8		Other	6
9	Social Relationship Management	Greeting	247
10		Valediction	45
11		Exclamation	21
12		Apology	21
13		Gratitude	423
14		Downplay	65
15		Closing Ritual	32
16		Rapport Building	82
17	Communication Management	Channel Checking	67
18		Pausing	219
19		Feedback	314

Problem of DA Classification

□ Supervised dialogue act (DA) classification



□ A problem existing DA classification approach

- ➢ Requires enough training data with labels
- **Sparseness: Lack of training data for rare and unusual words**
- It is critical in open-domain task such as reference interview

We need handle OOVs

Subword Approach

☐ Words can be divided into Subwords

➢ **Can reduce OOVs**

- Character-ngram
- Byte Pair Encoding, etc.
  [Gage et al. 1994, Sennrich et al. 2016]

Unquestionably

Un + **question(stem)** + **able**+ **ly**

prefix **stem** **suffix**

Table. Variation of units for text tokenization

Units	Reduce OOVs	Consideration of word structure
Word	×	×
Characters	◎	×
Character-ngram	○	△
Byte Pair Encoding	◎	○

Byte Pair Encoding Compression

Bottom-up character merging

- **Recursively merges most frequent consecutive symbols into one symbol**
- Starting point: character-level representation
- Hyper parameter: when to stop the merge operation
  vocabulary size = number of merges + unique characters

- E.g. training data = {_low, _lowest, _newer, _wider}
- ✓ Start = {_ l o w, _ l o w e s t, _ n e w e r, _ w i d e r }

1. _ l → _l
2. _l o → _lo
3. _lo w → _low
4. e r → er

- **Can segment the any text using merge operation rule**
- e.g. _lowly
- ✓ Tokenized: _low | l | y

