國立政治大學資訊管理學系

碩士學位論文

Text Segmentation and Name Entity Recognition for Memorials from the Qing Dynasty with Transformer-based Multitask Learning

基於Transformer之多任務學習用於清代奏摺斷句斷詞命名實體識別

指導教授：蔡瑞煌 博士、黃瀚萱 博士

研究生：薛卉吟 撰

中 華 民 國 110 年 11 月

DOI:10.6814/NCCU202101726

謝辭

一路上因為有許多人的幫助，才能順利完成這篇論文。感謝我的指導教授蔡瑞煌老師、黃翰萱老師，總是耐心的聆聽，協助我找出論文的方向，精準的給出修正的建議，並時常給予鼓勵，讓我有信心面對所有挑戰及難題。感謝我的口試委員金仕起老師、謝舒凱老師、林國平處長，細心的解答我的疑問，亦花很多時間閱讀我的論文，您們的專業彌補了我對奏摺與古文領域的不熟悉，讓這篇論文更加完整。感謝黃宇暘學長、周維強博士，在我對奏摺還一無所知的時候，幫助我建立相關知識，並給了我挑選奏摺的重要方向。感謝曾守正老師，雖然不是我的口試委員，仍然願意花時間了解我的論文內容，給予建議。感謝我的父母，無私的奉獻，總是支持與信任我的決定。最後，感謝政大，提供我優秀的學習環境，在這裡我獲益良多，不但開拓了眼界，專業能力與心境也成長許多，未來會帶著這些收穫，砥礪前行！

DOI:10.6814/NCCU202101726

ABSTRACT

Memorials are important materials for research on policy implementation and the formation of legal institutions. Although the memorials of Qing palace and the Grand Council had been accomplished with image scanning, the application is still not popular in academia. One of the reasons is that classical Chinese will often take a lot of historian's time to determine the segmentation of sentences and the meaning of words. The use of natural language processing (NLP) tools for analyzing classical Chinese remains an emerging topic in the digital humanity community. For classical Chinese, there are few NLP tools, and the performance of artificial intelligence (AI) models is not the same after learning the data of different dynasties. To address the challenges regarding the memorials of Qing dynasty, this study proposes a classical Chinese analysis model with transformer-based single task learning (STL) and multitask learning (MTL) that simultaneously copes with three tasks for classical Chinese: word segmentation, sentence segmentation, and the joint task for part-of-speech (POS) tagging and named entity recognition (NER). To accomplish the goal, the labels have three parts: (1) BOE format tags for sentence segmentation, (2) BIES format tags for word segmentation, and (3) the joint tags for POS and NER. For evaluating the proposal, this study focuses on the Yong-zheng (雍正) emperor and the Qing's memorials dataset annotated with new tagging schemes by Chinese professionals is collected. The research results show that method MTL performs significantly better on both sentence segmentation task and word segmentation task than method STL. And on POS+NER task, there is no significant difference between the two methods. The prediction of the memorials can help scholars to read memorials easily and reduce the probability of misinterpretation of word meaning.

Keywords: Memorial, Qing Dynasty, Transformer, BERT, Sentence segmentation, Word segmentation, Name entity recognition, Multitask learning, Classical Chinese, NLP

DOI:10.6814/NCCU202101726

摘要

奏摺，是研究清代政策實施和法制建設的珍貴的史料。雖然存於國立故宮博物院的清代宮中檔及軍機處的奏摺已完成數化，但應用仍然不普及，原因之一是辨識古典漢語的斷句、斷詞和詞義需花費歷史學家大量的時間。對於古典漢語，很少有有用的自然語言處理（NLP）工具，並且先進的人工智能（AI）模型學習不同朝代的訓練數據後，其性能也不盡相同。此外，沒有合適的NLP工具來分析清代的奏摺。為了解決有關於分析清代奏摺的挑戰，本研究探索一種基於Transformer之單任務學習（STL）及多任務學習（MTL）之模型，該模型可同時應付以下三個任務：斷句、斷詞、詞性（POS）標記和命名實體識別（NER）。為了完成此任務，本研究建議的標記方案包括三個部分：（1）用於斷句的BOE格式標籤；（2）用於斷詞的BIES格式標籤；以及（3）用於POS和NER的聯合標籤。為了評估該提案，本研究著重於雍正皇帝時期之奏摺，並收集並建立由中文專業人士參照新標籤標記方案所標註的清朝宮中檔奏摺數據集。研究結果顯示，斷句及斷詞任務中，多任務學習效能顯著優於單任務學習，兩個學習方法在詞性標記和命名實體識別則無顯著差異。模型的斷句結果可以達到輔助初學者們閱讀奏摺，斷詞以及詞性的標注結果則可以協助學者辨認詞義，減少對詞義誤讀的可能。

關鍵字：清代奏摺、斷詞斷句、命名實體識別、多任務學習、自然語言處理、機器學習、古文

DOI:10.6814/NCCU202101726

INDEX

DOI:10.6814/NCCU202101726

TABLE INDEX

DOI:10.6814/NCCU202101726

FIGURE INDEX

DOI:10.6814/NCCU202101726

1 INTRODUCTION

Archival documents are not only records produced by government agencies in their administrative activities, but also important materials for research on policy implementation and the formation of legal institutions. A memorial, which is also called “zou-zhe” (奏摺) in Qing dynasty, was the most important form of document sent by an official to the Emperor [30]. The memorials record the local events, contain very rich local historical records such as local traditions, fiscal economics, crimes, etc. These provide valuable information for area studies of the Qing dynasty [1]. The National Palace Museum has a collection of over 150 thousand Qing palace memorials and over 190 thousand items in the archives of the Grand Council. Although the memorials of Qing palace memorials and Grand Council had been accomplished with image scanning, the application is still not popular in academia. One of the reasons is that classical Chinese is too hard to read and understand, and then ambiguity is often inherited. It will often take a lot of historian’s time to determine the segmentation of sentences and the meaning of words. Another problem with the database is that the original metadata fields that can be searched are based on the needs of the manager. It has seven items: document number, name of the writer, official position of the writer, time of the memorials sent, red comment date, red comment, and content summary. It needs huge human resources to fill information into the new metadata schema which has 57 fields at present. To solve this problem, the automatic system with the documents from the Qing dynasty is necessary [3]. And even if it's completed to fill information into the new metadata schema, scholars still cannot use full-text search to obtain information.

The modern punctuation mark of Chinese appeared in the 20th century [2], thus, the classical Chinese texts from the end of the Spring and Autumn period (early 5th century BC) to the end of the Han dynasty [21], almost have no punctuation marks. Researchers need to segment themselves and the ambiguity still happens sometimes. If the natural language processing (NLP) model can segment sentences automatically, it can help readers read faster for analyzing classical Chinese documents.

The “word” is the basic semantic unit in NLP, unlike English which has a natural boundary between the words, the various length of characters can be a word in Chinese. Moreover, the word composition and the meaning of the word between classical Chinese and modern Chinese were constantly different. To define meaning of words, part-of-speech (POS) and name entities (NE) is also important in classical Chinese text analysis.

DOI:10.6814/NCCU202101726

Nowadays, many tools of modern word segmentation are available, such as Jieba[1]. Baidu's lexical analysis system published on Baidu AI open platform can be used for three individual analysis tasks: word segmentation, POS tagging, and named entity recognition (NER). And they also constructed an open-source toolkit LAC[2] public on GitHub to jointly accomplish all three tasks with deep bidirectional-gate recurrent unit (GRU)- conditional random fields (CRF) Network [17]. The CKIP (Chinese knowledge and information processing) Lab released CKIP Tagger[3] and CKIP Transformers[4] for word segmentation, POS, and NER tasks. But there are few widely used tools for classical Chinese, and models with different dynasty trained data have dissimilar performance, and no suitable tool to analyze memorials of Qing dynasty.

The bidirectional encoder representations from transformers, as known as BERT, has released the Chinese version of pretrain model in 2018. It allows the same pre-trained model to successfully tackle various NLP tasks, and also can be used as an embedding layer in learning-based models [11]. Dai et al. [10] have used the BERT-BiLSTM-CRF model for Chinese Electronic Health Records, and have obtained a good performance.

This work builds a Chinese NLP model with transformer-based single task learning (STL) and multitask learning (MTL) for documents from the Qing dynasty with three tasks: sentence segmentation, word segmentation, and POS+NER. For training this model, the annotated Qing's memorials dataset with new tagging scheme is also necessary to built, and the Yong-zheng（雍正） emperor reigned was been focus on. The tagging scheme has three parts: (1) BOE format tags for sentence segmentation, (2) BIES format tags for word segmentation, and (3) the joint tags for POS and NER.

1 Jieba. Retrieved on March 17 2021, from: https://github.com/fxsjy/jieba

2 Jiao, Z., Sun, S., & Sun, K. (2018). Chinese lexical analysis with deep Bi-GRU-CRF network. *arXiv preprint.* https://arxiv.org/abs/1807.01882

3 Li, P. H., Fu, T. J., & Ma, W. Y. (2020). Why Attention? Analyze BiLSTM Deficiency and Its Remedies in the Case of NER. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, 34*(05), 8236-8244. https://doi.org/10.1609/aaai.v34i05.6338

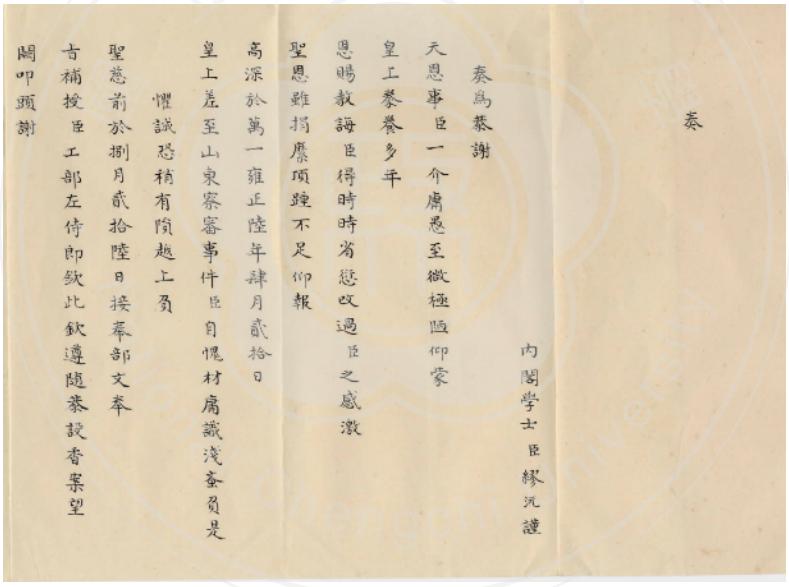
4 Yang, M., Ma,W. Y. (2020). *CKIP Transformers*. CKIP Lab. https://github.com/ckiplab/ckip-transformers

DOI:10.6814/NCCU202101726

2 PREVIOUS WORKS

2.1 Qing Palace Memorials of National Palace Museum

The [1] published by the National Palace Museum describes the Qing palace memorials detilly. The Qing palace memorials currently in the collection of the National Palace Museum are mainly memorials made by the emperors and Archives of the Grand Council's ministers of the Qing dynasty. In brief, the memorials shown in Figure 1, which is also called "zou-zhe" (奏摺) are the communications from the ministers to the emperors. From ancient times, the official communications from ministers to the emperors has a lot of format, such as biao (表), zhang (章), shu (疏), zhou (奏), shu (書), and fong-shi (封事). In Ming dynasty, the format of zhou-ben (奏本) and ti-ben (題本) were formulated. The early Qing dynasty's zou-zhe was an adaptation of Ming dynasty's zhou-ben. And till the K'ang-hsi (康熙) emperor reigned, the system of the memorial officially appeared [1].


