Contextual Embeddings for Hypernym Discovery

Geonil Yun

rjsdlf45@gmail.com

19th Oct, 2022

Contents

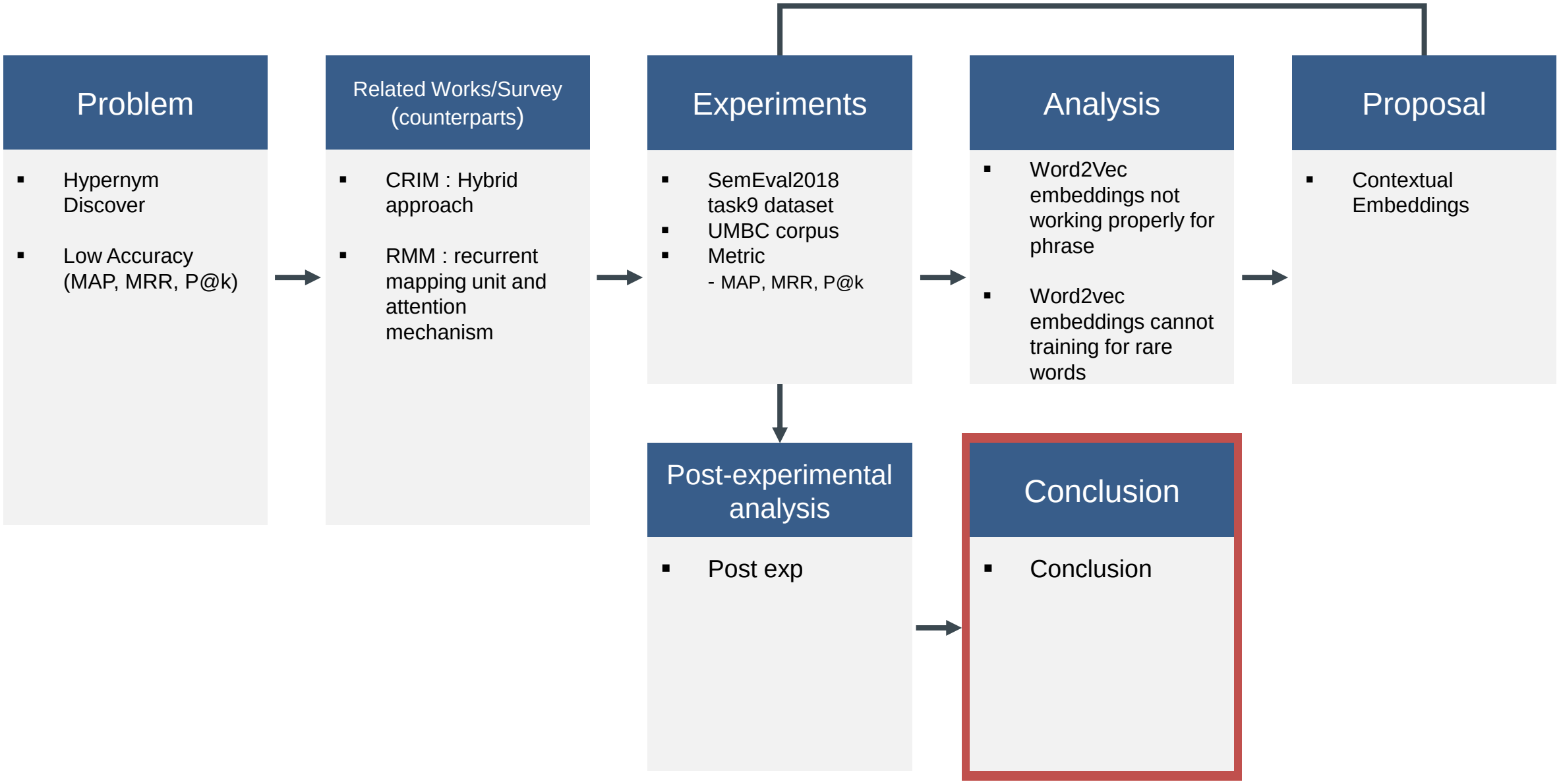
To Do List from Previous Seminar

- ✓ Experiments – repeat 30 times
  - ➢ 1A, 2A, 2B for Proposed model
  - ➢ it will take some time

- ✓ Explain what vectors are good for hypernym discovery
  - ➢ Difference between contexture embedding and word2vec embedding

- ✓ Pretraining은 우선 Appendix로 이동, Finetuning만으로 논문 마무리

Overall Progress



Hypernym Discovery

Datasets

- Corpus, training, validation, test, vocabs => Learning word embeddings(for query and vocabs) in corpus
- Train, validation, test queries(**bold**) are mutual exclusive.
- Gold hypernyms(not bord) are in the given vocabulary

Corpus

Johnson says paying bonuses to employees is not an unusual practice , especially as the state tries to compete with the private sector to keep good employees . He says recent media coverage shows Howard was " unfairly scrutinized " by politicians and pundits who accused her of " emptying the drawers " before her successor came into office . " The fact that she ' s simply done what the Legislature told her to do probably puts an unfair burden on her ," Johnson said , citing a senate bill passed earlier this year by lawmakers allowing department heads to reward leftover salary money to employees at their discretion . Overall , State Controller Keith Johnson says he is concerned the recent scrutiny over bonuses will make agency heads more cautious about rewarding deserving employees when there is leftover money in their salary drawers . …

Train

Carly Fiorina : person
Kim Il-sung University : academy university educational institution
phonetic transcription : written language written communication
⋮

Valid

Burger King : eating house eating place restaurant
cingulate gyrus : neural structure anatomical structure
Pinot blanc : grapevine grape vine white wine
⋮

Test

oil pollution : misfortune disaster danger catastrophe slick
XML : data file computer file software program
intermittent claudication : claudication disorder sickness
⋮

Vocabulary

a thousand million
a$
a&e
a.k.a.
a.m.
a.r.
a/2
…

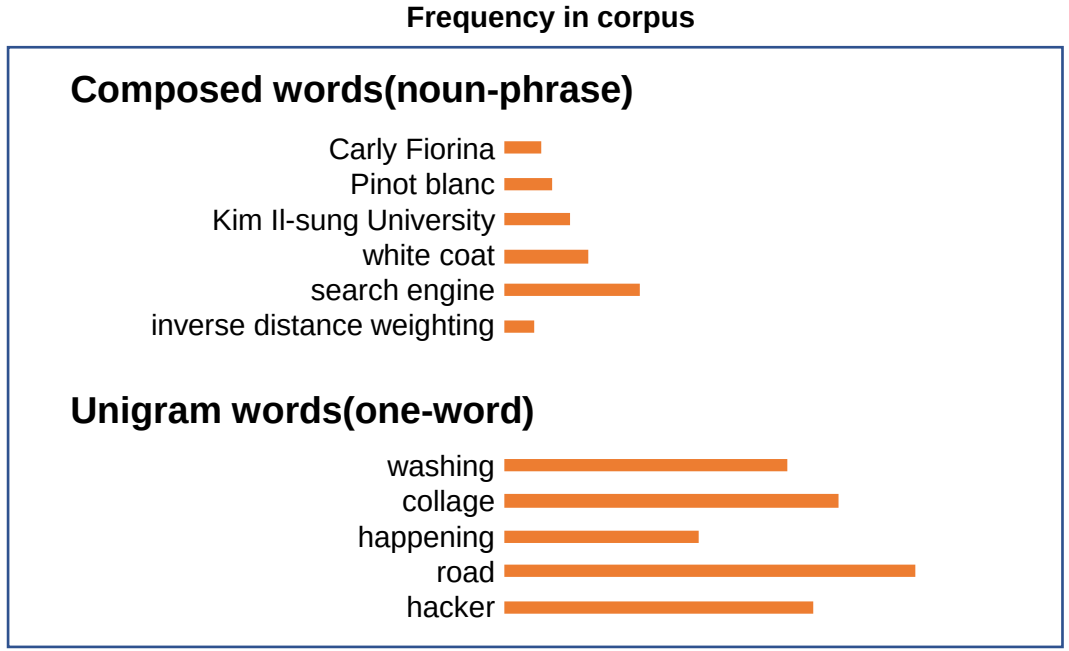
war
war baby
war bond
war bride
war chest
war cloud