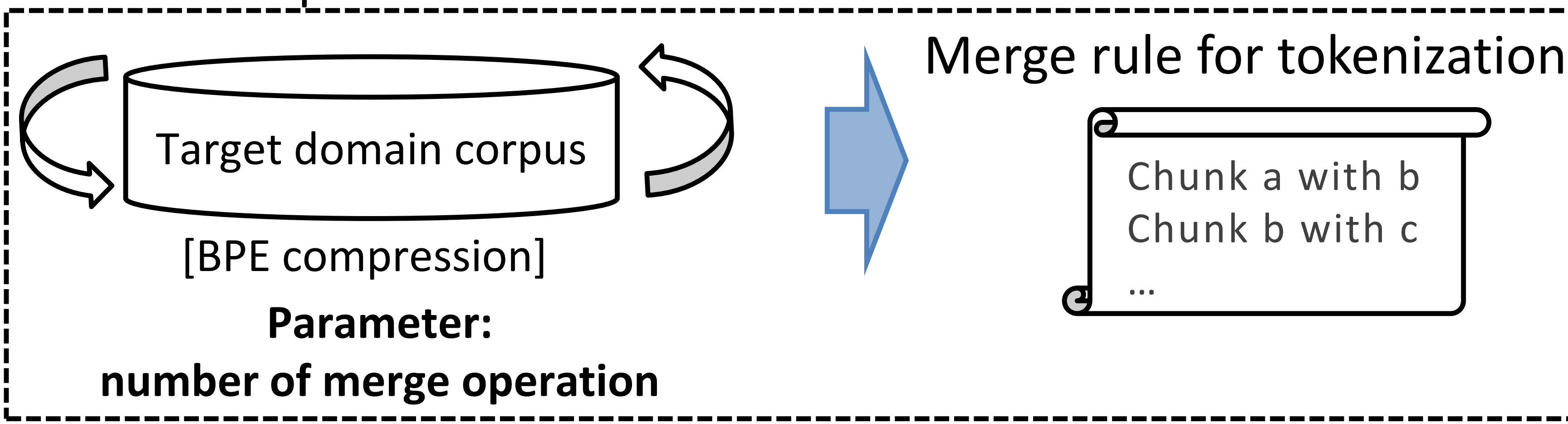
Solution

☐ Handle OOVs & Build an better vocabulary units

➢ Apply the Byte Pair Encoding for subword tokenization

- **BPE is regarded as a domain-dependent feature extractor**

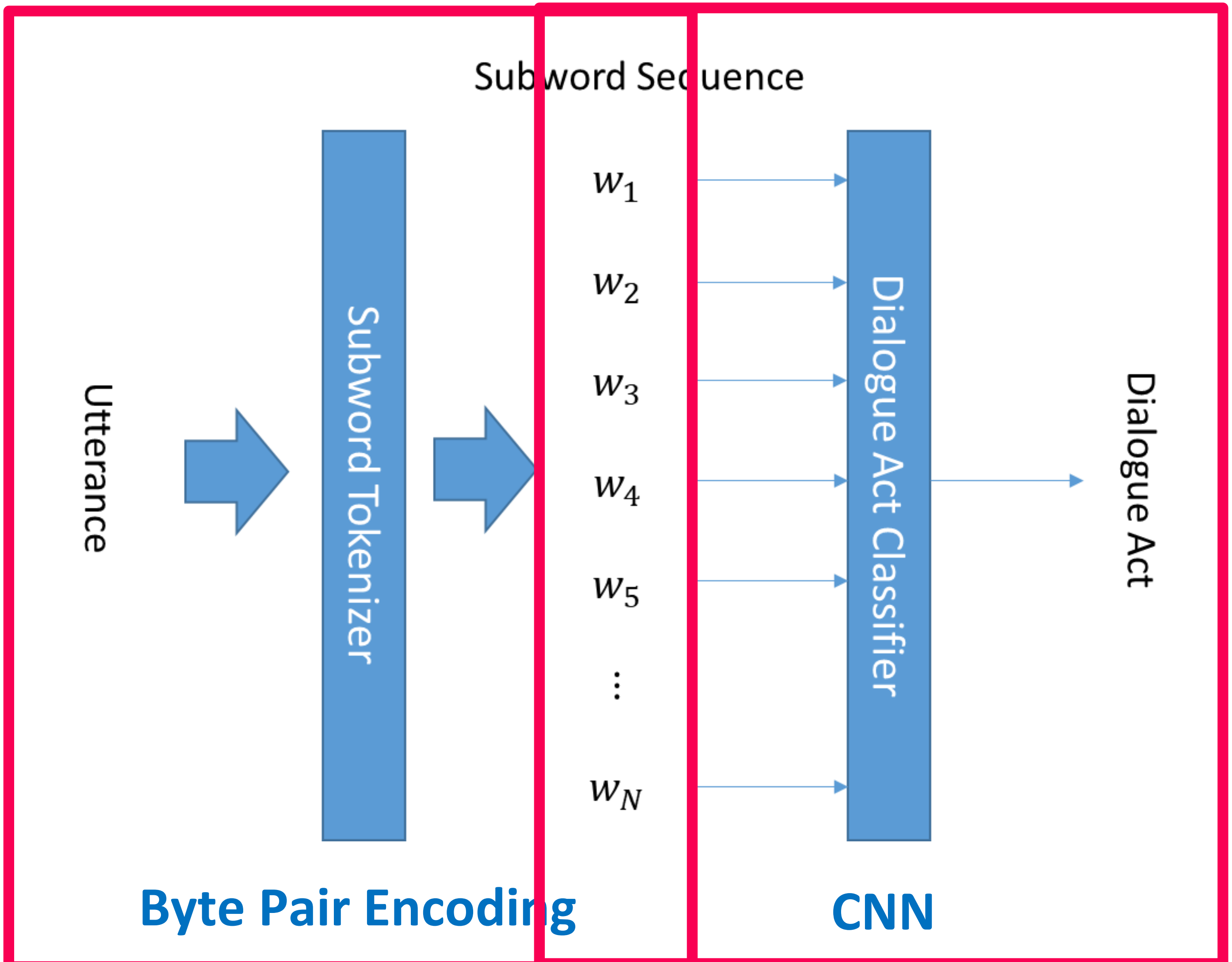
Domain adaption of Subword tokenizer



☐ Adaptation to neural dialogue act classification model

- Applied for a simple CNN-based classifier