奏

内閣学士臣椿泥謹

奏爲恭謝

天恩事臣一介庸愚至微極陋仰蒙

皇上參養多年

恩賜教誨臣得時時省愆改過臣之感激

聖恩雖捐糜頂踵不足仰報

高深於萬一雍正陸年肆月貳拾日

皇上差至山東察審事件臣自愧材庸識淺奎負是

恩誠恐稍有隕越上負

聖慈前於捌月貳拾陸日接奉部文奉

古補授臣工部左侍郎欽此欽遵隨恭設香案望

闕叩頭謝

Figure 1: The Qing's memorial from National Palace Museum

The memorials are not only the communication document, but also the important means for emperors to collect opinions and gather information from all over the country. Whether it is a public or private incident, ministers must report truthfully as a reference for the emperor's decision. The memorials record the local events in great detail because the main resource of it are the expatriate officers. Thus, memorials contain very rich local historical records such as local traditions, fiscal economics, crimes, etc. These provide valuable information for area studies of the Qing dynasty [1].

DOI:10.6814/NCCU202101726

After rebuilding, the National Palace Museum organized the Qing palace memorials actively, sorted the number, date, position, name, abstract, the replay by emperor, and the category index and name index for easy look up. The National Palace Museum has a collection of over 150 thousand Qing palace memorials, the most well-preserved reigned is the Yong-zheng (雍正) emperor reigned, about 22 thousand memorials [1].

2.2 Chinese Text Classification Tasks

There are four text Classification tasks in this paper: sentence segmentation, word segmentation, POS tagging, and NER [15].

Sentence Segmentation Unlike modern Chinese text, classical Chinese texts seen in historical documents almost have no punctuation marks. Readers must segment themselves during the reading process but the ambiguity still arises. Thus, the sentence segmentation task is needed specially in classical Chinese documents. Huang, Sun and Chen [15] used the CRF segmenter applied on classical Chinese sentence segmentation. Gu, Wu and Zhang [13] post the architecture of CNN to resolve Chinese sentence classification tasks. Han et al. [14] added a component of radical embedding and a long short-term memory (LSTM)-CRF model to improve performances.

Word Segmentation Many Asian languages including Chinese do not put space between words. The "word" is seen as a basic semantic unit for NLP, and in Chinese, different lengths of characters can be a word. Thus, word segmentation is one of the prerequisite tasks in NLP for these languages [4]. In 2003, ICTCLAS, a hierarchical hidden markov model (HHMM) based Chinese lexical analyzer, had competitive performance [32]. Huang and Wu [16] proposed an approach to deal with classical Chinese word segmentation without any marked-up corpus. Ma, Ganchev and Weiss [19] found that a bidirectional LSTM model had good performance for the task of Chinese word segmentation.

Part-Of-Speech Tagging The statistical methods like Hidden Markov models (HMM) , Maximum Entropy models (MEMM), and CRF models were used in Chinese POS tagging initially [6][20][24]. Ng and Low [20] proved feature representation in word-based and character-based within a maximum entropy framework had different advantages in Chinese POS tagging. Because Chinese has no natural boundary between words, word segmentation is often done with POS tagging, even resolving jointly [26][27][28]. Shao et al. [25] used Bidirectional recurrent neural networks (RNN)-CRF to train character-based joint segmentation and POS tagging.

DOI:10.6814/NCCU202101726

Named Entity Recognition Named Entities (NE) were defined as proper names and quantities of interest, which consisted of three subtasks (entity names, temporal expressions, number expressions) in MUC-7[5]. NER is a sub-task of information extraction, and is widely used in natural language understanding applications. Since Chinese NER is more complicated and difficult, the approaches that are successfully applied in English cannot be simply used to deal with the problems of Chinese NER [33]. Wu et al. [31] used the deep neural network applied on Chinese clinical NER tasks. The NLP department of Baidu used Bi-GRU-CRF to train for joint lexical analysis tasks of word segmentation, POS tagging and NER [17]. Gong, Zhang and Chen [12] used BiLSTM and Transformer based on pretrain entity recognition models for Chinese electronic medical records.

Tagging Schema Han et al. [14] changed the BIO format of sentence segmentation tag into BOE format, which indicates the beginning of the sentence, others , the end of the sentence, respectively, and placed punctuation marks between E and B. The "BIES" is a popular labeling scheme of word segmentation [29][19]. The form of tags used by Jiao et al. [17] was "*t*-[BI]", which *t* was "B" or "I" indicating the character position of the word, and [BI] would be the joint tag of POS and NER tasks. They didn't involve the "O" tag because there was no outside character when POS and NER tasks were handled jointly. The POS tagging set of Ancient Chinese Corpus [7] shows as Table 1.

Table 1: POS set of ancient Chinese corpus [7]

#	Tag	POS	Example (Chinese_English Trans)
1	a	adjective	大_big
2	c	conjunction	則_then
3	d	adverb	不_not
4	f	locative	前_front
5	j	combined	焉_at there
6	m	number	一_one
7	n	noun	人_human
8	nr	person	孔子_Confucius
9	ns	location	齊_Qi (state name)
10	p	prepositional	於_at

[5] Chinchor, N. (1997). MUC-7 Named Entity Task Definition.

DOI:10.6814/NCCU202101726

11	q	classifier	匹_classifier for horse and wolf
12	r	pronoun	吾_me
13	s	onomatopoeia	譆譆_LOL
14	t	time	五月_the Fifth month
15	u	aux	之_of
16	v	verb	如_go
17	y	modal	乎_interrogative

2.3 Bidirectional Encoder Representations from Transformers

BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning in all layers, and can easily be fine-tuned with just one additional output layer. It's allowing the same pre-trained model to successfully tackle various NLP tasks [11]. There are two model sizes has been reported:

- $BERT_{BASE}$ (L=12, H=768, A=12, Total Parameters=110M)
- $BERT_{LARGE}$ (L=24, H=1024, A=16, Total Parameters=340M)

The architecture of the BERT is a multi-layer bidirectional Transformer encoder, and uses bidirectional self-attention. Unlike the traditional left-to-right or right-to-left language models, for example, GPT Transformer can only attend to previous tokens in the self-attention layers, BERT can use both left and right context at every layer [11].

The BERT had been pre-trained by two unsupervised tasks:

- Masked Language Modeling (MLM)
- Next Sentence Prediction (NSP)

The MLM learned to understand the relationship between words, and the NSP learned to understand the relationship between sentences. Therefore, for a given token, its input representation is the summarization of the corresponding token, segment, and position embeddings. Figure 2 gives an illustration of the construction [11].

DOI:10.6814/NCCU202101726

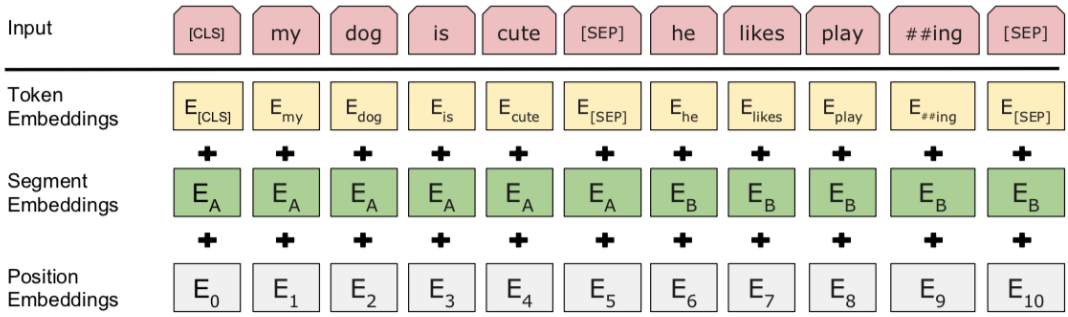


Figure 2: BERT input representation [11]

There are two approaches using the BERT, fine-tuning approach, and the feature-based approach. The fine-tuning approach, where a simple classification layer is added to the pre-trained model, and jointly fine-tuned all parameters on a downstream task (Figure 3). The feature-based approach has certain advantages within extracted features from the pre-trained model: no-needed to add task-specific model architecture, using the expensive pre-compute representation to run many experiments with cheaper models on top of this representation. In the previous work, fine-tuning the entire model was only better than the method that concatenated the token representations from the top four hidden layers of the pre-trained Transformer [11] by 0.3% of f1-score.

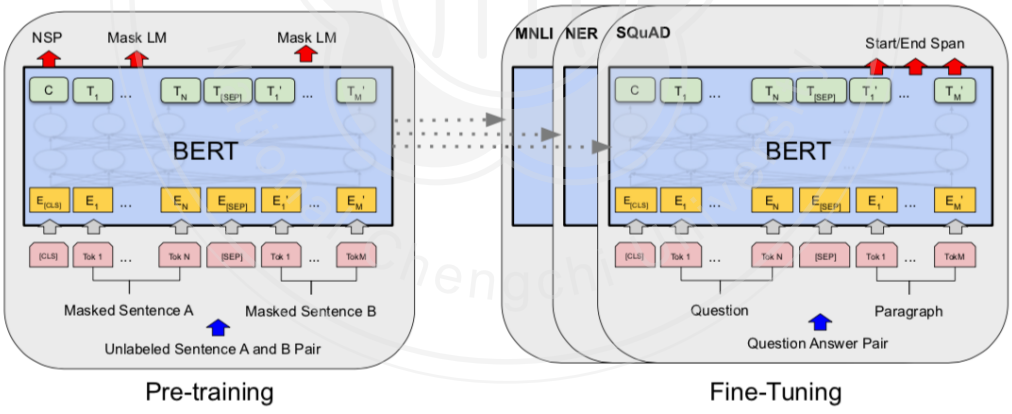


Figure 3: BERT pre-training and fine-tuning [11]

Dai et al. [10] proposed the NER task for Chinese electronic health records with transformer-based model, which contains a BERT pre-trained embedding layer, a BiLSTM and a CRF layer. They have compared several neural models for NER tasks, including CNN-LSTM, BiLSTM, BiGRU-CRF, BiLSTM-CRF, W2V-BiLSTM-CRF, BERT-BiLSTM-CRF. And they found that the BERT-BiLSTM-CRF model was the outperformed one which achieved approximately 75% F1 score during the tests. Qin, Zhao and Liu [23] used the BERT-BiGRU-CRF Model for entity recognition of Chinese electronic

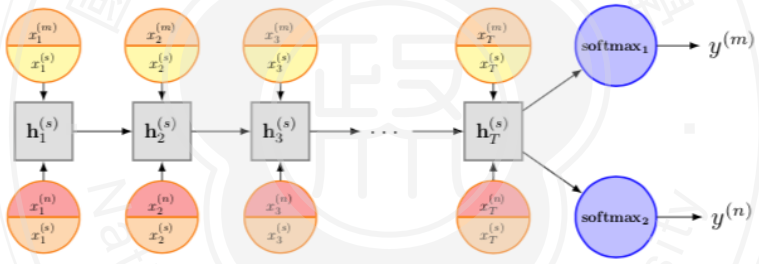
DOI:10.6814/NCCU202101726

medical records in 2021. That model used the BERT layer to use the context information and obtained the optimal tagging sequence through the BiGRU-CRF.