…
¢
¥
©
°
∂

Related Works

CRIM, RMM

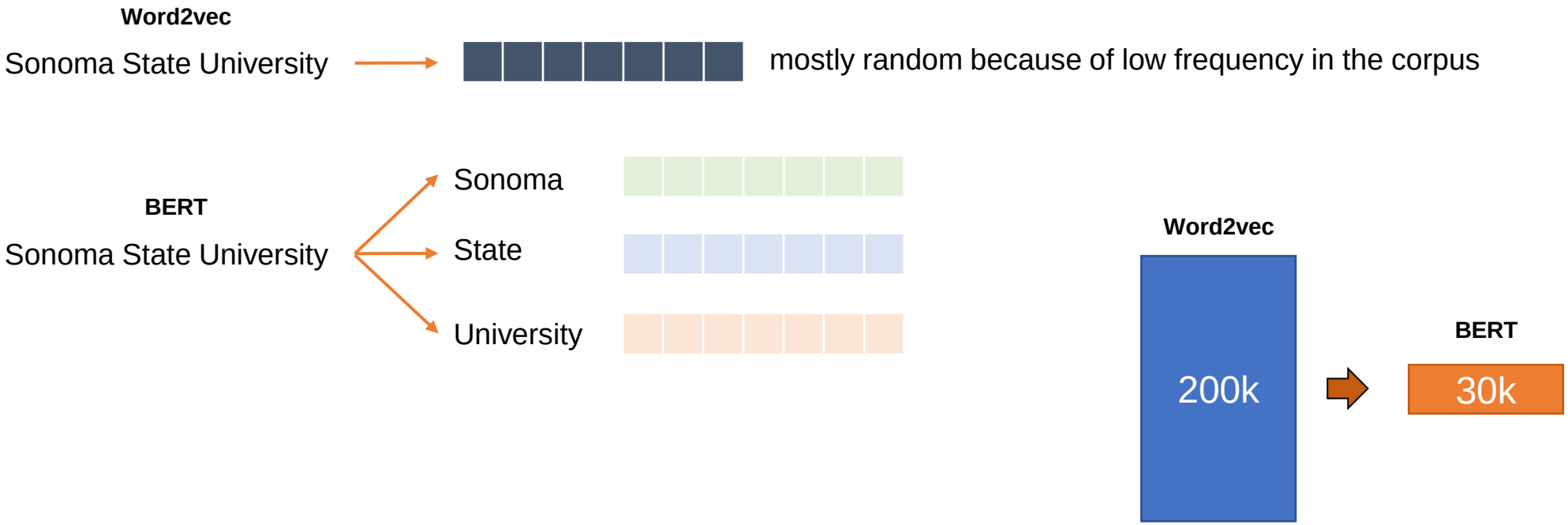
- Training train, validation, test queries and vocabulary words with word2vec embeddings(skip-gram)
- Problem of word2vec is cannot learn properly for composed words(noun-phrase)
- Word2vec needs to learn all words in vocabs(about 200k) for word representation



Proposed Method

Contextual Embeddings

- Using contextual embeddings for phrase representation
- BERT can tokenize noun-phrase to sub-word => Not need to learn all vocabs



Proposed Method

Why pre-training?

- Original pre-trained BERT was trained for the language modeling
- Need to learn hypernym-relation knowledge for BERT
- Using masked hearst pattern modeling for adaption hypernym relation

Corpus

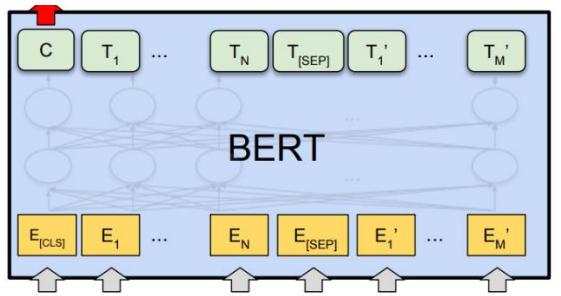
Johnson says paying bonuses to employees is not an unusual practice , especially as the state tries to compete with the private sector to keep good employees . He says recent media coverage shows Howard was " unfairly scrutinized " by politicians and pundits who accused her of " emptying the drawers " before her successor came into office . " The fact that she ' s simply done what the Legislature told her to do probably puts an unfair burden on her ," Johnson said , citing a senate bill passed earlier this year by lawmakers allowing department heads to reward leftover salary money to employees at their discretion . Overall , State Controller Keith Johnson says he is concerned the recent scrutiny over bonuses will make agency heads more cautious about rewarding deserving employees when there is leftover money in their salary drawers . …

Find hearst patterns

	Pattern
Hearst	Y such as (and|or) X
	such Y as (and|or) X
	X (and|or) other Y
	Y (including|especially) (and|or) X

Dog is an **animal**
city such as **Incheon** and **Seoul**

Dog is an **[MASK]**
city such as **[MASK]** and **Seoul**



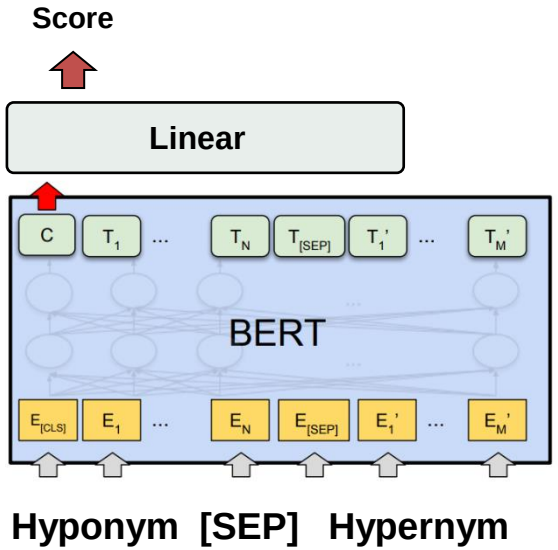
Dog is an **animal**
city such as **Incheon** and **Seoul**

Dog is an **[MASK]**
city such as **[MASK]** and **Seoul**

Proposed Method

Fine-tuning

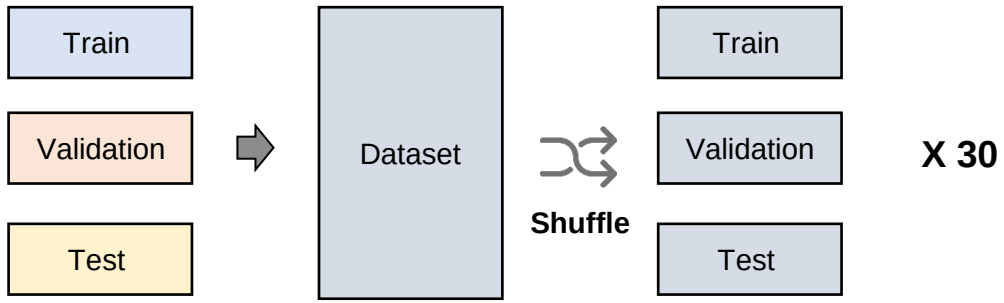
- Fine-tune the Hearst-BERT(before pretrained by hearst pattern denoising) for hypernym discovery task
- Using negative sampling for the training



Experimental Settings

Dataset

- The experiment was repeated 30 times with the shuffled dataset .
- We used 3 sub-tasks of SemEval2018 task9-Hypernym Discovery : 1A(general), 2A(Medical), 2B(Music)
- We employed 3 metrics for evaluation : MRR, MAP, P@k



Experimental Results

Dataset	Metric	Proposed	$CRIM_{r1}$	RMM
1A General	MRR	**32.19 ± 4.79**	**37.49 ± 0.92**	
	MAP	**20.38 ± 2.15**	**25.63 ± 0.65**	
	P@1	**22.65 ± 5.43**	**28.21 ± 1.00**	
	P@3	**17.79 ± 2.61**	**22.87 ± 0.59**	
	P@5	**17.87 ± 2.13**	**22.86 ± 0.60**	
	P@15	**23.86 ± 1.67**	**29.24 ± 0.83**	
2A Medical	MRR	**51.48 ± 8.54**		
	MAP	**38.54 ± 5.46**		
	P@1	**38.01 ± 9.30**		
	P@3	**34.79 ± 6.64**		
	P@5	**35.30 ± 5.98**		
	P@15	**44.28 ± 4.27**		
2B Music	MRR	**52.35 ± 5.95**		
	MAP	**38.80 ± 3.58**		
	P@1	**38.43 ± 7.66**		
	P@3	**35.87 ± 4.47**		
	P@5	**36.52 ± 4.05**		
	P@15	**43.27 ± 3.44**		

To Do List

- ✓ Pretraining LM(language model) with masked hearst pattern modeling
- ✓ Check the frequency of unigram, bigram, trigram words in the corpus
- ✓ Check the appearance of test queries in the corpus