Diagram of Proposed Method



CNN based DA Classifier

☐ BPE-Unit-Level Convolutional Neural Network

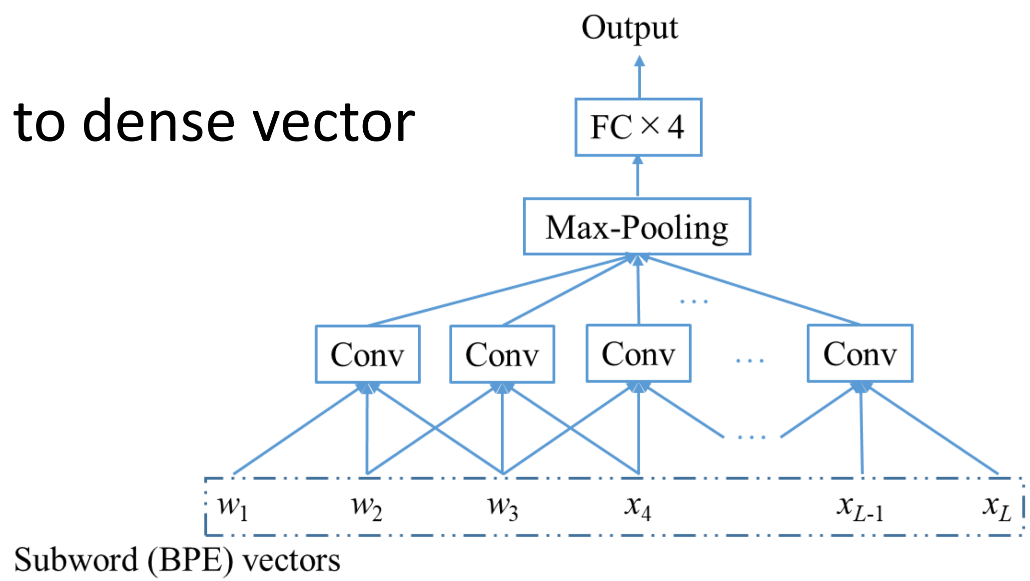
➢ Embedding Layer

- Convert one-hot BPE vector to dense vector

➢ 1D Convolution layer

➢ Global Max-Pooling Layer

- pool size = input length



☐ Character vs. word vs. Subword (BPE-Unit)

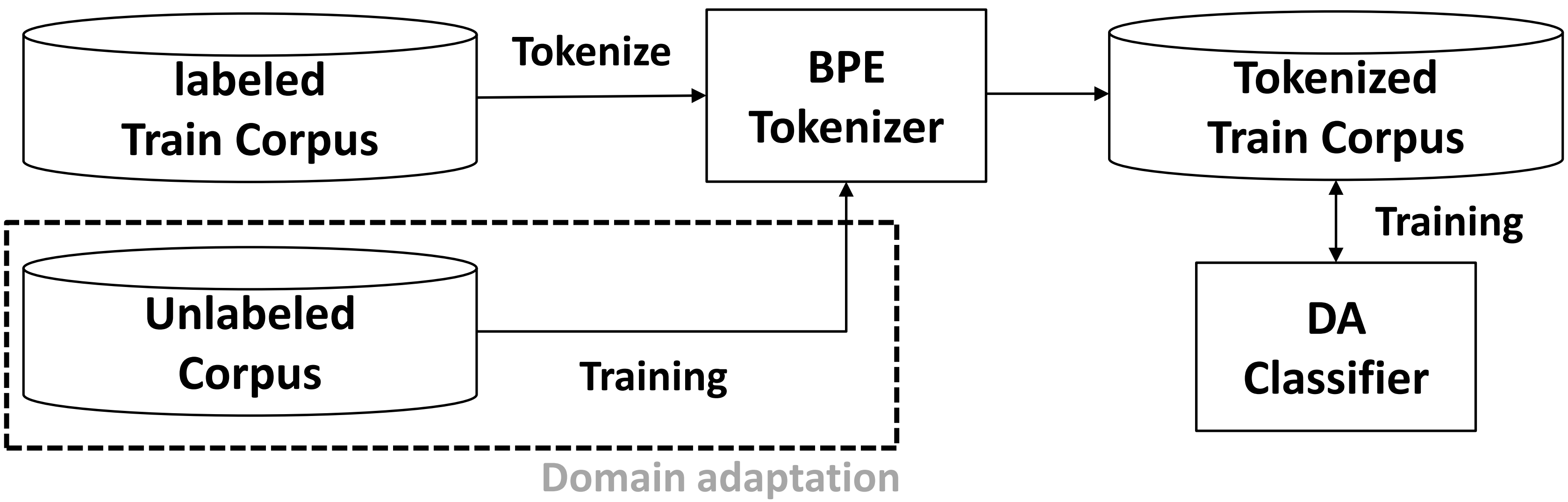
- Comparison with a simple model
- Adapt the simple CNN