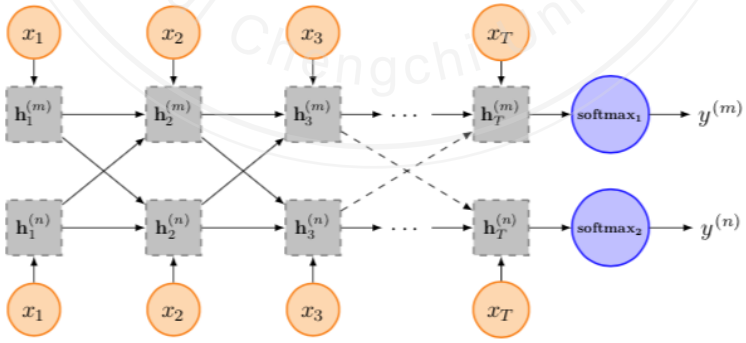
2.4 RNN-based Multi-Task Learning

Multi-task learning (MTL) is one way of sharing some lower layers to determine common features to achieve inductive transfer between *related tasks*. The MTL nets learn tasks in parallel while sharing some lower layers and representations; what is learned for each task can help [5]. After the shared layers, the remaining layers are splited for the multiple specific tasks.

Liu, Qiu and Huang [18] proposed three multi-task architectures of sharing information with RNN, which learns to map arbitrary text into semantic vector representations with both task-specific and shared layers in 2016. Showing that joint learning can improve the performance of each task relative to learning them separately. These models are shown in Figure 4.

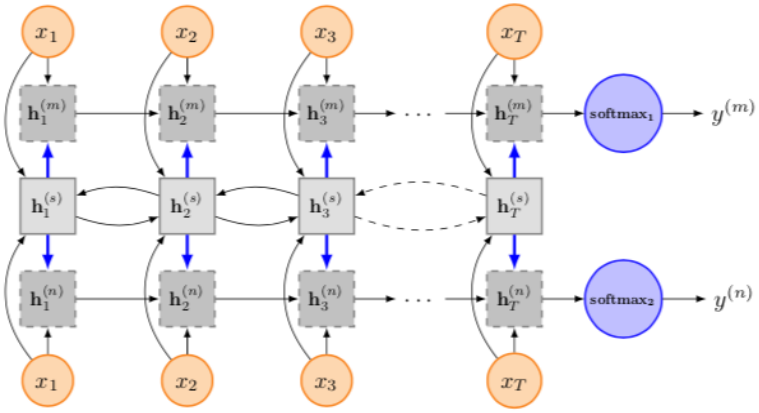


(a) Model-I: Uniform-Layer Architecture



(b) Model-II: Coupled-Layer Architecture

DOI:10.6814/NCCU202101726



(c) Model-III: Shared-Layer Architecture

Figure 4: Three architectures for modelling text with multi-task learning [18]

- **Uniform-Layer Architecture**: Using just one shared LSTM and embedding layer for all the tasks.
- **Coupled-Layer Architecture**: Using different LSTM layers for different tasks, but each layer can read information from other layers.
- **Shared-Layer Architecture**: Assigning a separate LSTM layer for each task, and a shared bidirectional LSTM layer to capture the shared information for all the tasks.

Each of these models was trained jointly for four text classification tasks in a single system. And in the research, the shared-layer architecture gives the best performances [18]. After that, Chen et al. [8] wrote the hidden states of shared layer and private layer precisely as follows:

$$\mathbf{h}_t^{(s)} = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}^{(s)}, \theta_s),$$
$$\mathbf{h}_t^{(k)} = \text{LSTM}(\begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}_t^{(s)} \end{bmatrix}, \mathbf{h}_{t-1}^{(k)}, \theta_k)$$

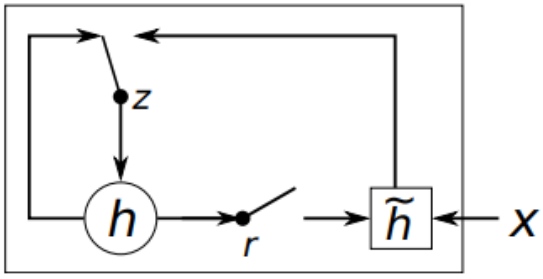
where $h_t^{(s)}$ and $h_t^{(k)}$ are hidden states of the shared layer and the private layer of task k at step t respectively; θ_s and θ_k denote their parameters. In task-specific output layers, they use softmax and CRF for text classification and sequence tagging.

2.5 Bidirectional Gate Recurrent Unit

The RNN is widely used in sequence to sequence tasks, and the LSTM and the GRU are commented on Chinese classification tasks [10][14][17][18][22][23][25][29], because they can involve the dependencies between elements of sequences. Cho et al. [9] improved GRU from LSTM but is much simpler to compute and implement. The GRU units combine the

DOI:10.6814/NCCU202101726

forget and input gates into a single “update gate”, and merge the cell state and hidden state. The GRU cell define as follows:



$$r_j = \sigma\left([W_r x]_j + \left[U_r h_{(t-1)}\right]_j\right) \quad (1)$$

$$z_j = \sigma\left([W_z x]_j + \left[U_z h_{(t-1)}\right]_j\right) \quad (2)$$

$$h_j^{(t)} = \varphi\left([Wx]_j + \left[U\left(r \odot h_{(t-1)}\right)\right]_j\right) \quad (3)$$

$$\tilde{h}_j^{\langle t \rangle} = \varphi\left([Wx]_j + \left[U\left(r \odot h_{(t-1)}\right)\right]_j\right) \quad (4)$$

Figure 5: GRU units. The update gate z selects whether the hidden state is to be updated with a new hidden state $h\tilde{}$. The reset gate r decides whether the previous hidden state is ignored. See Eqs. (1)–(4) for the detailed equations of r, z, h and $h\tilde{}$ [9].

Bidirectional GRU (BiGRU) which combines the forward GRU and reversed direction GRU and concatenates their results as output , is an extension to GRU [17]. Like the BiLSTM, the BiGRU involves the dependencies of both the next and previous element of the sequence.

DOI:10.6814/NCCU202101726

3 EXPERIMENT DESIGN

3.1 Models

According to the previous work [10][11], the transfer learning with the Chinese pre-train BERT, and the feature-based approach is adopted in this paper. The available Chinese pre-trained model "BERT-Base, Chinese[6]" is based on $BERT_{BASE}$ with 12-layer, 768-hidden, 12-heads, and 110M parameters. This study also revises BERT-BiGRU-CRF model [23] to create two model structures as two methods: one method is for single task learning (STL) and another is for MTL with the shared-layer architecture [8]. Table 2 shows these 2 different methods in this study.

Table 2: Two methods of this study

Method	Model Structure
STL	BERT-BiGRU-CRF with STL
MTL	BERT-BiGRU-CRF with MTL

In the model structure of method STL shown in Figure 6, the pre-trained Chinese BERT is the embedding layer to yield the Chinese character embedding vector, then follows the BiGRU layer and the CRF layer. Figure 7 shows the model structure of method MTL, after the pre-trained Chinese BERT embedding layer is the shared BiGRU layer, followed by the distinctive GRU and CRF layers for each of three tasks, respectively. Especially, the residual connections, which is the mechanism that the input of a layer is element-wise added to the output before feeding to the next layer, is useful to solve gradient problems. In the MTL model-III structure shown in previous Figure 4 [18], the output of the shared BiLSTM layer and the pre-trained embedding are combined as the input vector of priviated LSTM layer. Thus, the model of this study also adopts this mechanism, the concatenated vectors of character embedding and shared BiGRU output are the inputs of private GRU layers.

[6] Turc, I., Chang, M. W., Lee, K., & Toutanova, K. (2019). Well-Read Students Learn Better: On the Importance of Pre-training Compact Models. *arXiv preprint.* https://arxiv.org/abs/1908.08962

DOI:10.6814/NCCU202101726

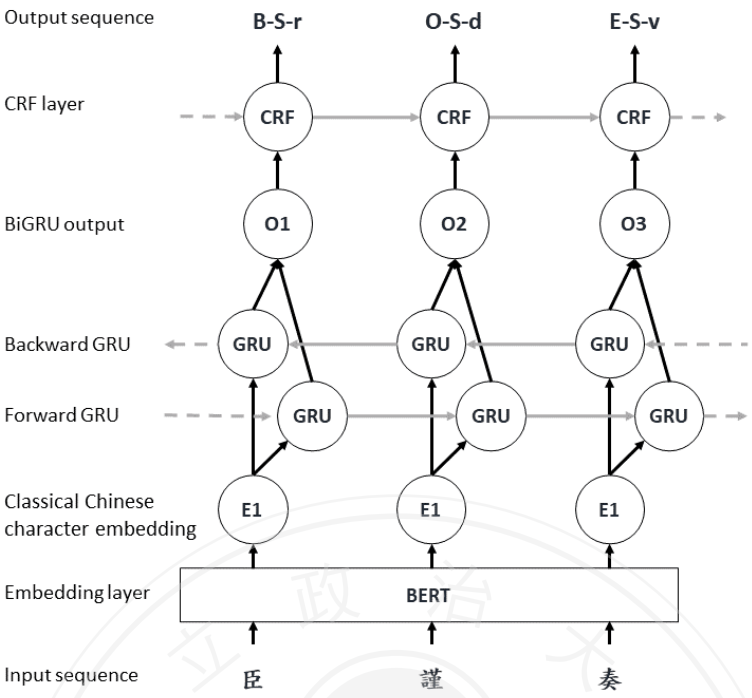


Figure 6: Model structure of method STL

The operation of the method STL is as follows:

1) In the embedding layer, each classical Chinese character in the sequence is mapped to a classical Chinese character embedding based on the pre-trained BERT.
2) These embeddings are the input to the BiGRU layer.
3) Input the output vectors of the BiGRU layer to the CRF layer. Then the CRF layer predicts the optimal output sequence by capturing the dependencies across adjacent labels.
4) Tuning the weight of the BiGRU layer and the CRF layer based on the loss between real labels and predicted labels.
5) Repeat steps 1) through 4) 50 epochs.