Thank you

Restoration of Homoglyph Attack on Spam Using BERT Masked Language Model

Hanyong Lee

glhy0718@gmail.com

19th Oct, 2022

Contents

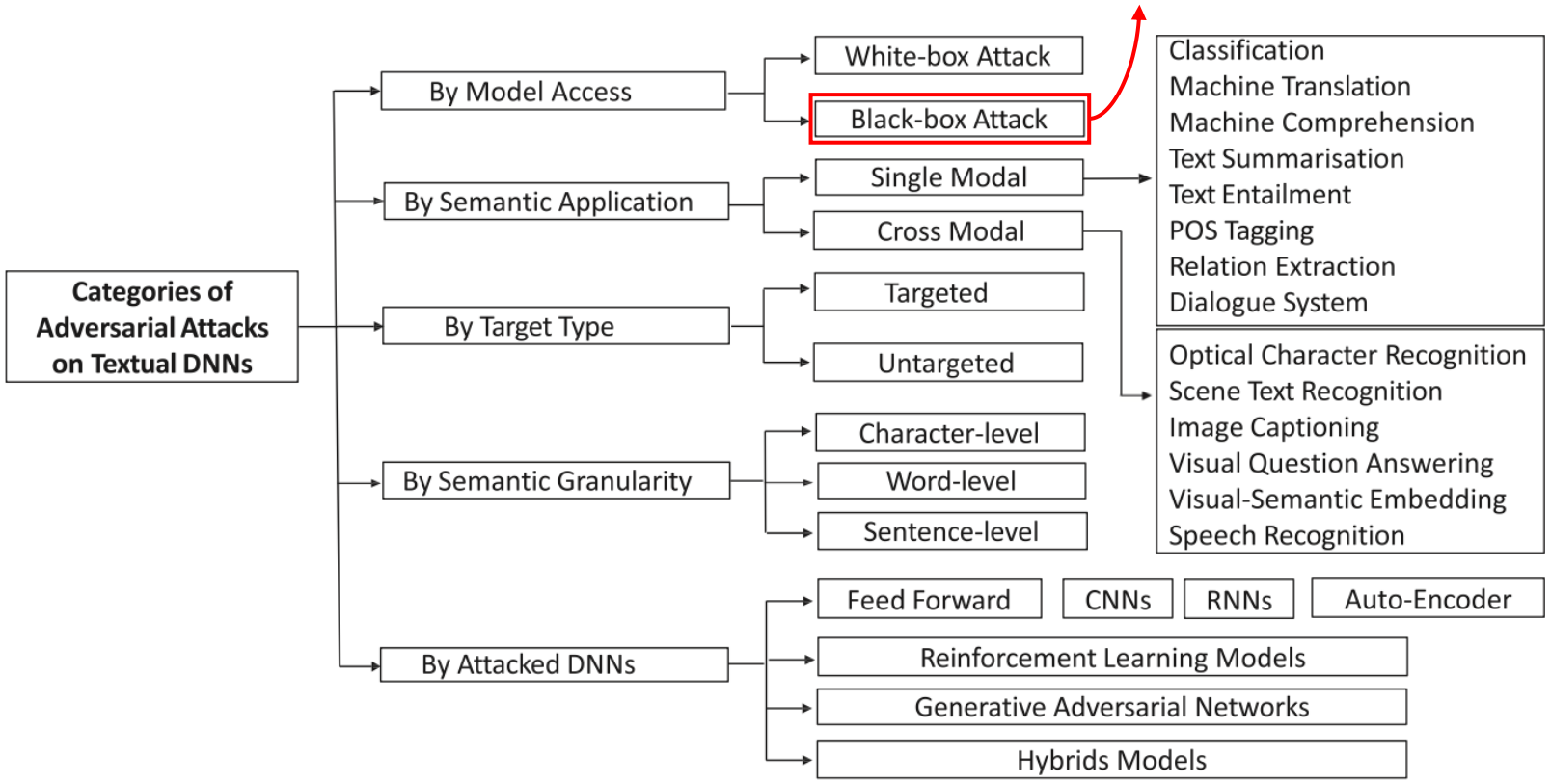
To-Do List from Previous Seminar

- 논문작성 계속 진행
- PPT 내용 수정
  - ✓ Friedman test의 critical value 넣기
  - ✓ Contribution에 in depth analysis 추가
- 특허 진행 상황
  - ✓ 출원지시 완료
- ICEIC논문 작성
  - ✓ 초안 작성완료

Introduction

NLP Attack

Adversarial examples in **Natural Language Processing**

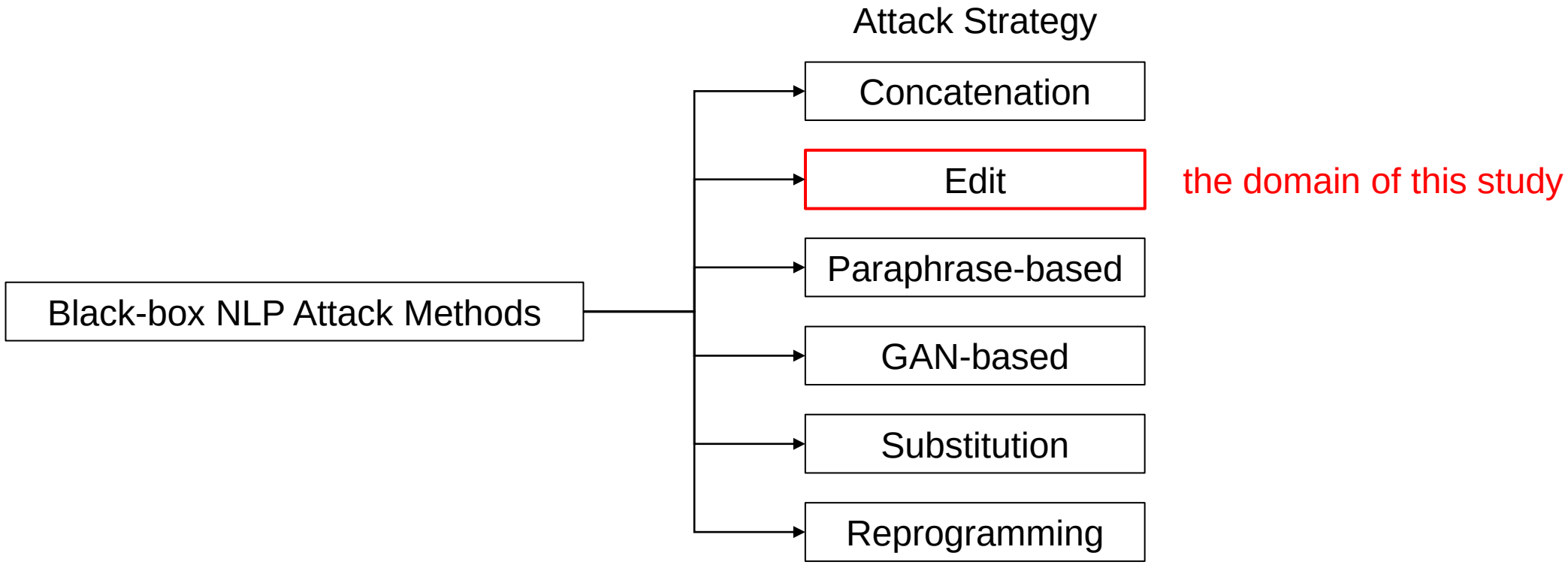


Zhang, W. E., Sheng, Q. Z., Alhazmi, A., & Li, C. (2020). Adversarial attacks on deep-learning models in natural language processing: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 11(3), 1-41.

Introduction

NLP Attack

Black-box NLP Attack Methods

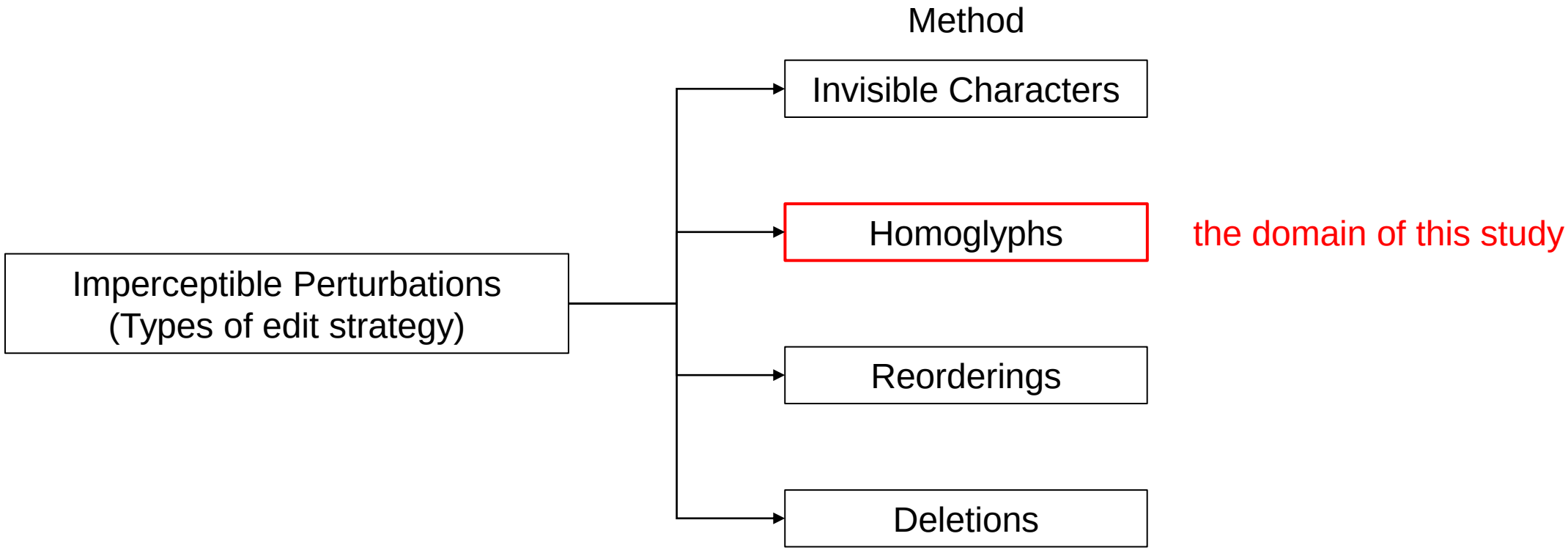


Zhang, W. E., Sheng, Q. Z., Alhazmi, A., & Li, C. (2020). Adversarial attacks on deep-learning models in natural language processing: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 11(3), 1-41.