Figure. The Input Generation to CNN

Experimental Setup

☐ Dataset: QuestionPoint transcripts

- Labeled 5,327 utterances, unlabeled 7, 307 utterances



☐ Evaluation

- Predict the 5 class & 19 class DA categories [Inoue 2013]
- 10-fold cross validation with paired t-test (5,327 utterances)

Comparison

☐ DA classifiers

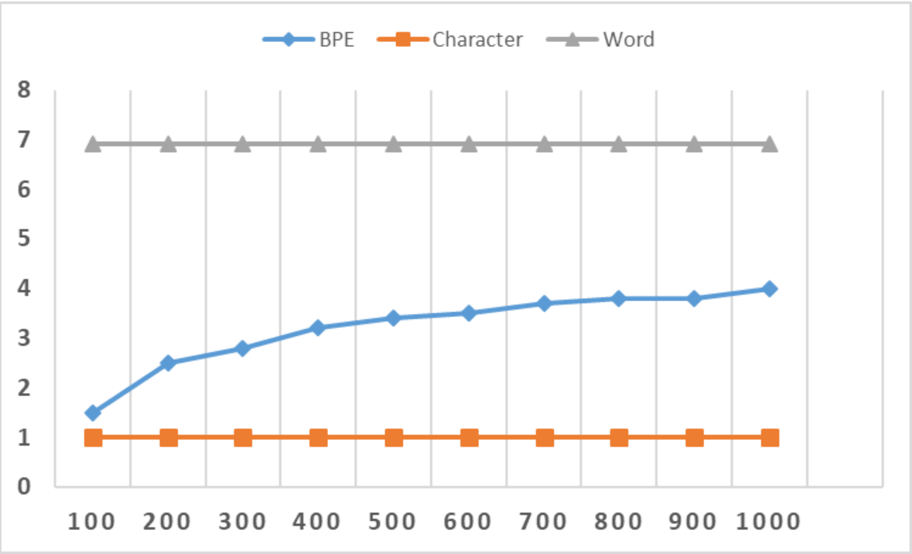
Method	Unit
BPE-Unit-Level CNN	BPE-Unit
Character-Level CNN	Character
Word-Level CNN	Word
Word-Level LSTM	Word
MLP	Word
MLP w/o addition	Word
RF	Word
RF w/o additional	Word

☐ Baseline features of RF & MLP

- Basic: bag-of-words (BoW), bag-of-bigrams (BoW)
- Additional : speaker, length of tokenized utterance, order of utterance