DOI:10.6814/NCCU202101726

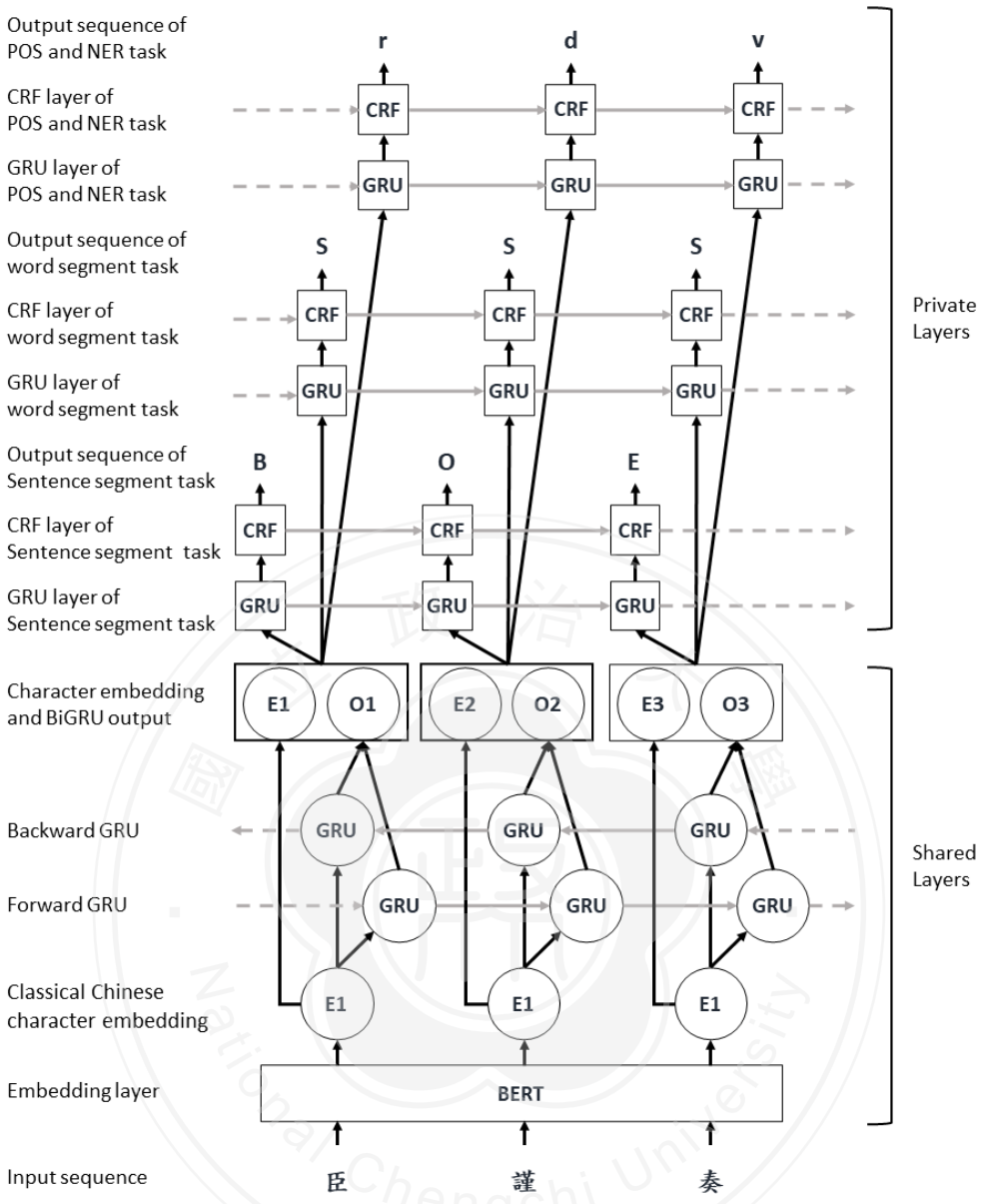


Figure 7: Model structure of method MTL

The operation of the method MTL is as follows:

1) The embedding layer processes each classical Chinese character in the input sequence into a classical Chinese character embedding based on the pre-trained BERT.
2) These embeddings are the input to the shared BiGRU layer, and with residual connection, the output of the shared BiGRU layer and the pre-trained embedding are combined into one vector.
3) Input the concatenated vectors to each priviated GRU of three tasks. Then the CRF layers predict the optimal output sequence by capturing the dependencies across adjacent labels.

DOI:10.6814/NCCU202101726

4) Tuning the weight of shared BiGRU layer and priviated GRU and CRF layers based on the loss between real labels and predicted labels.
5) Repeat steps 1) through 4) above 50 epochs.

The hyperparameters used in the models are modified according to [23] as shown in Table 3 below, and will be tuned in the experiment phase**.**

Table 3: Hyperparameters

Hyperparameter	Version
Embedding dimension	768
GRU dimension	200
Batch size	16
Learning rate	7e-5
Dropout	0.3
Sequence length	200
Epoch	50

3.2 Input X

The input x for the model are token sequences, a Chinese character will be a token. Each article in the dataset will be splitted into different parts which length is less than or equal to 200. And these parts are used as input sequences with the model.

3.3 Output Tags

Three tags used in this study : Sentence segmentation tag *S*, word segmentation tag *W*, and joint tag of POS and NER *J*. The tag *S* has 3 classes, the tag *W* has 4 classes, and the tag *J* which has two schemes are 18 and 24 classes, respectively. In method STL, a new tagging mechanism that integrates the tagging scheme of sentence segmentation task, word segmentation task, POS+NER tasks has been explored. Each Chinese character will be labeled with a joint label. The joint label will show as *S*-*W*-*J*, and the total number of each unique tag is 216 and 288.

3.3.1 Sentence Segmentation Tags

In previous work [14], they found there are significant features both of the beginning and the end of a sentence. And changed the popular NER tagging method BIO format into BOE tags. The “B” tag means this character is at the beginning of the sentence, “O” indicates others, “E” means this character is at the end of the sentence. This study followed it and used BOE tags as our sentence segmentation tags.

DOI:10.6814/NCCU202101726

Table 4: Sentence segmentation tags of this study

Tag	Notes
B	This character is at the BEGINNING of the sentence.
O	Others
E	This character is at the END of the sentence.

3.3.2 Word Segmentation Tags

This research handled POS and NER tasks jointly, thus there is no outside character [17]. Therefore, the BIES tagging scheme, same as the previous work, was adopted [29] to recognize boundaries between words. The "B" tag means this character is at the beginning of the word, "I" indicates inside, "E" means this character is at the end of the word, and "S" means this character is a word.

Table 5: Word segmentation tags of this study

Tag	Notes
B	This character is at the BEGINNING of the word.
I	The character is inside the word.
E	This character is at the END of the word.
S	This character is the word. (Single)

3.3.3 Joint Tags of POS and NER

The tagging schema designed for the ancient Chinese corpus of Zuozhuan [7] has 17 joint tags of POS and NER. The Qing's memorials are one of the official documents, so some particular words are often used in it. This study wants to know if more detailed annotations will have better results. In order to find the more suitable tagging scheme for memorials, two schemes are used to label the Qing's dataset.

Regular Tagging Scheme

This study adds another "np" tag to represent the "position" based on the POS tagging set of Ancient Chinese Corpus [7], since the positions in the Qing dynasty were meaningful. This scheme can be used regularly for the classical Chinese text. The details of the regular tagging scheme shown in Table 6 below.

Table 6: Regular tagging scheme of POS and NER

#	Tag	Part-of-speech	Example (Chinese_English Trans)
1	a	adjective	大 (da) : big
2	c	conjunction	則 (ze) : then

DOI:10.6814/NCCU202101726

3	d	adverb	不 (bu) : not
4	f	locative	前 (qian) : front
5	j	combined	焉 (yan) : at there
6	m	number	一 (yi) : one
7	n	noun	人 (ren) : human
8	nr	person	孔子 (Kong-zi) : Confucius
9	ns	location	齊 (Qi) : state name
10	np	position	內閣學士 (nei-ge-xue-shi) : secretary of the Grand Secretariat
11	p	prepositional	於 (yu) : at
12	q	classifier	匹 (pi) : classifier for horse and wolf
13	r	pronoun	吾 (wu) : me
14	s	onomatopoeia	譆譆 (xi-xi) : LOL
15	t	time	五月 (wu-yue) : the Fifth month
16	u	aux	之 (zhi) : of
17	v	verb	如 (ru) : go
18	y	modal	乎 (hu) : interrogative

Memorial Tagging Scheme

In addition to the "np" tag added in regular tagging scheme, experts helps to sort out 6 more tags for memorials : "nl" , "na", "nt", "nk", "mw", and "ow", representing the "law", "agency", "title", "knighthood", "modest words", and "official words". For the reason that the memorials are the official documents in the Qing dynasty, they have the specific format. The details of the memorial tagging scheme shown in Table 7 below.

Table 7: Memorial tagging scheme of POS and NER

#	Tag	Part-of-speech	Example (Chinese_English Trans)
1	a	adjective	大 (da) : big
2	c	conjunction	則 (ze) : then
3	d	adverb	不 (bu) : not
4	f	locative	前 (qian) : front
5	j	combined	焉 (yan) : at there
6	m	number	一 (yi) : one
7	n	noun	人 (ren) : human
8	nr	person	孔子 (Kong-zi) : Confucius
9	ns	location	齊 (Qi) : state name
10	np	position	內閣學士 (nei-ge-xue-shi) :

DOI:10.6814/NCCU202101726

			secretary of the Grand Secretariat
11	nl	law	律 (lv) : law
12	na	agency	司 (si) : agency name
13	nt	title	怡賢 (Yi-xian) : title of the prince
14	nk	knighthood	親王 (qin-wang) : prince
15	p	prepositional	於 (yu) : at
16	q	classifier	匹 (pi) : classifier for horse and wolf
17	r	pronoun	吾 (wu) : me
18	s	onomatopoeia	譆譆 (xi-xi) : LOL
19	t	time	五月 (wu-yue) : the Fifth month
20	u	aux	之 (zhi) : of
21	v	verb	如 (ru) : go
22	y	modal	乎 (hu) : interrogative
23	mw	modest words	竊 (qie) : the modest word
24	ow	official words	奏 (zou) : the official word

3.3.4 Example

In method MTL, there are three different tags for each character : (1) sentence segmentation tag, (2) word segmentation tag, and (3) joint tag of POS and NER**.**

Table 8-1: Sentence segmentation tags of method MTL

奏	內	閣	學	士	裏	行	走	臣
B	B	O	O	O	O	O	O	O
俞	兆	晟	謹	奏				
O	O	O	O	E				

Table 8-2: Word segmentation tags of method MTL

奏	內	閣	學	士	裏	行	走	臣
S	B	I	I	E	S	B	E	S
俞	兆	晟	謹	奏				
B	I	E	S	S				

Table 8-3: POS and NER tags of method MTL

奏	內	閣	學	士	裏	行	走	臣
ow	np	np	np	np	n	np	np	r
俞	兆	晟	謹	奏				
nr	nr	nr	ow	ow				

DOI:10.6814/NCCU202101726

The accurate label of method STL for each character shown in Table 9. One character corresponds to one joint label *S-W-J*. The *S* represents the sentence segmentation tag, the *W* represents the word segmentation tag, and the *J* is joint tags of POS and NER.

Table 9: Accurate joint labels of method STL

奏	內	閣	學	士	裡	行	走	臣
B-S-ow	B-B-np	O-I-np	O-I-np	O-E-np	O-S-n	O-B-np	O-E-np	O-S-r
俞	兆	晟	謹	奏				
O-B-nr	O-I-nr	O-E-nr	O-S-ow	E-S-ow				

3.4 Dataset

This paper obtained ancient texts from the National Palace Museum‘s Catalogue Database of Qing Palace Memorials and Archives of the Grand Council[7]. In this paper, the Qing dynasty that the Yong-zheng (雍正) emperor reigned was been focus on. The 50 memorials which obtain 34,390 characters were selected as the Qing’s dataset. Each article in the dataset will be splitted into different parts as input sequences whose length is less than or equal to 200, and the total number of sequences is 198.