Introduction

Homoglyph Attack

Imperceptible Perturbations

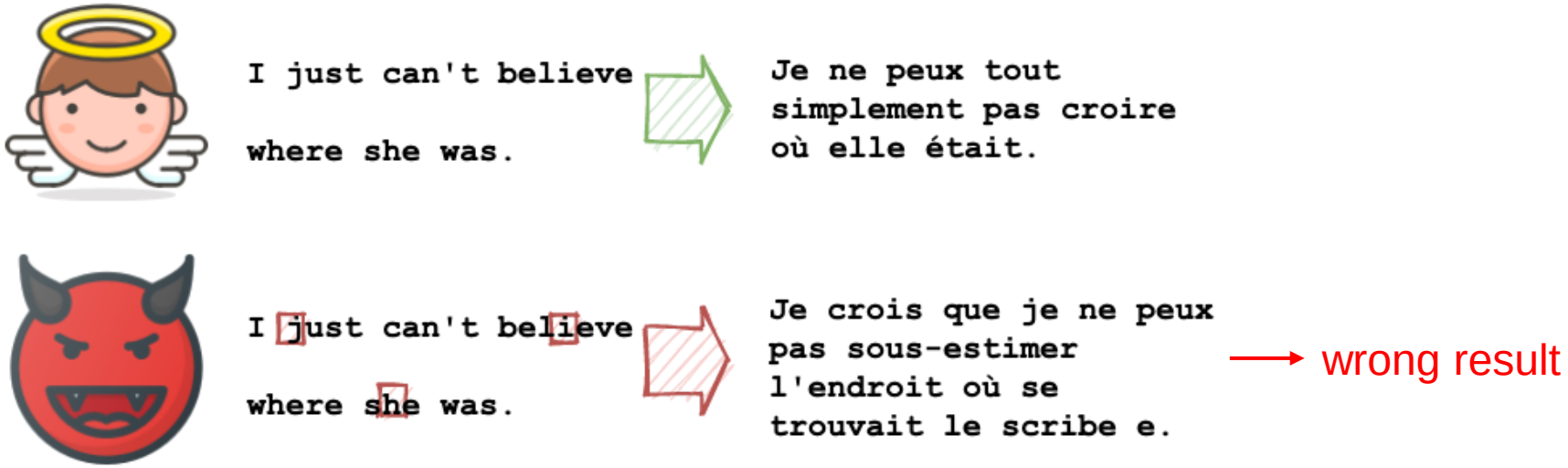


Boucher, N., Shumailov, I., Anderson, R., & Papernot, N. (2022, May). Bad characters: Imperceptible nlp attacks. In 2022 *IEEE Symposium on Security and Privacy (SP)* (pp. 1987-2004). IEEE.

Related Work

Homoglyph Attack

Example of an attack using homoglyph to confuse English-Franch translation task.

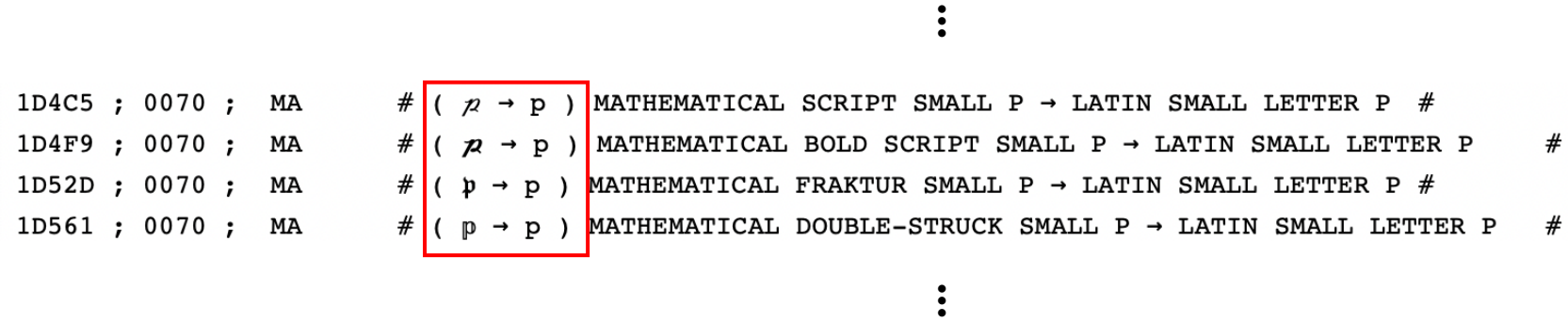


“j” is replaced with U+3F3(“j”), “i” with U+456(“i”), and “h” with U+4BB(“h”).

Boucher, N., Shumailov, I., Anderson, R., & Papernot, N. (2022, May). Bad characters: Imperceptible nlp attacks. In 2022 *IEEE Symposium on Security and Privacy (SP)* (pp. 1987-2004). IEEE.

Defense Methods of Homoglyph Attack

Translate table for confusable characters



```
1D4C5 ; 0070 ;  MA      # ( 𝓅 → p ) MATHEMATICAL SCRIPT SMALL P → LATIN SMALL LETTER P  #
1D4F9 ; 0070 ;  MA      # ( 𝓹 → p ) MATHEMATICAL BOLD SCRIPT SMALL P → LATIN SMALL LETTER P    #
1D52D ; 0070 ;  MA      # ( 𝔭 → p ) MATHEMATICAL FRAKTUR SMALL P → LATIN SMALL LETTER P #
1D561 ; 0070 ;  MA      # ( 𝕡 → p ) MATHEMATICAL DOUBLE-STRUCK SMALL P → LATIN SMALL LETTER P   #
```

It doesn’t contain some pairs.

- μ → u
- $ → S
- 5 → S
- …

Unicode Consortium 2022, *Unicode Utilities: Confusables.* accessed 10 September 2022, <https://www.unicode.org/Public/security/15.0.0/confusables.txt>.

Related Work

Defense Methods of Homoglyph Attack

SimChar database

H. Suzuki et al. used the following formula to calculate the similarity of two bitmap images of characters.

$$\Delta = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |I_1(i,j) - I_2(i,j)|$$

pixel of image I_1, I_2 at coordinate (i, j)

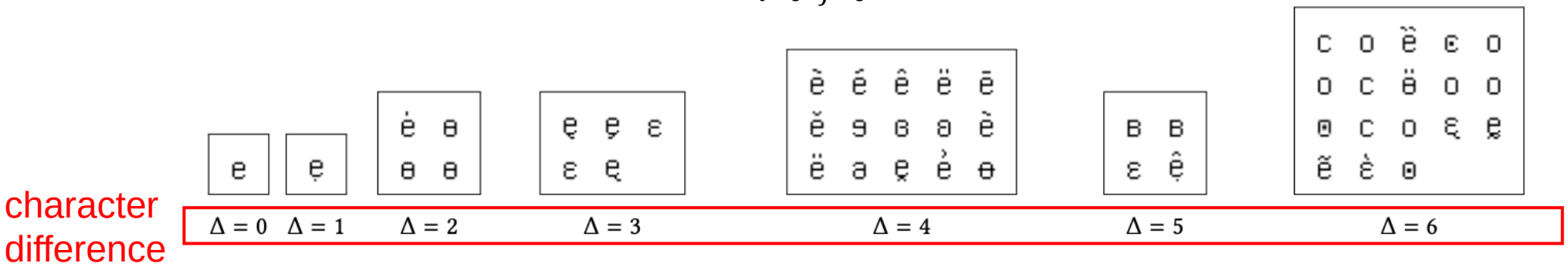


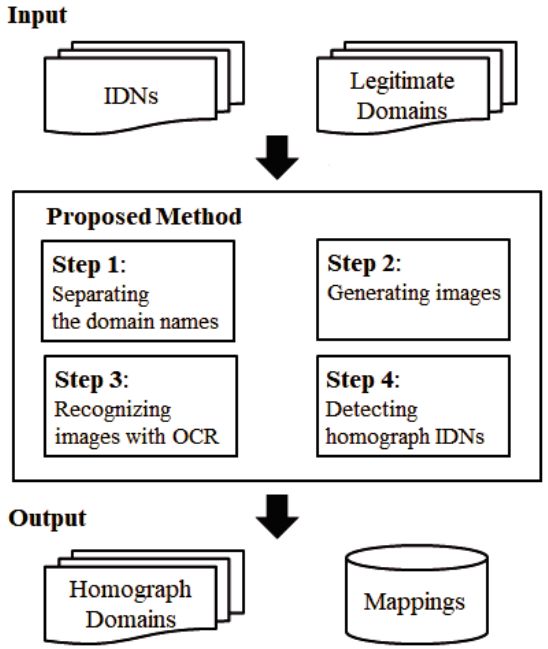
Figure 6: Basic Latin lowercase letter 'e' and characters under different values of the threshold Δ. In this work, characters with $\Delta \leq 4$ are detected as homoglyphs.

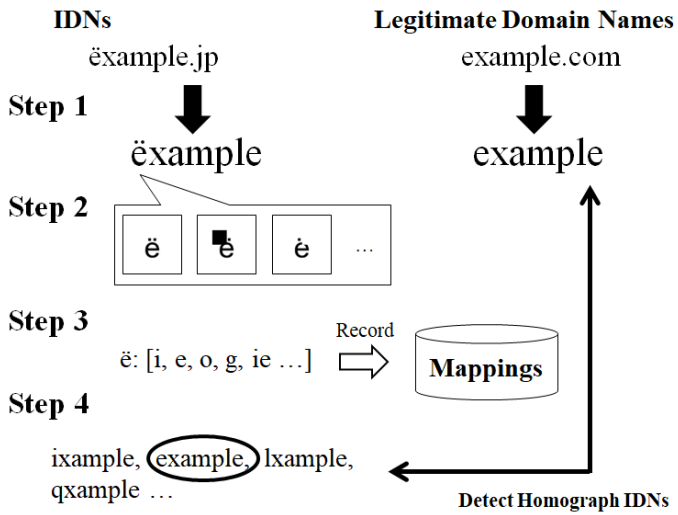
Suzuki, H., Chiba, D., Yoneya, Y., Mori, T., & Goto, S. (2019, October). ShamFinder: An automated framework for detecting IDN homographs. *In Proceedings of the Internet Measurement Conference* (pp. 449-462).