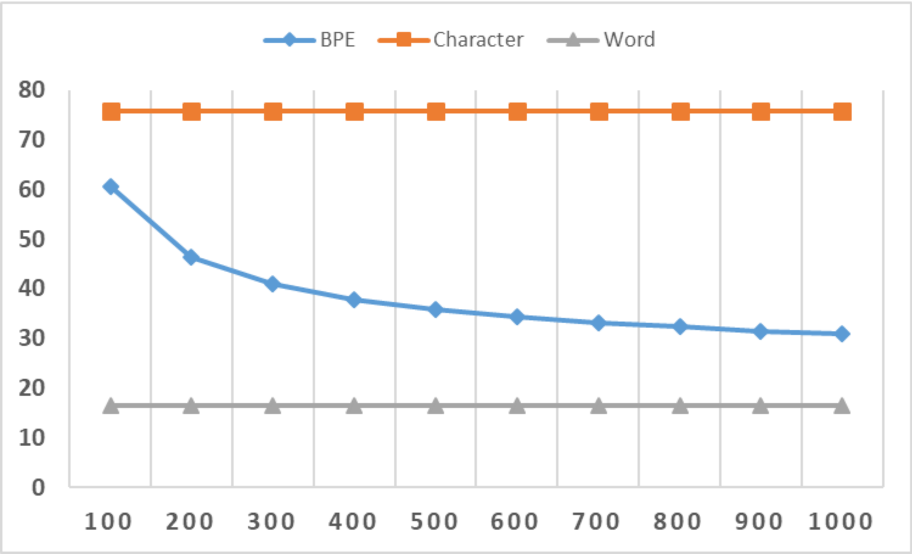
Statistics of BPE Tokenization

☐Number of characters per token



Vocabulary size of BPE

☐Length of tokenized utterances



Vocabulary size of BPE

☐Average number of OOVs

- ➢ Word-Unit
- • **334 OOVs**

- ➢ Character & BPE-Unit
- • **Less OOVs (< 2)**

Experimental Results

Method	DAF (5 class)	DAD (19 class)
BPE-Unit-Level CNN (v=200)	**0.8684 *****	**0.7256 ***
Character-Level CNN	0.8538	0.7124
Word-Level CNN	0.8438	0.6937
Word-Level LSTM	0.8286	0.6745
MLP	0.8515	0.7145
MLP w/o additional	0.8498	0.7119
RF	0.8367	0.7008
RF w/o additional	0.8292	0.6790