This dataset was separated randomly into three parts: (1) training data, (2) validation data, and (3) testing data. The exact amount of these three parts is shown in Table 10. The 9-fold cross validation is used in this study, so that the training and validation part is randomly selected during training. In Section 3.4.1, how to collect the Qing’s dataset will be explained, and the data labeling would be discussed in Section 3.4.2.

Table 10: Qing’s datasets details of this study

	Number of sequences	**Number of characters**	**Total number of sequences**	**Total number of characters**
Training Data	160 (80%)	31209	198	34390
Validation Data	19 (10%)			
Testing Data	19 (10%)	3181		

3.4.1 Data Collection for Qing’s Dataset

In Yong-zheng emperor reigned, there were more than twelve thousand different official positions. This paper focuses on memorials belonging to “the grand secretary of the grand secretariat”, which was the assistance secretariat of the emperor. The query result of this

[7] Catalogue Database of Qing Palace Memorials and Archives of the Grand Council. National Palace Museum. Retrieved on February 2 2021, from: http://npmhost.npm.gov.tw/tts/npmmeta/GC/indexcg.html

DOI:10.6814/NCCU202101726

range showed 323 memorials, and in 45 different topics. Memorials with the same topic have similar article structure, thus, selecting more occurring topics would be more helpful for processing Qing's memorials in the future. The top 13 topics are shown in Table 11, total number is 124.

Table 11: The high-frequency occurrence topics of this study

Topic	Numbers
學政 (xue-zheng) : staff of educational affairs	19
請安 (qing-an) : greet	14
漕運 (cao-yun) : river transportation	14
遷調 (qian-diao) : postion transfer	13
地方事務 (di-fang-shi-wu) : local affairs	11
條陳 (tiao-chen) : propose and display	9
雨雪糧價 (yu-xue-liang-jia) : weather condition and price of grains	8
繳批 (jiao-pi) : submit for review	8
倉儲 (cang-chu) : stack and storage	7
河工 (he-gong) : river management	6
刑案 (xing-an) : criminal cases	5
錢糧 (qian-liang) : money and grains/ tax	5
吏制 (li-zhi) : bureaucratic system	5

The memorials of 請安 (qing-an) topic were not selected because they almost were short and the same for sending respects to the emperor. And so do for 繳批 (jiao-pi) topic that the contents had few common parts. The 漕運 (cao-yun) topic is also not selected because words used in this topic like quantifier, date, location, grain were common with 雨雪糧價 (yu-xue-liang-jia), 河工 (he-gong), 倉儲 (cang-chu), 錢糧 (qiang-liang) topics. And the 雨雪糧價 (yu-xue-liang-jia), 河工 (he-gong), 錢糧 (qiang-liang) topics are selected to learn these words. In the same topic, the memorials with longer length are selected, the accurate quantity is shown in Table 12 below.

Table 12: The selected topics of this study

Topic	Quantity
地方事務 (di-fang-shi-wu) : local affairs	9
條陳 (tiao-chen) : propose and display	8
雨雪糧價 (yu-xue-liang-jia) : weather condition and price of grains	7

DOI:10.6814/NCCU202101726

學政 (xue-zheng) : staff of educational affairs	6
遷調 (qian-diao) : postion transfer	5
刑案 (xing-an) : criminal cases	5
吏制 (li-zhi) : bureaucratic system	4
河工 (he-gong) : river management	3
錢糧 (qian-liang) : money and grains/ tax	3
Total	**50**

3.4.2 Data Labeling for the Qing's Dataset

In this paper, the new Qing's dataset which involved 50 Qing's memorials would be built. The memorials of the National Palace Museum's Catalogue Database of Qing Palace Memorials and Archives of the Grand Council were saved in image format. Therefore, the first thing that has to do is type these images into texts. After that, the texts would be labeled following the tagging scheme shown in Section 3.3 by Chinese professionals. There are some extra rules be followed during the labeling process:

1) Each word is separated by a half width space.
2) Each clause is separated by fullwidth "、".
3) Adding the full stop "。" at the end of each sentence.
4) Saved with UTF-8 encoding, .txt format.

The following is an example :

1) Original sentence :
   奏內閣學士裏行走臣俞兆晟謹奏
2) Labeled：
   奏/n 。 內閣學士/np 裏/n 行走/np 臣/r 俞兆晟/nr 謹/d 奏/v 。
3) Preprocessed :
   - Method MTL：奏内阁学士里行走臣俞兆晟谨奏:B,B,O,O,O,O,O,O,O,O,O,O,O,O,O, E:S,B,I,I,E,S,B,E,S,B,I,E,S,S:n,np,np,np,np,n,np,np,r,nr,nr,nr,d,v
   - Method STL：奏内阁学士里行走臣俞兆晟谨奏:B-S-n,B-B-np,O-I-np,O-I-np,O-E-np,O-S-n,O-B-np,O-E-np,O-S-r,O-B-nr,O-I-nr,O-E-nr,O-S-d,E-S-v

DOI:10.6814/NCCU202101726

3.4.3 Statistical Description of the Qing's Dataset

Table 13 and 14 shows statistical descriptions of two schemes with joint tags of POS and NER. These tables reveal that "f", "j", and "s" tags are not used in Qing's dataset, and the largest proportion tags are "n" and "v", which represent "noun" and "verb".

Table 13: Dataset statistical description with regular tagging scheme

#	Tag	Part-of-speech	Number	Percentage (%)
1	a	adjective	1802	5.24
2	c	conjunction	997	2.90
3	d	adverb	3433	9.98
4	f	locative	0	0.00
5	j	combined	0	0.00
6	m	number	1502	4.37
7	n	noun	**7779**	**22.62**
8	nr	person	1138	3.31
9	ns	location	1409	4.10
10	np	position	1799	5.23
11	p	prepositional	817	2.38
12	q	classifier	619	1.80
13	r	pronoun	1292	3.76
14	s	onomatopoeia	0	0.00
15	t	time	1754	5.10
16	u	aux	900	2.62
17	v	verb	**8984**	**26.13**
18	y	modal	165	0.48

Table 14: Dataset statistical description with memorial tagging scheme

#	Tag	Part-of-speech	Number	Percentage (%)
1	a	adjective	1766	5.14
2	c	conjunction	997	2.90
3	d	adverb	3267	9.50
4	f	locative	0	0.00
5	j	combined	0	0.00
6	m	number	1499	4.36
7	n	noun	**7273**	**21.15**
8	nr	person	1133	3.29
9	ns	location	1414	4.11
10	np	position	1664	4.84
11	nl	law	119	0.35
12	na	agency	380	1.10

DOI:10.6814/NCCU202101726

13	nt	title	19	0.06
14	nk	knighthood	17	0.05
15	p	prepositional	815	2.37
16	q	classifier	616	1.79
17	r	pronoun	1277	3.71
18	s	onomatopoeia	0	0.00
19	t	time	1754	5.10
20	u	aux	901	2.62
21	v	verb	**8646**	**25.14**
22	y	modal	165	0.48
23	mw	modest words	222	0.65
24	ow	official words	446	1.30

3.5 Experiment Environment

The "BERT-Base, Chinese" pre-train model with Chinese was released on Github by Google. And NLP researchers from HuggingFace[8] made a PyTorch version of BERT available which is compatible with these pre-trained checkpoints and is able to originalize our results. The Transformer library of HuggingFace was used in this paper, it provides thousands of pretrained language models to perform tasks on texts such as classification, information extraction, question answering, summarization, translation, text generation. With the PyTorch version pre-train BERT, Python 3.8.5 and PyTorch 1.6.0 setted up in our environment. This experiment also used a graphic processing unit GeForce GTX 1080 to Implement parallel computing, and the CUDA Toolkit version is 10.2.89.

Table 15: Tools of this study

Tool	Version
Python	3.8.5
PyTorch	1.6.0
CUDA Toolkit	10.2.89
GPU	GeForce GTX 1080

[8] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., …Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38-45.

DOI:10.6814/NCCU202101726

3.7 Evaluation

There are three methods to evaluate the performance of the model :

1) Loss function value for model.
2) Precision, recall, f1-score for the following three tags :
    - Sentence segment tag
    - Word segmentation tag
    - POS tag + NER tag
3) Confusion Matrix for POS tag + NER tag

To evaluate, there are different comparisons in this study :

- Residual connection
- Compare with other models
- Compare with other Chinese NLP tools
- Different tagging scheme of POS+NER
- Different granularity of word segmentation

DOI:10.6814/NCCU202101726

4 EXPERIMENTS

4.1 Preprocessing

This study takes 50 memorials which obtain 34,390 characters as the Qing's dataset. Firstly, convert all characters from traditional Chinese into simplified Chinese. Secondly, convert all skips that have full and half stop meaning in the sentences into " ◦ ". This experiment only focuses on the segmentation of sentences, and will not discours the type of punctuation marks. Furthermore, each article in the dataset will be splitted into different parts as input sequences whose length is less than or equal to 200, and the total number of sequences is 198. These sequences are split randomly into a training set (containing the validation set), and a testing set. The proportions are 90%, and 10%, respectively, and saved with the form that is easier to use by model as txt files finally.

4.2 Training

Table 16 shows 2 different methods compared in this study : (1) method STL , whose model structure is BERT-BiGRU-CRF with STL, and (2) method MTL , whose model structure is BERT-BiGRU-CRF with MTL. And because of the opinions of experts and the demand for memorials, the memorial tagging scheme is finally adopted in the experiment. The method STL consists of the BERT layer, 2 Bi-GRU layers (2 forward GRU and 2 reversed GRU), and a CRF layer. The method MTL consists of the BERT layer, 2 shared Bi-GRU layers, the distinctive GRU and CRF layers for each of three tasks.

Table 16: Methods of this study

Method	Model Structure	Tagging Scheme of POS+NER
STL	BERT-BiGRU-CRF with STL	Memorial
MTL	BERT-BiGRU-CRF with MTL	Memorial

In the experiment of this study, the hyperparameters are modified according to [23] shown in Table 17 below. To select better hyperparameters for the models, different values of hyperparameters, such as GRU dimension, learning rate, dropout, batch size, and epoch are tried during the experimentation. After that, the hyperparameters which can achieve better performance are used in the models. The character embeddings dimension is 768, and GRU layers have 200-dimensional hidden units. All the weight matrices in GRUs are initialized with random matrices, and optimization is performed using adam. The learning rate is set to

DOI:10.6814/NCCU202101726

7e-05, the dropout rate from the BERT layer is set to 0.3, the batch size is set to 16, and train the models in 50 epochs.

Table 17: Hyperparameters of this study

Hyperparameter	Version
Embedding dimension	768
GRU dimension	200
Optimizer	Adam
Learning rate	7e-05
Dropout	0.3
Batch size	16
Epoch	50
Sequence length	200

Figure 8 shows the model loss of two different methods. These two methods are effective for the reason that model losses have decreased. Method MTL has a faster rate of convergence than method STL. Both of the model loss reduction becomes slow after around epoch 30.