Defense Methods of Homoglyph Attack

OCR-based method

Sawabe et al. used the OCR-based Method to detect Homoglyph Attack.





Sawabe, Y., Chiba, D., Akiyama, M., & Goto, S. (2019). Detection method of homograph internationalized domain names with OCR. *Journal of Information Processing*, *27*, 536-544.

Defense Methods of Homoglyph Attack

Spellchecker-based method

Imam, N. H et al. used a spellchecker-based method to restore the homoglyph attacks.

Table 2
Some of the adversarial text examples used in experiments.

Original	Flipping	Swapping	Deletion
Busy	Bu$y	Bsuy	Bsy
Caller	C@1ler	Claler	Cller
Reply	Rep1y	Rpely	Rply
winner	w!nner	wniner	wnner

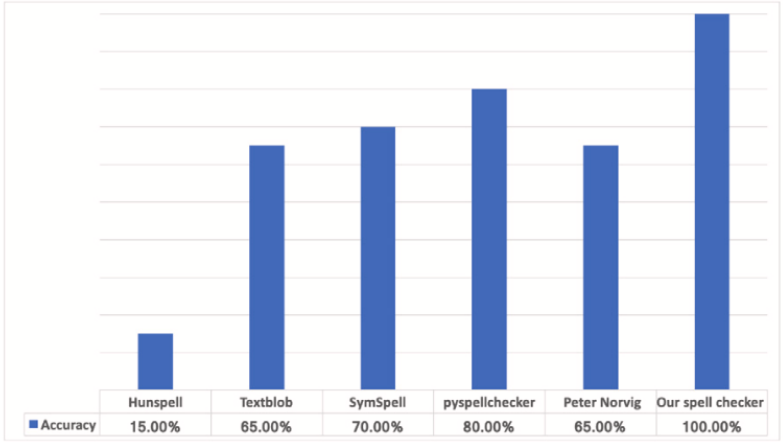


Fig. 9. Flipping.

Imam, N. H., Vassilakis, V. G., & Kolovos, D. (2022). OCR post-correction for detecting adversarial text images. *Journal of Information Security and Applications*, *66*, 103170.

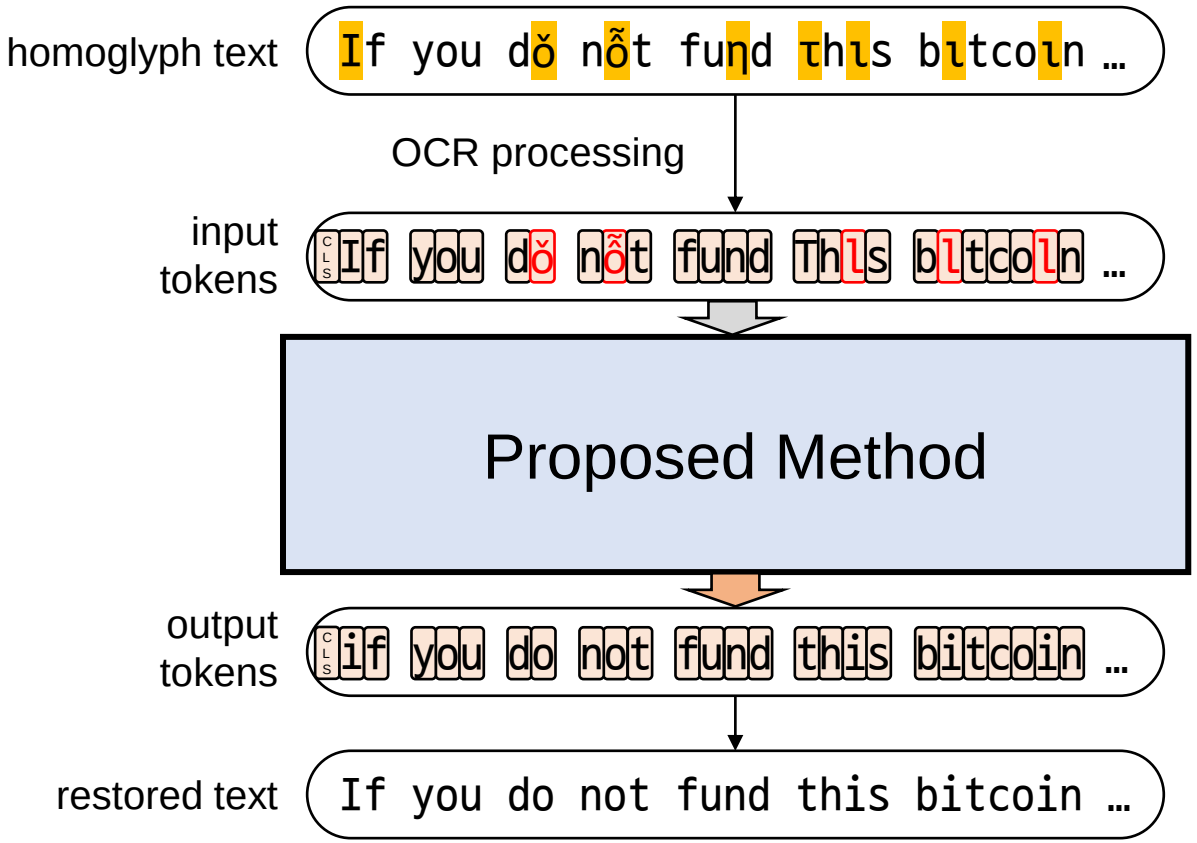
Related Work

Pros & Cons of related works

Algorithm	PROS	CONS
BERT	• Slightly robust on new homoglyphs	• Pre-raining the model takes a long time. • Word-based tokenization is not helpful for the task.
OCR	• Robust on new homoglyphs • Automated workflow	• OCR is not performing well on homoglyphs • Inability when the input & output lengths are different
SimChar DB	• Automated workflow on finding homoglyph pairs	• Not robust on new homoglyphs • Must manually add homoglyph pairs to the dictionary • Inability when the input & output lengths are different
Spell Checker	• Robust on new homoglyphs	• Must manually add words to the dictionary