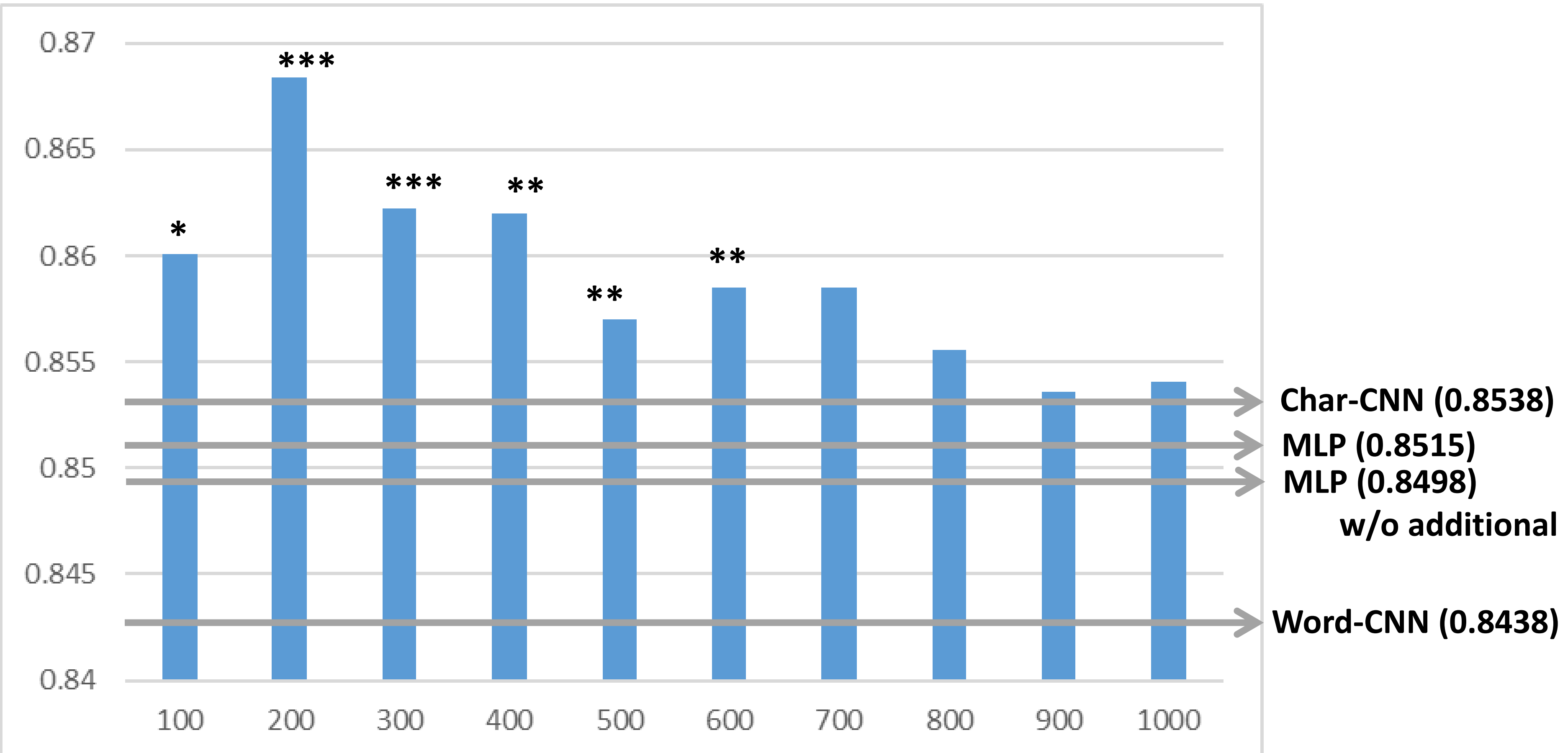
best of baselines

＊ $p < 0.05$ ＊＊ $p < 0.01$ ＊＊＊ $p < 0.001$ ： Comparison with MLP

BEP-Unit Level CNN ＞Character-Level CNN ＞＞Word-Level CNN

Results of Several BPE Settings

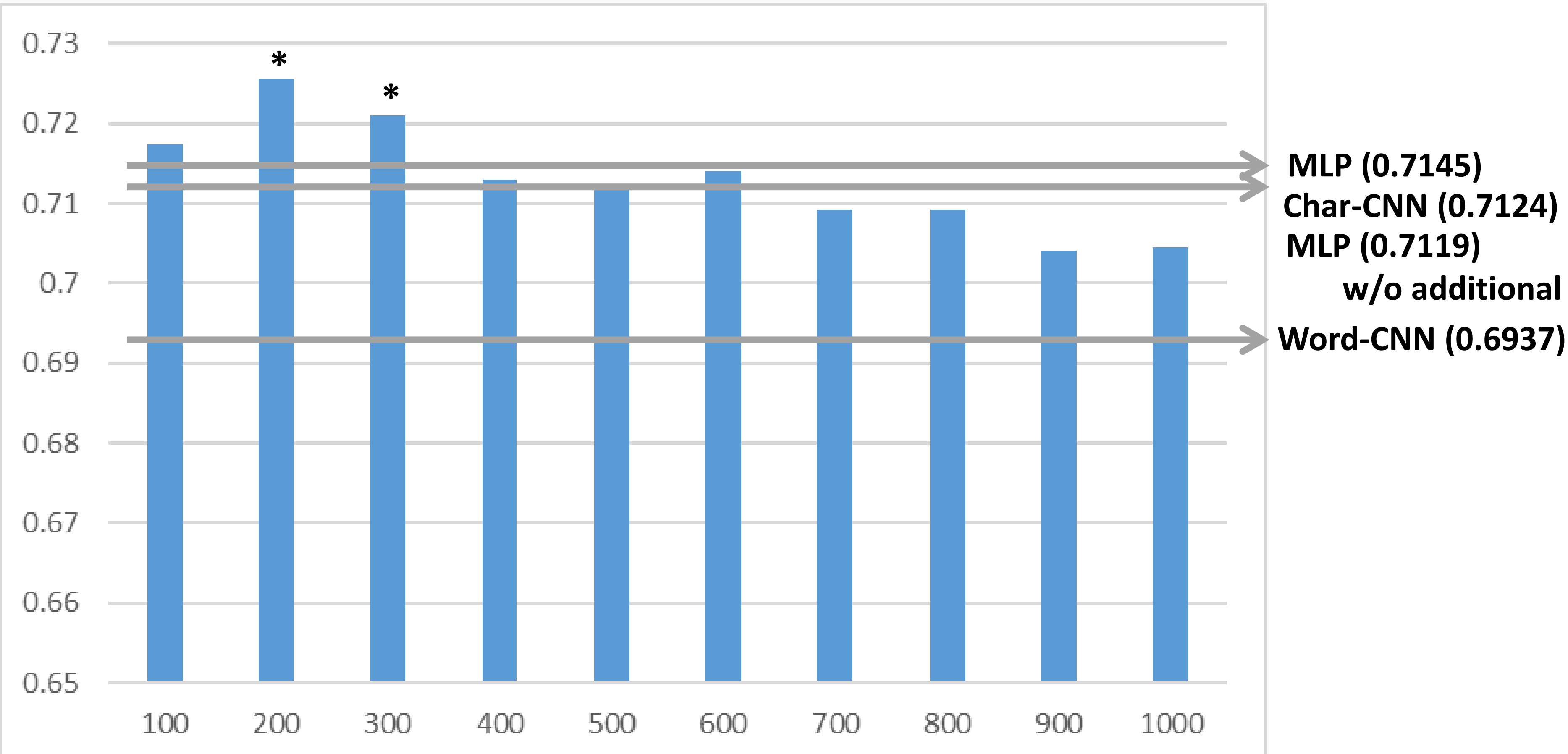
□ DAF (5 class) results



＊ $p < 0.05$ ＊＊ $p < 0.01$ ＊＊＊ $p < 0.001$ ：Comparison with MLP

Results of Several BPE Settings

□ DAF (19 class) results



＊ $p < 0.05$ ＊＊ $p < 0.01$ ＊＊＊ $p < 0.001$ ：Comparison with MLP

Conclusion

- **We proposed dialogue act classification model in reference interview using CNN with Byte Pair Encoding**
  - Achieved the best performance without complicated feature engineering and additional features

- **CNN with Character vs. word vs. Subword (BPE-Unit)**
  - BPE-Unit > Character >> Word
  - **BPE-Unit-Level CNN improved accuracy than Character-Level CNN**
  - ✓ Possibility of eliminating sparseness & acquiring the better unit

Future work

☐ More Improve the dialogue act classification model

- ➢ Automatic parameter decision in BPE
- ➢ Combining some additional information (e.g. dialogue history)

☐ Improve the current annotation scheme

- ➢ e.g. apply the ISO-24617-2 [Bunt et al. 2013, Yoshino et al. 2018]
- ➢ Define the dialogue state for the reference interview

☐ Apply the reinforcement learning

- ➢ Understanding unknown reward structure in a reference interview
- E.g. Inverse reinforcement learning (IRL)

End Slide.

Example of BPE Tokenization

e.g.

we need simple explanations for the nervous and lymphatic system .

- **Vocabulary size= 100**

_we _n e ed _s i m p le _ e x p l an tion s _for _the _n er v o u s _and _l y m p h a t i c _s y st e m .

