ISLab at the NTCIR-17 QA Lab-PoliInfo-4: Models for Automatically Identifying Politicians' Stances on Bills

Guan-Yu Chen
Department of Computer Science and Information Engineering
National Kaohsiung University of Science and Technology
Kaohsiung, Taiwan, R.O.C
C109151123@nkust.edu.tw

Yu-Cheng Liu
Department of Computer Science and Information Engineering
National Kaohsiung University of Science and Technology
Kaohsiung, Taiwan, R.O.C
F110151106@nkust.edu.tw

Tao-Hsing Chang
Department of Computer Science and Information Engineering
National Kaohsiung University of Science and Technology
Kaohsiung, Taiwan, R.O.C
changth@nkust.edu.tw

Fu-Yuan Hsu
Research Center for Psychological and Educational Testing
National Taiwan Normal University
Taipei, Taiwan, R.O.C
kevin@rcpet.ntnu.edu.tw

ABSTRACT
This paper aims to design a model that can determine whether the politician's stance is approved or disapproved the bill based on the politician's utterance on a specific bill in the parliament. This study proposed two frameworks for determining the stance in utterances. The first framework involves concatenating BERT model with Bi-LSTM model to form a comprehensive decision-making model while the second framework is concatenating Curie model with ChatGPT model. This paper used the dataset provided by Stance Classification 2 task in NTCIR-17 for model training and testing, and GPT-based model this paper proposed achieved an accuracy of 0.932.

KEYWORDS
BERT, Bi-LSTM, GPT-3, ChatGPT, Curie, Stance Classification.

TEAM NAME
ISLab

SUBTASKS
Stance Classification-2 (Japanese)

1 INTRODUCTION
This study, as a subtask of the QA Lab-PoliInfo-4 [1] task of the NTCIR-17 conference under the “Stance Classification 2” subtask, aims to design a model that can determine whether the politician's stance is approved or disapproved the bill based on the politician's utterance on a specific bill in the parliament. This subtask is similar to the subtask in NTCIR-15 [2], but the types of processing texts subtask are different; Figure 1 illustrates the difference between these subtasks. Figure 1(a) shows the original text of an utterance, which is the form of the text processed in NTCIR-15. This utterance focuses on a specific politician's opinion on Proposal 57. The politician believes that amendments to the bill will increase people's medical expenses before eventually becoming unaffordable and hence is disapproving the proposal.

Figure 1(b) shows a new text formed by masking words related to the politician's stance in Figure 1(a) with a [STANCE] marker. As shown in Figure 1(a), when the word “反対(disapproval)” appears in rows 2 and 19, as marked in red, it is replaced with [STANCE], which generates the text in Figure 1(b), i.e., the text processed in the subtask “Stance Classification 2.” As explicit stance-indicative words are replaced with the same marker, the stance-determining model must rely on other information within the text to determine the conveyed stance of the given utterance.

We proposed two model frameworks for determining the stance in utterances. The first framework involves concatenating a semantic space model with a sequential data processing model to form a comprehensive decision-making model (hereafter referred to as the bidirectional encoder representations from transformers (BERT) [3]–long short-term memory (LSTM) [4] model). Our idea is to first transform the text through a language space model into a semantic space coordinate representing the text, anticipating that the text of approved and disapproved stances would form separate clusters in the semantic coordinate space. Therefore, by training a simple classifier to differentiate stances, we can convert the utterance to be judged into a semantic coordinate and then use the classifier to determine the stance. This framework has been widely applied to various classification issues with considerable success.
The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.

The second framework is concatenating GPT-3's [5] two sub-models, i.e., Curie[6], with the summarization model of ChatGPT[7] (hereafter referred to as the GPT-based model). Large language models (LLMs), also known as foundation models, can be fine-tuned to meet the requirements of downstream tasks. We treated the "Stance Classification 2" subtask as a question-answering (QA) issue. In other words, we provided the model with an utterance and asked it to determine the stance of the utterance. Therefore, we used ChatGPT to summarize the utterance before fine-tuning the QA model Curie with the summarized text, thus enabling the model to determine the stance of the given text.
feasible to establish a stance detection model with rule-based methods that require a high level of proficiency in the Japanese language to analyze the linguistic rules. Therefore, this paper primarily adopts methods based on deep-learning neural networks to design the stance detection model.