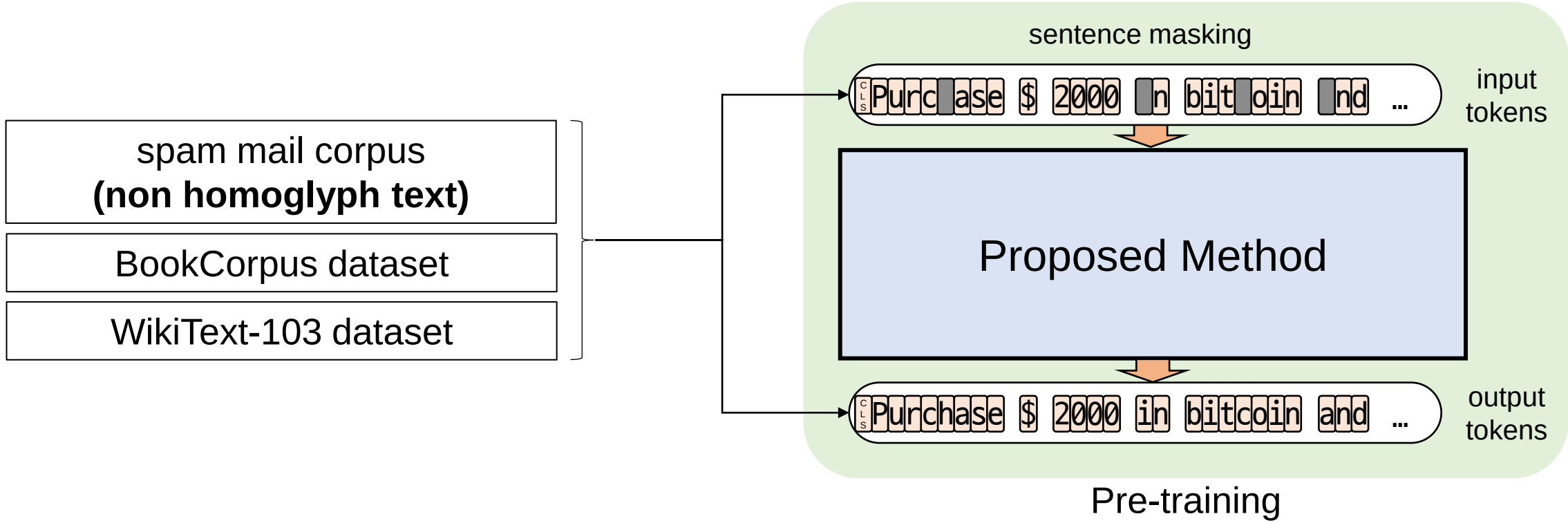
Proposed Method

Restoring Process Overview



Proposed Method

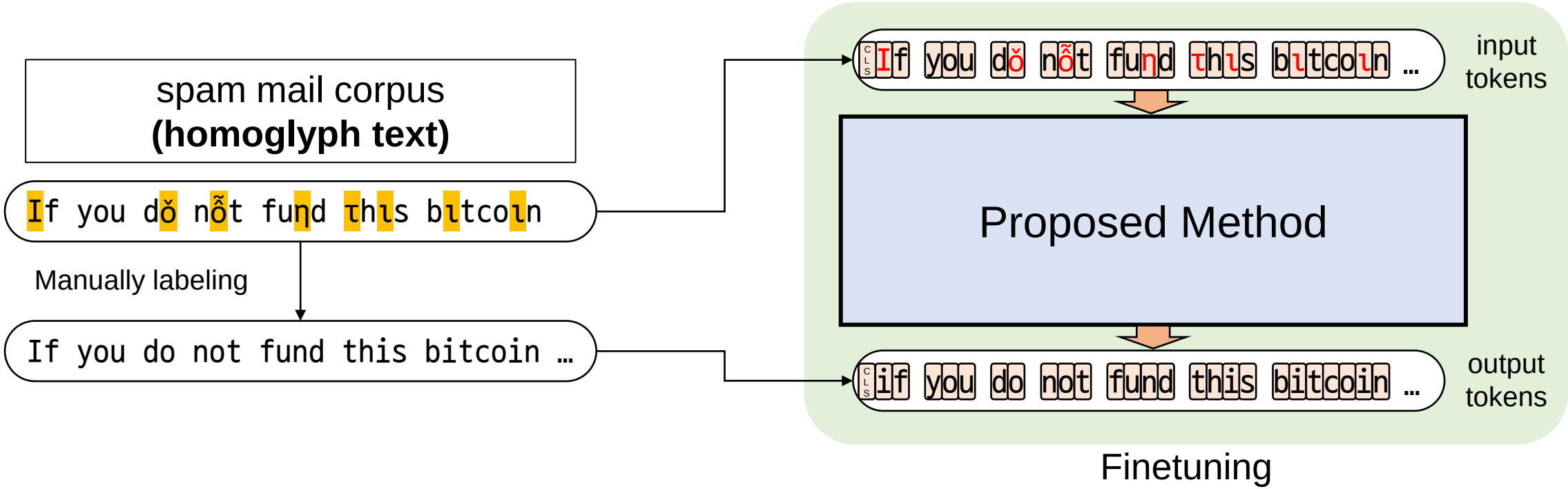
Pre-training



We pretrained the proposed model as a masked language model (MLM).

Proposed Method

Finetuning



Evaluated with a blind test (Train/Test set split ratio = 8:2 / Iteration = 50 times)

Experimental Settings

Dataset

		# of Examples		# of Tokens	
	Dataset	Train Set	Validation Set	Train Set	Validation Set
Pre-training dataset	WikiText-103	1,165,029	2,461	285,324,545	612,967
	BookCorpus (about 2% of the total dataset)	1,480,085	3,126	79,346,399	143,939
	bitcoinabuse.com dataset (non-homoglyph)	299,239	633	23,152,956	54,680
Finetuning dataset	bitcoinabuse.com dataset (homoglyph)	21,342	5,336	random split	random split

Experimental Settings

Model Parameters

Parameter	Value
Input Max Sequence	512
Vocab Size	881
# of Attention Heads	12
# of Hidden Layers	12
# of Model Parameters	86,720,369

Experimental Settings

Pre-training Setting

Parameter	Value
# of Epochs	25
Train / Validation split	refer the slide 14.
Used Datasets	• WikiText-103 • BookCorpus (about 2% of the total dataset) • bitcoinabuse.com dataset (non-homoglyph sentences)
Metric	Accuracy

Experimental Settings

Finetuning Setting

Parameter	Value
# of Epochs	5
Train / Validation split	8 : 2
# of Test Iteration	50
Used Datasets	• bitcoinabuse.com dataset (homoglyph sentences)
Metric	Accuracy

Experimental Results

Finetuned Model Evaluation - Accuracy

Method	Word Accuracy
Proposed	**0.9959 ± 0.0008**
BERT	0.6713 ± 0.0029 ▼
OCR	0.6518 ± 0.0028 ▼
SimChar DB	0.5512 ± 0.0046 ▼
Spell Checker	0.9087 ± 0.0011 ▼

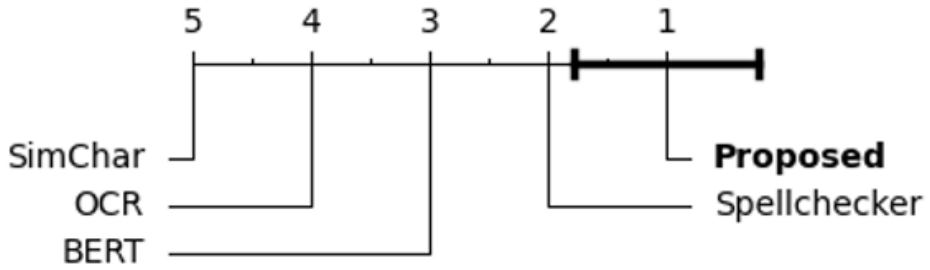
Word level accuracy and T-test with the accuracy of the proposed method
▼ means that the algorithm has a significantly lower performance than the proposed algorithm based on the t-test.
▼: $p < 0.01$

Experimental Results

Finetuned Model Evaluation – Friedman test

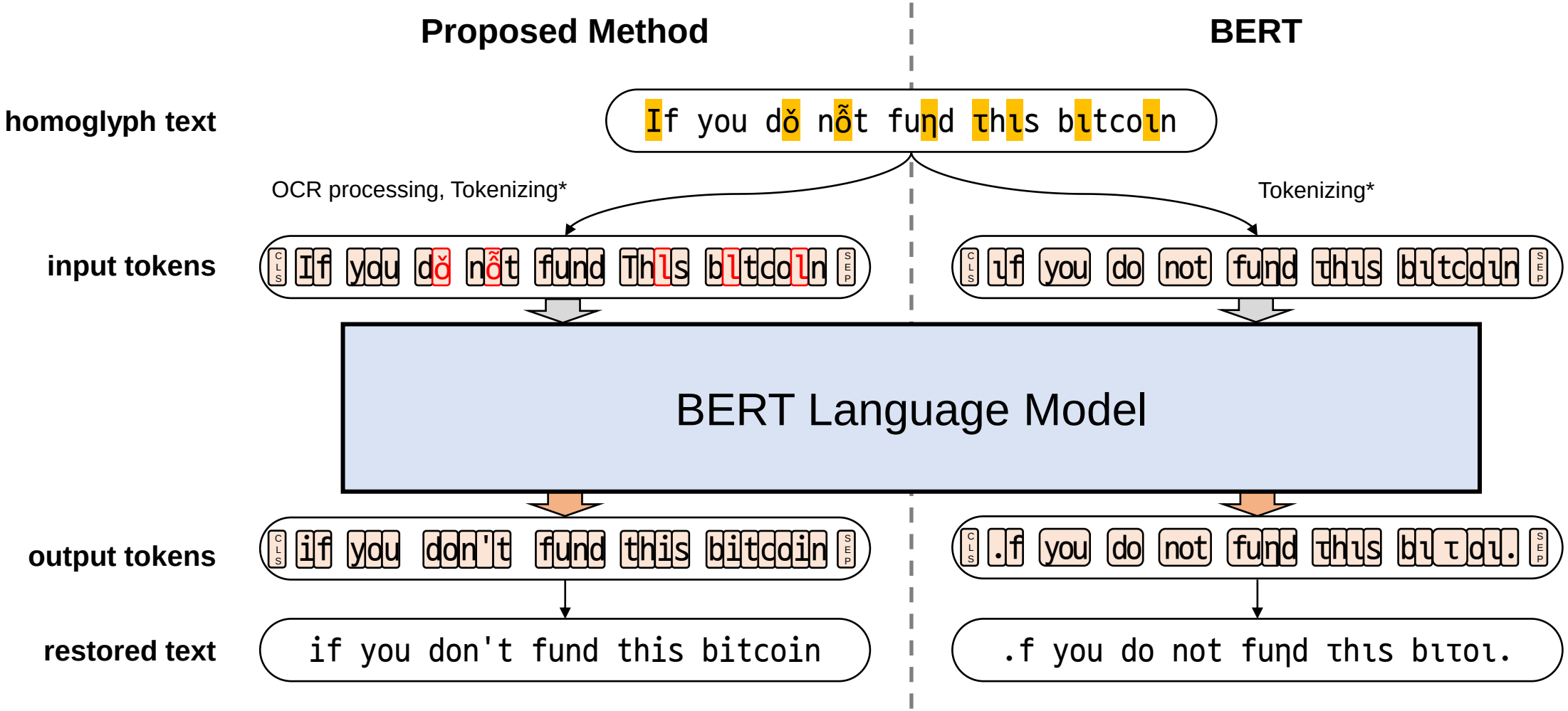
Statistic	p value	critical value
100.0	2.3643e-05	113.145

Finetuned Model Evaluation – Bonferroni–Dunn test (α = 0.05, CD = 0.7899)



Experimental Results

Analysis – A comparison of the proposed method and BERT



* Tokenizing has a Unicode normalization process.