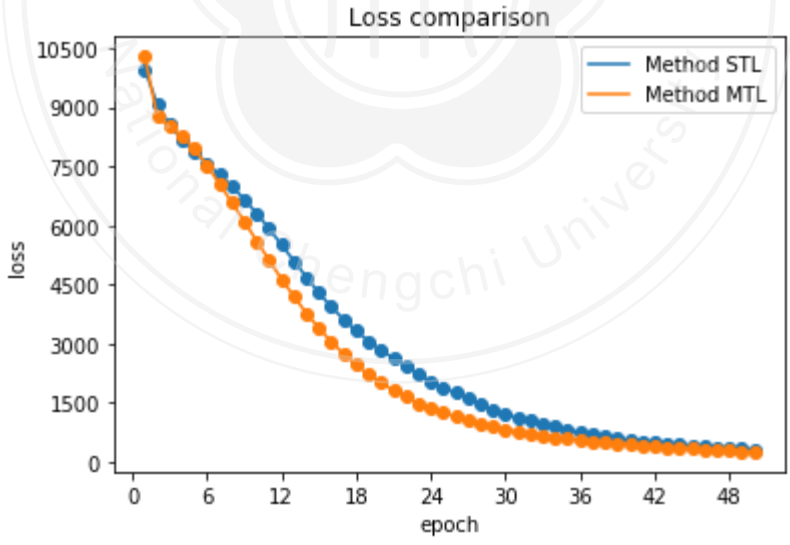


Figure 8: The train parse of two method

4.3 Evaluation

Table 18 shows the f1-score of sentence segmentation task, word segmentation task, and POS+NER task of two methods. As shown in the table, the f1-scores of sentence segmentation and word segmentation are higher with method MTL, which are 0.90538 and

DOI:10.6814/NCCU202101726

0.87991, respectively, and the p-values of these two tasks are lower than 0.05 ; the score of POS+NER is higher with method STL, but the p-value of POS+NER is higher than 0.05. It indicates method MTL performs significantly better on both sentence segmentation task and word segmentation task than method STL; and on POS+NER task, there is no significant difference between the two methods.

Furthemore, to evaluate the NER and POS task separately, the POS tag (“v”, “a”, “d”, tec.) replaced with “o” in NER task, and NER tag (“na”, “ns”, “np”, etc.) replaced with “n” in POS task. Table 18 indicates that both methods perform better on NER task than POS task, and the performance are not significantly different between method STL and method MTL.

Table 18: The f1-scores of two methods

Method	Sentence segmentation	Word segmentation	POS+NER	POS	NER
STL	0.89689	0.86514	0.85225	0.86577	**0.92361**
MTL	**0.90538**	**0.87991**	0.84156	0.85665	**0.91512**
p-value	7.73E-05	5.32E-05	2.33E-01	8.95E-01	5.26E-01

Thus, this study adopts method MTL to implement sentence segmentation task, word segmentation task, and POS+NER task on Qing’s dataset. Table 19 shows the prediction example of the testing data.

Table 19: Prediction example of the testing data

Original	奏/ow 內閣學士/np 兼/v 禮部/na 侍郎/np 福建/ns 觀風整俗使/np 臣/r 劉師恕/nr 謹奏/ow 。為/p 恭謝/v 天恩/n 事/n 。臣/r 濛/v 我/r 皇上/np 特/d 達/a 之/u 知/n 。屢/d 加/v 簡/v 擢/v 。
Prediction	奏/ow 內閣學士/np 兼/v 禮部/na 侍郎/np 福建/ns 觀風整俗使/np 臣/r 劉師恕/nr 謹奏/ow 。為/p 恭謝/v 天恩/n 事/n 。臣/r 濛/v 我/r 皇上/np 特/d 達/v 之/u 知/n 屢/d 加/v 簡擢/v 。

The confusion matrix of method MTL shown in Table 20 makes it easy to see whether the model is confusing two POS+NER tags. The tag “n” is confused with the tag “a” 41 times and “v” 59 times in 626 characters, about 6.55% and 9.42% respectively ; the tag “v” confuses with the tag “d” 37 times and “n” 36 times in total 821, about 4.51% and 4.38% respectively. The model performs best on tag “q”, and worst on tag “nt” and “nk”. It might be because there are few “nt” and “nk” tags in the training data, and the model is not learning well with these two tags.

DOI:10.6814/NCCU202101726

Table 20: The confusion matrix of method MTL regarding POS+NER tags

		Predicted Tag																							
		a	**c**	**d**	**f**	**j**	**m**	**n**	**nr**	**ns**	**np**	**nl**	**na**	**nt**	**nk**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**y**	**mw**	**ow**
Actual Tag	**a**	96	0	12	0	0	5	23	1	2	4	0	0	0	0	0	0	6	0	0	0	15	0	0	0
	c	0	71	5	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	1	2	0	0	0
	d	14	10	224	0	0	1	5	0	0	0	0	0	0	0	0	0	1	0	9	2	26	0	0	0
	f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	m	0	0	0	0	0	84	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
	n	**41**	1	7	0	1	1	493	3	2	4	0	2	0	0	0	4	4	0	3	0	**59**	0	0	1
	nr	0	0	0	0	0	0	2	98	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	ns	0	0	0	0	0	0	5	8	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	np	0	0	0	0	0	0	8	0	2	213	1	1	0	0	0	0	0	0	0	0	2	0	0	0
	nl	1	0	0	0	0	0	1	1	0	0	6	0	0	0	0	0	0	1	0	0	0	1	0	0
	na	0	0	0	0	0	1	3	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0
	nt	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	nk	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	p	0	2	0	0	0	0	1	0	0	0	0	0	0	0	75	0	0	0	0	3	2	0	0	0
	q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0
	r	1	2	1	0	0	3	3	0	0	0	0	0	0	0	0	0	108	0	0	3	1	2	0	0
	s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	t	4	0	2	0	0	0	4	0	0	0	0	0	0	0	0	0	1	0	177	0	0	0	0	0
	u	0	1	1	0	0	0	2	0	0	0	0	0	0	0	3	0	5	0	0	67	1	0	0	0
	v	17	2	**37**	0	0	1	**36**	3	0	1	0	0	0	0	4	0	2	0	0	2	712	0	1	3
	y	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	11	0	0
	mw	1	2	2	0	1	0	3	2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	10	0
	ow	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	46

DOI:10.6814/NCCU202101726

4.4 Comparisons

4.4.1 Residual Connection

The model structure of method MTL presented in Figure 7 adopts residual connection especially. The outputs of the shared BiGRU layer and the character embeddings from the pre-train BERT are combined into the concatenated vectors as input of priviated GRU layer. To observe the effect of this mechanism, models with residual connections or not are compared in this section. Table 21 shows the f1-scores of sentence segmentation task, word segmentation task, and POS+NER task. The model with residual connection performs better in the POS+NER task than the model without residual connection, which are 0.84708 and 0.78907 respectively. It might be because the POS+NER task is more complicated than the other two taks, and the model with residual connection can perform better on it.

Table 21: The f1-scores of method MTL with/without residual connection

Residual Connection	Method	Sentence segmentation	Word segmentation	POS+NER
N	2	0.90174	0.86087	0.78907
Y	2	0.90538	0.87991	**0.84156**

4.4.2 Compare with Other Models

Table 22 shows the f1-scores of method MTL and three other baseline models : (1) CRF (2) Bi-GRU (3) BERT. The CRF model has the lowest f1-scores among these four models on sentence segmentation task, word segmentation, and POS+NER task. Although the BERT model performed well, the f1-scores of methods MTL proposed by this study are higher of all three tasks than it. It reveals the model of this study has the best performance among these four models.

Table 22: The f1-scores of four different models

Model	Sentence segmentation	Word segmentation	POS+NER
CRF	0.80443	0.67305	0.61261
BiGRU	0.81354	0.70849	0.75868
BERT	0.89940	0.85848	0.83742
our model	**0.90538**	**0.87991**	**0.84156**

DOI:10.6814/NCCU202101726

4.4.3 Compare with Other Chinese NLP Tools

To ensure the word segmentation availability of our method MTL suggested in this study, the model compared with other two existing Chinese NLP tools mentioned earlier : (1) Jieba (2) CKIP Tagger. The f1-scores of these two tools and method MTL are shown in Table 23 below. Our model of this study has the obviously highest score 0.87991 among these three tools. These two existing tools are built for modern Chinese, so that they couldn't achieve better performance when dealing with ancient Chinese. Therefore, using the dedicated model while processing memorials is important and necessary.

Table 23: Word segmentation f1-scores of three different Chinese NLP tools

Tool	Word segmentation
Jieba (Accurate Mode)	0.53603
CKIP Tagger	0.65427
our model	**0.87991**

4.4.4 Different Tagging Scheme of POS+NER

Because of the opinions of experts and the demand for memorials, the memorial tagging scheme is adopted in the experiment. As shown in Table 24, the f1-scores of POS+NER tasks with the POS+NER regular tagging scheme is higher than the other two with the memorial tagging scheme. That may be because the dataset with the memorial tagging scheme has more complicated NER and POS tags, which makes the model learning more difficult. Thus, no matter method STL or method MTL, the model with the memorial tagging scheme performs worse than the model with regular tagging scheme for the POS+NER task.

Table 24: The f1-scores of different methods and tagging schemes

Method	Tagging Scheme of POS+NER	Sentence segmentation	Word segmentation	POS+NER
STL	memorial	0.89689	0.86514	0.85225
	regular	0.90101	0.86738	**0.85827**
MTL	memorial	0.90538	0.87991	0.84156
	regular	0.90792	0.87398	**0.85732**

DOI:10.6814/NCCU202101726

4.4.5 Different Granularity of Word Segmentation

The granularity[9] refers to the size in which data fields are subdivided. The finer granularity means the level of detail is higher. In the experiment, different granularities are used to segment the words. Table 25 shows an example of word segmentation with fine and finer granularity. In this example, the word segmentation with fine granularity is more in line with human's reading habits than the word segmentation with finer granularity.

Table 25: Example of word segmentation with fine and finer granularity

Finer granularity	茲/u 復/d 蒙/v 特/a 旨/n 實授/v 學士/np 。 奴才/r 現在/t 統領/v 黑龍江/ns 兵馬/n ， 遵/v 旨/n 於/p 四月/t 青/a 草/n 發/v 生/v 之/u 時/t ， 帶/v 領/v 前/v 進/v ， 自/d 當/d 竭/v 盡/v 駑/a 力/n ， 奮勉/a 前/v 驅/v ， 以/c 期/v 仰/v 答/v 深/a 恩/n 於/u 萬/m 一/m 耳/y 。
Fine granularity	茲/u 復/d 蒙/v 特旨/n 實授/v 學士/np 。 奴才/r 現在/t 統領/v 黑龍江/ns 兵馬/n ， 遵旨/v 於/p 四月/t 青草/n 發生/v 之/u 時/t ， 帶領/v 前進/v ， 自/d 當/d 竭盡/v 駑力/n ， 奮勉/d 前驅/v ， 以/c 期/v 仰答/v 深/a 恩/n 於/u 萬一/m 耳/y 。

Table 26 shows the word segmentation f1-score of two granularity levels with our model and other existing Chinese NLP tools. First, our model proposed by this study achieves a higher score with finer granularity than fine granularity. It's because under finer granularity, there are mose 'S' of the word segmentation tags. And the lower complexity of word segmentation tags, the simpler for model learning. However, two other existing Chinese NLP tools achieve higher scores with fine granularity than finer granularity. It proves that using fine granularity to segment words is more reasonable.

Table 26: Word segmentation f1-scores of two granularity levels

Tool	Finer granularity	Fine granularity
Jieba	0.41743	0.53603
CKIP Tagger	0.45690	0.65427
our model	**0.90230**	**0.86318**

[9] Keet C.M. (2013). *Encyclopedia of Systems Biology*. New York: Springer.