3 Methodology
The motivations behind the design of the two proposed frameworks, i.e., the BERT–LSTM and GPT-based models, are explained in Section 1. Sections 3.1 and 3.2 elaborate on the details of these two methods, respectively.

3.1 BERT-LSTM Model
This model concatenates BERT and bidirectional LSTM (Bi-LSTM) [14] models. We believe that utterances can be transformed into semantic vectors: texts with the same stance should be closer in the vector space and those with opposing stances should be farther apart. Therefore, we can use a classifier to divide the semantic vectors of utterances into two groups: "賛成" (approval) and "反対" (disapproval). Based on this concept, this paper employed the BERT model to transform text into semantic vectors. However, as BERT can only process a limited number of words in a single text, we broke longer texts that exceeded BERT’s limits into several segments, with each segment entered into BERT to generate its respective semantic vector and then integrated using Bi-LSTM.

Bidirectional Encoder Representations from Transformers (BERT) is a language model comprising a multilayer encoder based on the transformer [15] architecture. The BERT model is pre-trained with copious unlabeled text and employs attention mechanisms that consider relationships between all words in the entire text at once; this mechanism allows BERT to precisely transform text and words into vector coordinates in the language space. BERT’s pre-training procedure employs two methods. The first approach is the Masked Language Model (MLM) method, which operates based on the principle of requiring the model to predict certain masked words during training. By predicting the masked words, the model can be adjusted to make more accurate predictions in the future, leading to more precise settings of hyperparameters affecting relationships between these words. The second pre-training method, known as the Next Sentence Prediction (NSP) method, requires the model to predict whether two input sentences are consecutive. The model is adjusted based on the prediction outcomes, which helps in setting hyperparameters affecting the semantic relationship between sentences more accurately.

The Long Short-Term Memory (LSTM) model is based on the recurrent neural network (RNN) architecture and addresses the vanishing and exploding gradient problems in RNNs. With the designed gate and cell memory mechanisms, LSTM can determine whether to retain previous computation results in the next step, thus allowing LSTM to effectively handle data with time sequence characteristics. As LSTM processes data sequentially, information processed earlier does not consider information later in the sequence during computation, which does not meet the requirement of this paper of considering the text as a whole when processing the information. Therefore, to fulfill the need for considering the context of the information, this paper employs Bi-LSTM, which processes the full sequence of text at once.

![Figure 2: Architecture of BERT-LSTM.](image)

Figure 2 illustrates the architecture of the BERT–LSTM model. Firstly, the model divided the utterances into several segments based on the maximum input length allowed by the BERT model. These segments were entered into BERT in a sequential order to obtain the semantic vectors for each segment. These semantic vectors were then entered sequentially into Bi-LSTM to generate the final vector output from the LSTM as the semantic vector of the utterance. This vector was then fed into the full connection classifier to determine the stance of the utterance.

3.2 GPT-based Model
This model concatenates summarization and QA models. As we considered the stance determination of utterances a QA problem, this paper attempted to employ the Curie QA model to assess the stance of utterance. Additionally, similar to BERT, the Curie model has limitations on the number of words it can process in a single text. As a result, this model divided longer texts over the QA model's limits into several segments; each segment underwent a generative summarization process through ChatGPT to shorten the text length in the segment without altering the meaning of the original content. Subsequently, the summarized texts of each segment were reassembled in sequential order to form a single summarized text that met Curie's length restrictions. This summarized text was then fed to Curie to determine the stance of the original utterance.

Curie and ChatGPT are constructed primarily based on the technology of GPT-3. GPT [16] was introduced to address the
GPT involves pre-training a base model with a large amount of unlabeled text and then fine-tuning the base model on a small set of labeled text specific to the task. Like BERT, GPT is a transformer-based model. However, GPT uses the decoder part of the transformer architecture, whereas BERT focuses on the encoder part. Figure 3 illustrates the architecture of GPT.

GPT-2 [17] is an improved version of GPT. GPT-2 involves moving the layer normalization module to the input of feed-forward and multi-self-attention networks; this layer normalization module is added after the last transformer block. While GPT-3 and GPT-2 possess the same architecture, GPT-3 alternates between dense and locally banded sparse attention models in transformer blocks. In the traditional transformer, during self-attention calculations, every two vectors need to compute attention with each other. Meanwhile, when alternating between dense and locally banded sparse attention models, the dense attention model calculates vectors at fixed intervals, whereas the locally banded sparse attention calculates a few vectors adjacent to the current position. This modification reduces the computational complexity of the transformer, thus enabling the efficient handling of longer texts.