Experimental Results

Analysis – qualitative analysis

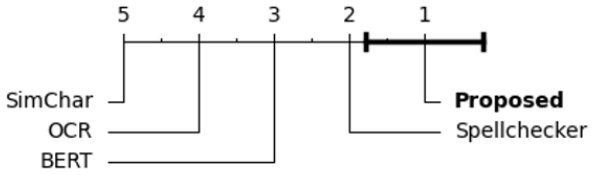
Original Text	Ground Truth	Proposed Method	BERT model
ī am well aware is yōūr pãss.	i am well aware is your pass.	i am well aware is your pass.	l am well as is your pass.
you have 48 hours to make the p ayment.	you have 48 hours to make the p ayment.	you have 48 hours to make the p ayment.	you have 48 hours to make the p ayment.
l have α unıque pixel wiтhiη тhıs e mαıl message	i have a unique pixel within this e mail message	i have a unique pixel within this e mail message	i have a unlqueid with the thls em all message
all íпgeиious ỉs sìмple.	all ingenious is simple.	all ingenious is simple.	all ingenlous ls simple.
you'll mακe тhe ραymeηt via βitco ıη to тhe below αddress	you'll make the payment via bitcoi n to the below address	you'll make the payment via bitcoi n to the below address	you. ll make the paymeat via bitc an to the above address

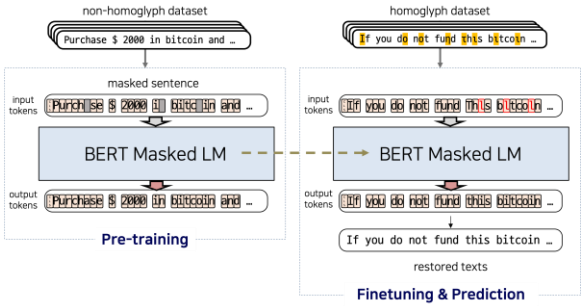
- homoglyph
- wrong word

- Since the BERT model Unicode normalization is being processed during the tokenization, it is partially effective for homoglyph restoration, but not fo r predefined homoglyphs.
- BERT's sub-word tokenization is basically not suitable for homoglyph restoration because one token contains one or more characters.

Conclusion

Contribution of the Study





1. We achieve higher performance than traditional methods such as OCR, SimChar DataBase, and spell checker.

2. We propose the Character-level BERT algorithm and introduce a dataset for the homoglyph restoration task.

3. Limitations of the existing methods were found, and in-depth analysis was conducted through statistical verification through the t-test, Freidman test, and Bonferroni–Dunn test to determine that the proposed model performed better than the existing model.

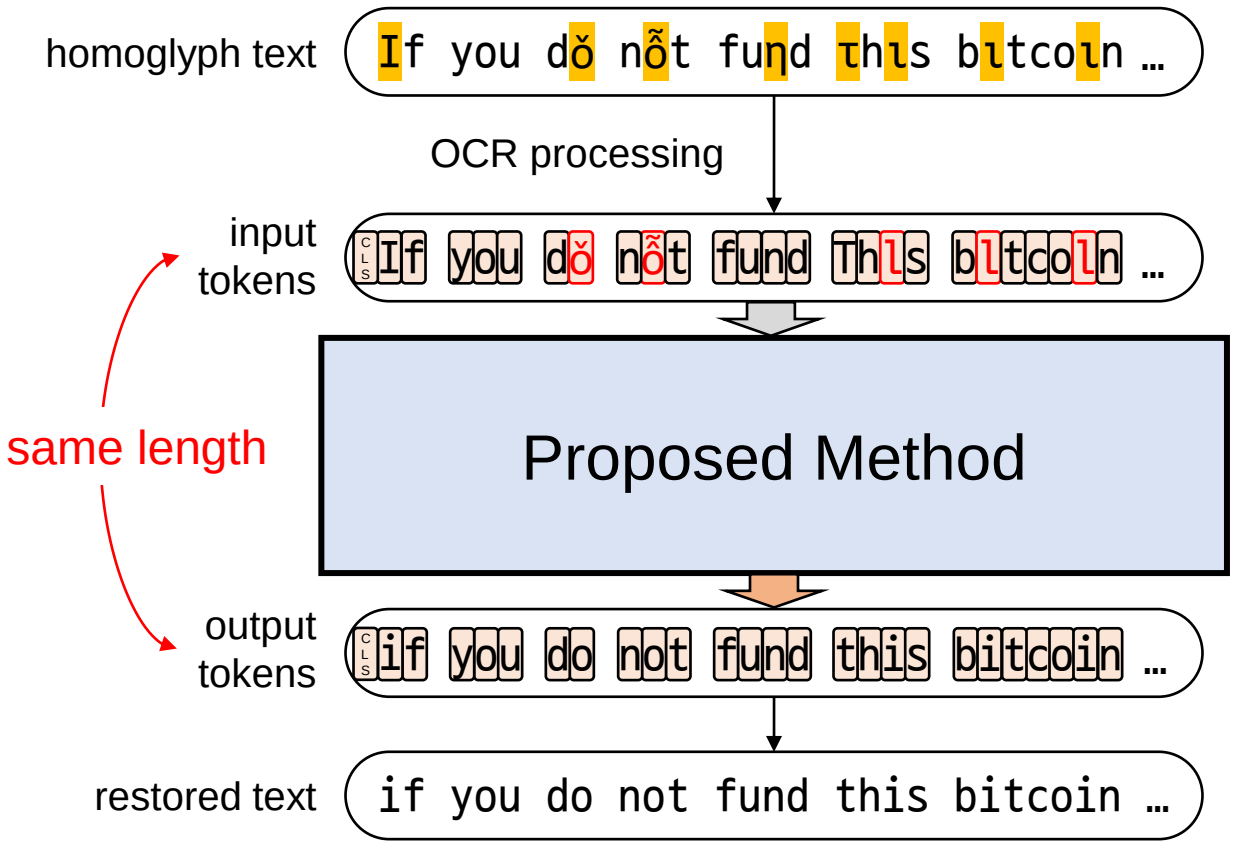
Conclusion

Limitation of Proposed Model

The proposed model is an encoder-only model, limited in that the input/output sentences should be the **same sequence length**.

- It can be used as a homoglyph of the alpha bet ‘m’ by concatenating the letters ‘r’ and ‘ n’. But the presented model does not solve this case.

The proposed model can be applied as a **sequence-to-sequence** model in future studie s.



References

1. Zhang, W. E., Sheng, Q. Z., Alhazmi, A., & Li, C. (2020). Adversarial attacks on deep-learning models in natural language processing: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 11(3), 1-41.
2. Boucher, N., Shumailov, I., Anderson, R., & Papernot, N. (2022, May). Bad characters: Imperceptible nlp attacks. In 2022 IEEE Symposium on Security and Privacy (SP) (pp. 1987-2004). IEEE.
3. Unicode Consortium, "Unicode Utilities: Confusables." unicode.org. https://www.unicode.org/Public/security/15.0.0/confusables.txt (accessed 10 September 2022).
4. Suzuki, H., Chiba, D., Yoneya, Y., Mori, T., & Goto, S. (2019, October). ShamFinder: An automated framework for detecting IDN homographs. In Proceedings of the Internet Measurement Conference (pp. 449-462).
5. Sawabe, Y., Chiba, D., Akiyama, M., & Goto, S. (2019). Detection method of homograph internationalized domain names with OCR. Journal of Information Processing, 27, 536-544.
6. Imam, N. H., Vassilakis, V. G., & Kolovos, D. (2022). OCR post-correction for detecting adversarial text images. Journal of Information Security and Applications, 66, 103170.
7. Norvig, P., "How to write a spelling corrector." norvig.com https://norvig.com/spell-correct.html (accessed 10 September 2022).
8. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
9. Merity, S., Xiong, C., Bradbury, J., & Socher, R. (2016). Pointer sentinel mixture models. arXiv preprint arXiv:1609.07843.
10. Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international conference on computer vision (pp. 19-27).

특허 출원 진행 상황

- P20220246, P20220244, P20220242: 출원 완료
- P20220254
  - 출원지시 완료

번호	관리번호	특허구분	명칭	발명자	출원번호	국가	특허사무소	처리상태
5	CAU20224499KR	국내특허	미미틱 알고리즘을 사용한 다중 라벨 특징 선택 방법...	이재성		대한민국	에스와이피(SYP) 특...	출원지시
4	CAU20224502KR	국내특허	실시간 사용자 활동 인식을 위한 다중시간 샘플링 시...	이재성	10-2022-0138362	대한민국	스카이특허법률사무...	출원완료
3	CAU20224503KR	국내특허	자동 음악 태그 분류 성능 향상을 위한 음악 특징 선...	이재성	10-2022-0138361	대한민국	스카이특허법률사무...	출원완료
2	CAU20224504KR	국내특허	미분 기반 신경망 구조 탐색 시스템 및 방법	이재성	10-2022-0138363	대한민국	스카이특허법률사무...	출원완료

ICEIC 논문

- 초안 작성 완료

To-Do List

- 안성 워크스테이션 IP 공개 관련 교수님께 문의
- ICEIC 논문 초안 검토
- 논문 가이드라인 채우기

Thank you

Appendix

A. Proposed Model Information

Character-level BERT

- Number of parameters: 86,720,369
- Max length of input tokens: 512
- Vocab size: 881

Appendix

B. Pre-trained Model Evaluation

Evaluated with the **whole test sentences**

Character level BERT is only pretrained with the **WikiText-103, BookCorpus & spam mail datasets**.

Algorithm	Word Accuracy
Proposed (not finetuned)	**0.9516 ± 0.1289**
Normal BERT	0.6713 ± 0.2679 ▼
OCR	0.6519 ± 0.2976 ▼
SimChar DB	0.5506 ± 0.3627 ▼
Spell Checker	0.9090 ± 0.1620 ▼

Word level accuracy and T-test with the accuracy of the proposed method
▼ means that the algorithm has a significantly lower performance than the proposed algorithm based on the t-test.
▼: $p < 0.01$

Appendix

C. Test Output Examples of Pretrained Model

Example 1

- input sentence L Neeԁ y0UR TOTɑı ɑTTeNTıOɴ fOR The c0MıNg TweNTY-fOUR hRS,
- output sentence : i need your total attention for the coming twenty - four hrs,

Example 2

- input sentence : ɑftêr thät, i hȧvë ṡtȧrtèd trȧcкiṅġ ẏȯur ịṅtérпët ȧctívítìêṡ.
- output sentence : after that, i have started tracking your internet activities.