DOI:10.6814/NCCU202101726

4.5 Discussion

In order to observe the predicted result of the methods proposed by this study and know what experts' suggestions are. The qualitative analysis with experts is taken in this section. Two memorials are selected to be discussed, one is from the training set, and the other is from the testing data. Table 27 and Table 28 show the original article and the prediction of these two memorials. Table 29 is the interview questions and feedback from historians.

Table 27: The memorial in testing data

Original	奏內閣學士兼禮部侍郎福建觀風整俗使臣劉師恕謹奏為恭謝天恩事臣蒙我皇上特達之知屢加簡擢但臣才質庸愚愆尤叢集洪慈保全至再至三訓誨指示極詳極切所降之級恩予復還且更特畀重任應追之銀恩予寬免且更特賞養廉大德生成極於天地逾於父母臣夙夜兢兢方恐從前過端未能補救再造弘仁未能報稱今於九月二十二日在歸化縣地方接準部咨欽奉俞旨補授內閣學士兼禮部侍郎寵命頻頒總超格外皇上之待臣者如此其優如此其重若臣第尋常供職即為有負聖恩感激彌深惶悚倍切臣自雍正七年五月內陛辭以來犬馬依戀之私常縈寤寐茲拜溫綸方以得覲天顏為喜續於十月初二日復接部文仍留觀風整俗使之任仰望闕廷愈增孺慕惟有勉竭蟻誠宣揚聖化以副我皇上牖迪斯民至意上報高厚於萬一為此繕摺恭謝天恩伏乞皇上睿鑒謹奏雍正玖年拾月拾柒日
Prediction	奏/ow 內閣學士/np 兼/v 禮部/na 侍郎/np 福建/ns 觀風整俗使/np 臣/r 劉師恕/nr 謹奏/ow 。 為/p 恭謝/v 天恩/n 事/n 。 臣/r 蒙/v 我/r 皇上/np 特/d 達/v 之/u 知/n 屢/d 加/v 簡擢/a 。 但/c 臣/r 才質庸愚/a 。 愆尤/d 叢/d 集/v 。 洪慈保全/d 至/v 再/d 至/v 三/d 訓誨/v 指示/v 極/d 詳/a 。 極/d 切/a 所/u 降/v 之/u 級恩/n 予/v 復/d 還/v 。 且/c 更/d 特/d 畀/v 重/a 任/n 。 應/d 追/v 之/v 銀/n 恩/n 。 予/v 寬免/v 。 且/c 更/d 特/d 賞養/v 廉/a 大/n 德/n 生/v 成/v 。 極/d 於/p 天地/n 逾/d 於/p 父母/n 。 臣/r 夙夜/t 兢兢/a 。 方/d 恐/v 從前/t 過/v 端/n 未能/d 補救/v 。 再/d 造/v 弘仁/n 。 未/d 能/d 報/v 稱/v 。 今/t 於/p 九月/t 二十二日/t 在/p 歸化縣/ns 地方/n 接/v 準/v 部咨/n 。 欽/d 奉/v 俞旨/n 補授/v 內閣學士/np 兼/v 禮部/na 侍郎/np 寵命/n 頻/d 頒/v 總/d 超格/v 外/n 皇上/np 之/u 待/v 臣/n 者/r 。 如此/c 其/u 優/a 。 如此/c 其/u 重/a 。 若/c 臣/r 第/d 尋常/d 供職/v 。 即/d 為/v 有/v 負/v 聖恩/n 。 感激/v 彌/d 深/a 。 惶/ow 悚倍切。 臣/np 自/v 雍正/na 七年/np 五月/ns 內陛辭以來/np 。 犬/r 馬依戀/nr 之私/ow 常/p 縈寤/v 寐。 茲/n 拜/n 溫/r 綸/v 方/r 以得/np 覲/d 天/v 顏/u 為/n 喜/d 。 續/v 於十/a 月/c 初/r 二日復接

DOI:10.6814/NCCU202101726

	/a 部文/d 。 仍/d 留/v 觀風整俗/d 。 使/v 之/d 任/v 仰/d 望闕/v 廷愈/v 增/d 孺/a 慕/d 。 惟/a 有/u 勉/v 竭/u 蟻。 誠/n 宣/v 揚/d 聖/v 化/c 。 以/d 副/d 我/v 皇/a 上/n 牖/d 迪/v 斯/v 民/n 至/n 意/v 上報/v 高/c 厚/d 於/d 萬一/v 。 為/a 此/n 繕/n 摺/v 恭/v 謝/d 天/p 恩。 伏/n 乞/d 皇/p 上睿/n 鑒/r 。 謹奏/t 。 雍正/a 玖/d 年/v 拾月/t 拾/v 柒/n 日。

Table 28: The memorial in training data

Original	奏內閣學士兼禮部侍郎奴才吳金謹奏為恭謝天恩仰祈睿鑒事本年正月拾陸日奴才接閱家信內閣學士德齡陞任員缺吏部開列額外學士阿克敦等職名欽蒙特旨將奴才實授內閣學士兼禮部侍郎等因奴才隨望闕叩頭恭謝天恩外伏念奴才於雍正肆年由兵部員外巡察黑龍江復命之日即蒙皇上格外栽培陞授戶部郎中交與怡賢親王試用奴才隨從貳年蒙怡賢親王朝夕訓導稍知趨向更荷殊恩超擢額外學士揣分已覺難安上年伍月內謝恩復蒙召入溫綸獎誨賜以上方珍硯奴才自念寒微隨祖父任生長南省後隨父入都迄今甫拾載有餘毫無知識遽陟崇班屢覲天顏備極榮寵清夜撫心銘感無極茲復蒙特旨實授學士奴才現在統領黑龍江兵馬遵旨於肆月青草發生之時帶領前進自當竭盡駑力奮勉前驅以期仰答深恩於萬一耳所有感激微忱理合繕摺奏謝奴才更有請者奴才雖係文職現今統兵在外可否邀恩賞給翎子以壯軍容伏侯睿裁為此謹奏雍正拾壹年正月貳拾貳日
Prediction	奏/ow 內閣學士/np 兼/v 禮部/na 侍郎/np 奴才/r 吳金/nr 謹奏/ow 。 為/p 恭謝/v 天恩/n 仰祈/v 睿鑒/v 事/n 。 本年/t 正月/t 拾陸日/t 奴才/r 接閱/v 家/a 信/n 。 內閣學士/np 德齡/nr 陞任/v 員缺/n 。 吏部/na 開列/v 額外學士/np 阿克敦/nr 等/a 職名/n 。 欽蒙/v 特旨/n 將/v 奴才/r 實授/v 內閣學士/np 兼/v 禮部/na 侍郎/np 等/n 因/n 。 奴才/r 隨/d 望/v 闕/n 叩頭/v 恭謝/v 天恩/n 外/n 。 伏念/mw 奴才/r 於/p 雍正/t 肆年/t 由/p 兵部/na 員外/np 巡察/v 黑龍江/ns 復命/n 。 之/u 日/t 。 即/d 蒙/v 皇上/np 格外/d 栽培/v 。 陞授/v 戶部/na 郎中/np 。 交與/v 怡賢/c 親王/nk 試用/v 。 奴才/r 隨從/v 貳年/t 蒙/v 怡賢/nt 親王/c 朝夕/t 訓導/v 。 稍/d 知/v 趨向/n 。 更/d 荷/v 殊恩/n 。 超擢/v 額外學士/np 。 揣/v 分/n 已/d 覺/v 難/d 安/v 。 上年/t 伍月/t 內/n 謝恩/v 。 復/d 蒙/v 召入/v 。 溫綸/n 獎誨/v 。 賜/v 以/u 上方/na 珍硯/n 。 奴才/r 自/r 念/v 寒微/n 。 隨/v 祖父/n 任/d 生長/v 南省/ns 。 後/d 隨/v 父/n 入/v 都/n 迄今/t 。 甫/d 拾載/t 有/v 餘/n 。 毫/d 無/v 知識/n 。 遽/d 陟/v 崇/a 班/n 。 屢/d 覲/v 天顏/n 。 備/d 極/d 榮寵/n 。 清夜/n 撫/v 心/n 。 銘感/v 無/v

DOI:10.6814/NCCU202101726

極/n 。 茲/u 復/d 蒙/v 特旨/n 實授/v 學士/np 。 奴才/r 現在/t 統領/v 黑龍江/ns 兵馬/n 。 遵旨/v 於/p 肆月/t 青草/n 發生/v 之/u 時/t 。 帶領/v 前進/v 。 自/d 當/d 竭盡/v 駑力/n 。 奮勉/d 前驅/v 。 以/c 期/v 仰答/v 深/a 恩/n 於/u 萬一/m 耳/y 。 所有/a 感激/v 微忱/n 理/n 合/d 繕/v 摺/n 奏謝/ow 。 奴才/r 更/d 有/v 請/v 者/r 。 奴才/r 雖/c 係/v 文職/n 。 現今/t 統兵/v 在/p 外/n 。 可否/d 邀恩/v 賞給/v 翎子/n 以/c 壯/v 軍容/n 。 伏侯/mw 睿裁/v 。 為/p 此/r 謹奏/ow 。 雍正/t 拾壹年/t 正月/t 貳拾貳日/t 。

Table 29: The interview questions and feedback from historians

Interview questions	Feedback from experts
Q1: These are the prediction tagging results of two memorials. What do you think about sentence segmentation, word segmentation, and POS+NER tags?	A1: The current sentence segmentation is good, and the POS+NER is not present as much of a problem; however, there are a lot of punctuation types in Chinese, which punctuations should be used for which sentence segmentation may also help to identify the specific meaning of individual words, especially in different contexts.
Q2: About the prediction tagging result of the precious two memorials, which one is better? Or do the two look similar?	A2: Yes, these two look similar.
Q3: As the current sentence segmentation results are concerned, can it achieve the effect of assisting scholars in reading memorials when there are only words and these predicted periods?	A3: Yes, the sentence segmentation prediction is helpful, especially for the beginners.
Q4: Can the current word segmentation and POS+NER tagging also assist scholars in identifying the meaning of words?	A4: Through the POS+NER, the possibility of misreading the meaning of the word can be reduced.
Q5: Is there anything else that can be improved for this study?	A5: After solving these segmentation problems, historians may still need to know, for example: Liu Shishu wrote this memorial to thank the emperor for his favor as usual (is that the convention for Qing's special officials?), or to find the opportunity

DOI:10.6814/NCCU202101726

to flatter the emperor. And Wu Jin asked the emperor for something because he was the new official, had to establish credibility , or just wanted to show off? Finally, we may still be curious, what is Yongzheng's reaction? These two memorials only record matters of one or two days. After reading a lot, the story of the Yongzheng period will become three-dimensional.

Going back to the original question, it is good to segment the sentence and confirm the POS+NER at present. But punctuations should be used, involving the identification of specific word meanings, that might be the next step.

According to the feedback from experts, firstly, the current sentence segmentation in the predictions is good, and the tags of POS+NER are acceptable. Second, the tagging results of trained memorials and untrained memorials have not much difference. Third, the prediction sentence segmentation is helpful in reading memorials; the prediction of word segmentation and POS+NER can help to reduce lexical ambiguity. Furthermore, the punctuations for sentence segmentation are important for judging the meaning of words, so defining exact punctuations is the next step in the future. Finally, two pieces of information historians want to know are the motivation of the ministers who wrote the memorials, and the action that emperor took.