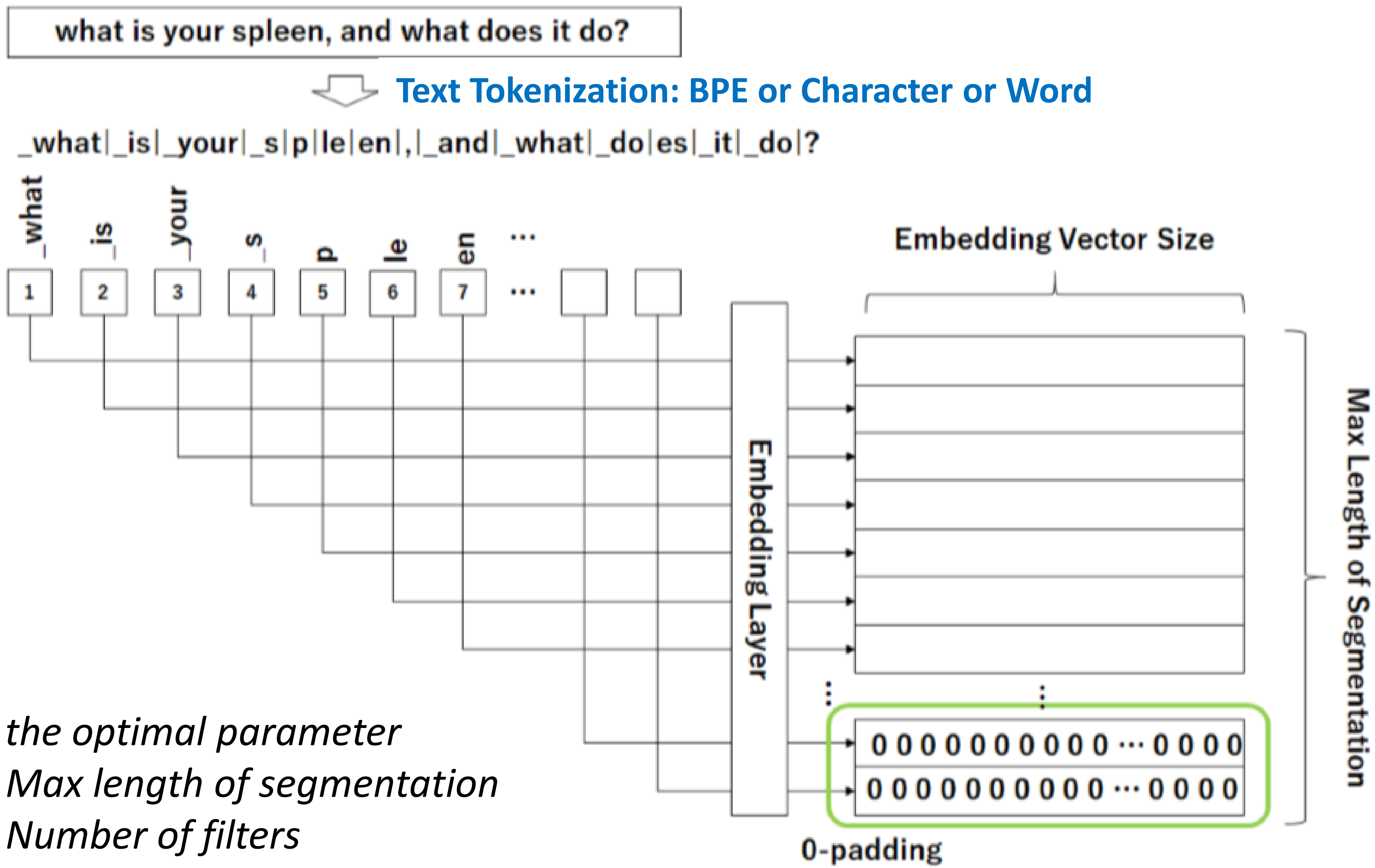
- **Vocabulary size = 500**

_we _need _s i m p le _ex pl an tion s _for _the _n er v o us _and _ly m p h at ic _system .

- **Vocabulary size = 1000**

_we _need _simple _ex pl an tion s _for _the _ne r v o us _and _ly m pha tic _system .

Input Generation to CNN



Set the optimal parameter

1. *Max length of segmentation*
2. *Number of filters*
3. *Kernel size of convolution*
4. *Stride length of convolution*

Figure. The Input Generation to CNN