Appendix

C. Test Output Examples of Pretrained Model

Example 3

- input sentence : do ṅȯt trɣ tȯ coṅtáct the ṗolḭćê àɴd ȯtɳêr sëcurity servḭces.
- output sentence : don't try to contact the police and other security services.

Example 4

- input sentence : yøuř åccøuňt wås åttåckěd.
- output sentence : yogr account was attacked.

Data Imputation for missing value

Hyejin Baek

bhj4208@gmail.com

23th Nov, 2022

Contents

1. To Do List from Previous Seminar
2. TabNet
3. Experiment
4. To Do List

To Do List from Previous Seminar

- ✓ TabNet 분석 후 실험 진행
  → 진행 완료 / 추가 분석 필요

- ✓ Experiment 코드 보완 및 연구 진행
  → 진행 완료

- ✓ Nearest Neighbors/Machine Learning 연구 진행
  → 진행중

TabNet

1. Breast_cancer (roc-auc-score)

	TabNet
20%	0.891
40%	0.912
60%	0.989
80%	0.910

Experiment

1. Breast_cancer (binary classification - accuracy)

	XGBoost		AdaBoost	
	zero	KNN	zero	KNN
20%	0.889±0.020	0.883±0.025	0.883±0.022	0.882±0.018
40%	0.806±0.027	0.803±0.032	0.803±0.031	0.796±0.029
60%	0.739±0.0244	0.737±0.023	0.739±0.019	0.716±0.011
80%	0.700±0.018	0.699±0.020	0.700±0.017	0.703±0.013

2. Abalone (regressor – Mean Squared Error)

	XGBoost		AdaBoost	
	zero	KNN	zero	KNN
20%	6.614±0.171	6.570±0.143	7.963±0.238	5.515±0.626
40%	7.798±0.270	7.788±0.228	9.012±0.109	6.138±0.471
60%	8.923±0.311	8.888±0.260	9.267±0.223	3.988±0.377
80%	9.679±0.169	9.656±0.218	10.496±0.502	2.537±0.263

To Do List

- ✓ TabNet cross validation 으로 진행
- ✓ 결측치 20%/모델 3개/imputation method 3개 추가해서 실험 진행

Thank you

Paper review on gpt

Chaelyn Lee

mylynchae@gmail.com

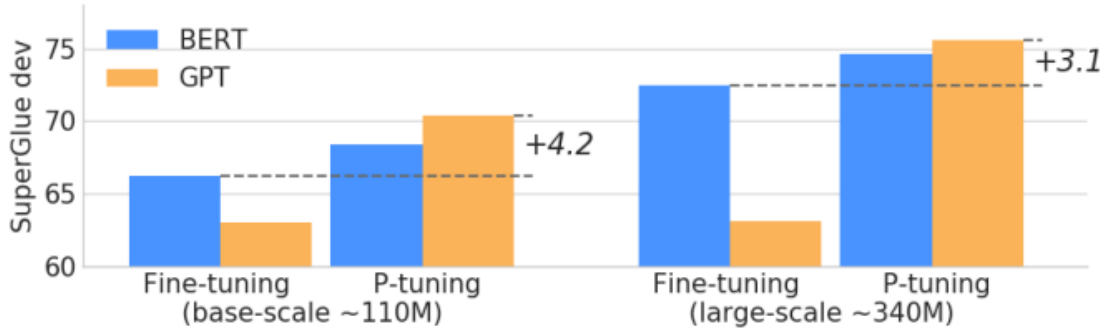
23rd Nov, 2022

To Do List from Previous Seminar

- ✓ Paper survey on new research
  - recommended for research related to undergraduate research.
  - If not, list up and read accepted papers from top conferences.

- ✓ Reading acl papers at nlp major conferences

motivation

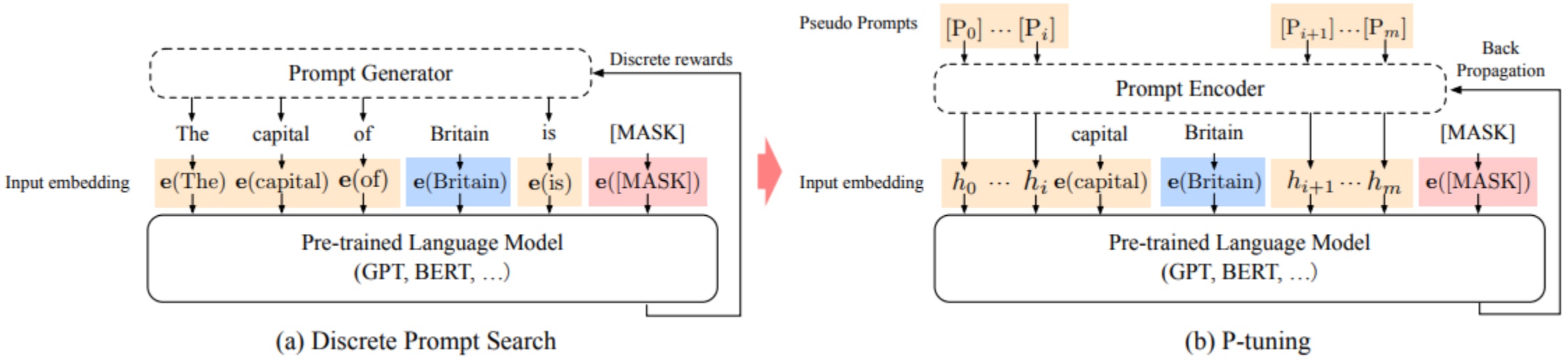


Prompt	P@1
[X] is located in [Y]. *(original)*	31.29
[X] is located in which country or state? [Y].	19.78
[X] is located in which country? [Y].	31.40
[X] is located in which country? In [Y].	51.08

- Previously, gpt performed poorly on NLU tasks compared to BERT.
  -> The paper suggests that gpt can outperform BERT through p-tuning.
- The table shows the difference in performance depending on the extent to which the prompt changes.
  -> Even a small change in Proft makes a big difference in performance.

GPT understands, too(*arXiv preprint arXiv:2103.10385,* 2021)

Method



기존 Prompt $\{\mathbf{e}([\mathrm{P}_{0:i}]), \mathbf{e}(\mathbf{x}), \mathbf{e}([\mathrm{P}_{i+1:m}]), \mathbf{e}(\mathbf{y})\}$

P-tuning $\{h_0, ..., h_i, \mathbf{e}(\mathbf{x}), h_{i+1}, ..., h_m, \mathbf{e}(\mathbf{y})\}$

Prompt Encoder $h_i = \mathrm{MLP}([\overrightarrow{h_i} : \overleftarrow{h_i}])$

$= \mathrm{MLP}([\mathrm{LSTM}(h_{0:i}) : \mathrm{LSTM}(h_{i:m})])$

GPT understands, too(*arXiv preprint arXiv:2103.10385,* 2021)

Experiments

Prompt type	Model	P@1
Original (MP)	BERT-base	31.1
	BERT-large	32.3
	E-BERT	36.2
Discrete	LPAQA (BERT-base)	34.1
	LPAQA (BERT-large)	39.4
	AutoPrompt (BERT-base)	43.3
P-tuning	BERT-base	48.3
	BERT-large	**50.6**

Model	MP	FT	MP+FT	P-tuning
BERT-base (109M)	31.7	51.6	52.1	52.3 (+20.6)
-AutoPrompt (Shin et al., 2020)	-	-	-	45.2
BERT-large (335M)	33.5	54.0	55.0	54.6 (+21.1)
RoBERTa-base (125M)	18.4	49.2	50.0	49.3 (+30.9)
-AutoPrompt (Shin et al., 2020)	-	-	-	40.0
RoBERTa-large (355M)	22.1	52.3	52.4	53.5 (+31.4)
GPT2-medium (345M)	20.3	41.9	38.2	46.5 (+26.2)
GPT2-xl (1.5B)	22.8	44.9	46.5	54.4 (+31.6)
MegatronLM (11B)	23.1	OOM*	OOM*	**64.2** (+41.1)

* MegatronLM (11B) is too large for effective fine-tuning.

Conclusion

- A large performance improvement was obtained by introducing a prompt-encoder with a relatively small size to GPT and BERT.

- Similar performance improvement can be expected by introducing p-tuning to models such as GPT3.

GPT understands, too(*arXiv preprint arXiv:2103.10385,* 2021)

GPT1

The challenges of learning important information from unlabeled data

- it is unclear what type of optimization objectives are most effective at learning text representations that are useful for transfer.
- there is no consensus on the most effective way to transfer these learned representations to the target task.

- gpt1 is a semi-supervised method using a combination of unsupervised pre-training and supervised fine-tuning.
- goal is to learn a universal representation that transfers with little adaptation to a wide range of tasks.

Improving Language Understanding by Generative Pre-Training (openAI, 2018)

To Do List

- ✓ Paper survey for selection of research topics
- ✓ Proceed in the order of finding the task first and studying the model in depth.

Thank you

Appendix