DOI:10.6814/NCCU202101726

5 CONCLUSION

This study proposes two methods of model structure (method STL and method MTL) and a new memorial tagging scheme of POS+NER, observes the impact of residual connection on the model, and builds the Qing's memorial dataset with different granularity and tagging schemes.

As the experiment result, first, method MTL, which uses the BERT-BiGRU-CRF with MTL model structure and tagging scheme memorial, performs better on sentence and word segmentation tasks than method STL, which uses the BERT-BiGRU-CRF with STL model structure and tagging scheme memorial. And two methods perform nonsignificant difference on the POS+NER task. Thus, this study adopts method MTL to implement all of three tasks with Qing's dataset. Second, the residual connection is proven to improve performance of method MTL on POS+NER task. Third, method MTL of this study has higher f1-scores of three tasks than other baseline models, and also has better performance on the word segmentation task than other existing Chinese NLP tools. It proved that existing tools for modern Chinese can't get good results when dealing with ancient Chinese, using the specific model while processing memorials is necessary. Furthermore, for the demand for memorials analysis and rationality, the scheme memorial and fine granularity are selected to build the dataset. Finally, according to the feedback from experts, current sentence segmentation in the predictions of this study is good, and the tags of POS+NER are acceptable. The result of sentence segmentation, word segmentation, and POS+NER task can help scholars to read memorials and define the meaning of words.

The main contributions of this study include: (1) build Qing's memorial dataset, (2) propose a new POS+NER tagging scheme for memorials, (3) build the classical Chinese NLP models for Qing's dataset, the prediction of this can help scholars to read memorials easily, reduce the probability of misinterpretation of word meaning, and (4) certain the residual connection can be effectively improved the POS+NER task performance of method MTL with Qing's dataset.

In future work, the research attempt is using different sentence segmentation tagging schemes to let the model identify the exact punctuations of the sentence segmentation, because the punctuations are important for judging the meaning of words. Besides, the AnchiBERT, which is the pre-trained model for ancient Chinese, has been proposed in 2020. Using it as the embedding layer might be able to achieve better performance. Moreover, after word segmentation and POS+NER, which are the preparation for development of the full-text

DOI:10.6814/NCCU202101726

search application, the full-text search system of memorials are establishable. Last but not least, the information historians want to know is not only the time of the memorial, people who were mentioned, but also the motivation of the ministers who wrote the memorials, and the action that the emperor took. It is very practical to further research the model that can be used for information extraction.

DOI:10.6814/NCCU202101726

REFERANCE

[1] 莊吉發 (1983)。故宮檔案述要。國立故宮博物院。

[2] 袁晖、管锡华、岳方遂 (2002)。汉语标点符号流变史。湖北教育出版社。

[3] 黃宇暘、郭鎮武、周維強、林國平、蔡瑞煌 (2021)。人工智慧在中文歷史文獻判讀領域應用初探：以國立故宮博物院典藏為例。科技博物，**25**(3)，5-26。

[4] Cai, D., & Zhao, H. (2016). Neural word segmentation learning for Chinese. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics , 1*, 409–420. https://aclanthology.org/P16-1039/

[5] Caruana, R. (1997). Multitask learning. *Machine learning*, *28*(1), 41-75. https://doi.org/10.1023/A:1007379606734

[6] Chang, C. H., & Chen, C. D. (1993). HMM-based part-of-speech tagging for Chinese corpora. *Very Large Corpora: Academic and Industrial Perspectives*. https://aclanthology.org/W93-0305

[7] Chen X., Li B., Feng M., Xu C., Xu R., Shi M., Yu L., Xiao L., & Wang Q. (2017). *Ancient Chinese Corpus LDC2017T14*. Philadelphia: Linguistic Data Consortium. https://doi.org/10.35111/ctjv-ez04

[8] Chen, J., Qiu, X., Liu, P., & Huang, X. (2018). Meta multi-task learning for sequence modeling. *Proceedings of the AAAI Conference on Artificial Intelligenc, 32*(1). https://ojs.aaai.org/index.php/AAAI/article/view/12007

[9] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation*. Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). https://arxiv.org/abs/1406.1078

[10] Dai, Z., Wang, X., Ni, P., Li, Y., Li, G., & Bai, X. (2019). Named entity recognition using bert bilstm crf for chinese electronic health records. *2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei)*, 1-5. IEEE.

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1,* 4171–4186. https://aclanthology.org/N19-1423

DOI:10.6814/NCCU202101726

[12] Gong, L., Zhang, Z., & Chen, S. (2020). Clinical Named Entity Recognition from Chinese Electronic Medical Records Based on Deep Learning Pretraining. *Journal of Healthcare Engineering, 2020*. https://doi.org/10.1155/2020/8829219

[13] Gu, C., Wu, M., & Zhang, C. (2017). *Chinese sentence classification based on convolutional neural network*. 2017 International Conference on Artificial Intelligence Applications and Technologies (AIAAT 2017), Hawaii, USA. https://iopscience.iop.org/article/10.1088/1757-899X/261/1/012008

[14] Han, X., Wang, H., Zhang, S., Fu, Q., & Liu, J. (2019). Sentence segmentation for classical Chinese based on LSTM with radical embedding. *The Journal of China Universities of Posts and Telecommunications, 26*(02). doi: 10.19682/j.cnki.1005-8885.2019.1001

[15] Huang, H. H., Sun, C. T., & Chen, H. H. (2010). *Classical Chinese sentence segmentation*. CIPS-SIGHAN joint conference on Chinese language processing. https://aclanthology.org/W10-4103/

[16] Huang, S., & Wu, J. (2018). A pragmatic approach for classical Chinese word segmentation. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. https://aclanthology.org/L18-1186

[17] Jiao, Z., Sun, S., & Sun, K. (2018). Chinese lexical analysis with deep Bi-GRU-CRF network. *arXiv preprint*. https://arxiv.org/abs/1807.01882

[18] Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi-task learning. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, 2873–2879. https://arxiv.org/abs/1605.05101

[19] Ma, J., Ganchev, K., & Weiss, D. (2018). State-of-the-art Chinese word segmentation with Bi-LSTMs. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4902–4908. https://aclanthology.org/D18-1529/

[20] Ng, H. T., & Low, J. K. (2004). Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based?. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 277-284. https://aclanthology.org/W04-3236

[21] Norman, J., & Jerry, N. (1988). *Chinese*. Cambridge University Press.

[22] Panchendrarajan, R., & Amaresan, A. (2018). Bidirectional LSTM-CRF for Named Entity Recognition. *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*. https://aclanthology.org/Y18-1061

DOI:10.6814/NCCU202101726

[23] Qin, Q., Zhao, S., & Liu, C. (2021). A BERT-BiGRU-CRF Model for Entity Recognition of Chinese Electronic Medical Records. *Complexity, 2021*. https://doi.org/10.1155/2021/6631837

[24] Ratnaparkhi, A. (1996). A maximum entropy model for part-of-speech tagging. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 133-142. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.5102

[25] Shao, Y., Hardmeier, C., Tiedemann, J., & Nivre, J. (2017). Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF. *Proceedings of the Eighth International Joint Conference on Natural Language Processing, 1*, 173–183. https://aclanthology.org/I17-1018

[26] Shi, M., Li, B., & Chen, X. (2010). CRF based research on a unified approach to word segmentation and POS tagging for Pre-Qin Chinese. *Journal of Chinese Information Processing, 2*(24), 39-46. http://jcip.cipsc.org.cn/CN/Y2010/V24/I2/39

[27] Tian, Y., Song, Y., Ao, X., Xia, F., Quan, X., Zhang, T., & Wang, Y. (2020). Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8286-8296. https://aclanthology.org/2020.acl-main.735/

[28] Tian, Y., Song, Y., & Xia, F. (2020). Joint Chinese Word Segmentation and Part-of-speech Tagging via Multi-channel Attention of Character N-grams. *Proceedings of the 28th International Conference on Computational Linguistics*, 2073-2084. https://aclanthology.org/2020.coling-main.187/

[29] Wang, Q., & Zeng, L. (2018). Chinese symptom component recognition via bidirectional LSTM-CRF. *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)* , 45-50. IEEE. doi: 10.1109/ICACI.2018.8377564.

[30] Wilkinson, E. P. (2000). *Chinese history: a manual*. Harvard Univ Asia Center.

[31] Wu, Y., Jiang, M., Lei, J., & Xu, H. (2015). Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network. *Studies in health technology and informatics, 216*, 624-628.

[32] Zhang, H. P., Yu, H. K., Xiong, D., & Liu, Q. (2003). HHMM-based Chinese lexical analyzer ICTCLAS. *Proceedings of the second SIGHAN workshop on Chinese language processing*, *17*, 184-187. https://doi.org/10.3115/1119250.1119280

[33] Zhang, H. P., Liu, Q., Yu, H. K., Cheng, X., & Bai, S. (2003). Chinese named entity recognition using role model. *International Journal of Computational Linguistics & SChinese Language Processing, 8*(2)*,* 29-60. https://aclanthology.org/O03-5002

DOI:10.6814/NCCU202101726

APPENDIX

Chinese Version of Interview and Feedback

Interview questions	Feedback from experts
Q1: 這是兩篇奏摺的原文以及預測結果，希望您可以從一位歷史學者之觀點，評論此兩篇奏摺之斷詞斷句效果之優劣處？	A1: 目前詞性的判斷問題不大，斷句也很好；不過，斷句該作「，」、「；」、「。」或「？」，可能還涉及了對個別的詞的具體意義，特別是不同脈絡中的意義的掌握。
Q2: 關於前面兩篇奏摺標記的結果，有哪篇是標註得比較好的嗎？還是兩篇看起來差不多呢？	A2: 是的，應當是差不多的。
Q3: 就目前的斷句結果來說，在只有文字和句號的情形下，可以達到輔助學者們閱讀奏摺的效果嗎？	A3: 斷句是有幫助的，特別是對初學者來說。
Q4: 另外，目前斷詞以及詞性的標註是不是也可以協助學者辨認詞義呢？	A4: 透過詞性的判斷，可以減少對詞義誤讀的可能。
Q5: 關於這個研究還有其他改進之處嗎？	A5: 解決了這些斷句問題後，歷史學者可能還需要知道：劉師恕寫這篇摺子謝恩，是因為照例（是不是清的特遣官員都這樣？）要表達的禮貌，還是找機會拍皇上馬屁。吳金是因為在黑龍江守邊，為了新官上任、文職統兵要服眾，才上摺子跟皇帝要東西，還是要拿皇帝的特賜炫耀。最後，我們可能還好奇，雍正的反應是什麼？從提供的原件看，劉師恕大概是拍馬屁，雍正一眼就看穿了。吳金的建議，雍正則是以過去沒那個例，回絕了所請（言下之意，就是要吳金自己看著辦，不要什麼都要皇帝罩著才行）。但這兩個摺子只是一兩天中的事，要大量看後，雍正時的故事才會慢慢立體起來。回到原來的問題，目前把句子斷開、詞性確認是很好的。但該斷成哪個符號，涉及具體詞義的辨認，那可能是下一步工作？

DOI:10.6814/NCCU202101726