In the second framework proposed in this paper, we employed ChatGPT as the summarization model. ChatGPT is a LLM built from GPT-3.5, which is GPT-3 fine-tuned with reinforcement learning from human feedback (RLHF) model training. RLHF, used by ChatGPT, consists of three steps. The first step is to record various human responses and use them as training data for a GPT-3.5 model in a supervised learning approach. In the second step, a set of questions is answered using various machine models, with the quality of the answers ranked manually; the sorted data are then used to train a reward model. The third step uses proximal policy optimization to fine-tune the hyperparameters of the model trained in the first step. Responses from the fine-tuned model are then scored by the reward model from the second step. The final scores are used to further refine the model; this process is iterated multiple times.

In the second framework proposed in this paper, the QA model employs Curie, a GPT-3 sub-model chosen from among the various GPT-3 sub-models provided by OpenAI owing to its faster computation speed and better performance in sentiment analysis tasks compared to other sub-models.
This model added prompts, which this paper designed, to each segment and then fed it to ChatGPT for summarization. To avoid extremely short summaries, which could result in significant semantic differences between the summarized segments, we used the following prompt:

"この記事を絶対に n 字以内で日本語でまとめる。"
(Summarize this text in Japanese with an absolute maximum of n characters.)

Where \( n \) was calculated using the following formula:

\[
 n = \text{floor}(t - a/b),
\]

where \( t \) denotes the total number of words in the utterance, \( a \) denotes the total number of words in sentences containing the [STANCE] marker within the utterance, and \( b \) denotes the number of segments.

Finally, the summaries of each segment and the sentences containing the [STANCE] marker were reorganized in sequence to create the summary text of the original utterance. The summary text was subsequently used to fine-tune Curie. Since the input of Curie requires a prompt in addition to the summary text, the prompt this paper created for Curie is as follows:

"記事中の[STANCE]は[賛成]とすべきか[反対]とすべきか?"
(Should the [STANCE] in the text be [approved] or [disapproved]?)

The two prompts mentioned above are selected after testing many different prompts. In Section 4, we provide further examples of how to select the prompts of ChatGPT and Curie from various prompts.

4 EXPERIMENTS

This paper used the dataset provided by Stance Classification 2 for training and testing. The dataset included 8,534 utterances in the training set and 2,240 utterances in the testing set, with each utterance manually labeled with a stance (approved or disapproved). Owing to an error in an item of the training dataset and the fact that this paper extracted the first 1,000 utterances out of 8,533 as the validation dataset, the actual number of utterances used for training was 7,533; apart from the aforementioned dataset, no additional datasets were used in this study. The utterances in the testing set did not contain manually labeled stances. The results of the proposed models on the testing set were uploaded on the NTCIR-17 website, where the website automatically calculates the accuracy of the proposed model.

In the GPT-based model, two crucial parameters, i.e., feeding prompts to ChatGPT and Curie, were considered. This paper experimented with the following three prompts for ChatGPT:

1. "日本語で n ワード以内で記事に要約する";
(Summarize the text content in Japanese with no more than n words.)

We randomly selected 100 utterances from the train dataset that need to be summarized by ChatGPT as the validation data. Then, we use the 9 combinations of the prompts mentioned above as the inputs of the proposed model to calculate the prediction accuracy of the model on the validation data. Table 1 lists the predicting accurate rates of the proposed model under various combinations. Table 1 shows that the prediction accuracy of the proposed model is higher when Curie uses prompt 3 as input. Moreover, the highest accuracy is obtained when the proposed model uses ChatGPT's prompt 3 and Curie's prompt 3. Therefore, the two prompts 3 (i.e. the prompts mentioned in Subsection 3.2) are adopted by the proposed model.

<table>
<thead>
<tr>
<th>Selected Prompts</th>
<th>ChatGPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prompt1</td>
<td>0.73</td>
</tr>
<tr>
<td>Prompt2</td>
<td>0.81</td>
</tr>
<tr>
<td>Prompt3</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 2 lists the stance determination results of the two proposed models on the testing set. As the pilot experiments suggested no significant differences in accuracy among different prompts, GPT-based model only employed the prompts mentioned in Section 3.2. Table 2 indicates that the GPT-based model outperformed the BERT-BiLSTM model in terms of accuracy.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT-BiLSTM</td>
<td>0.916</td>
</tr>
<tr>
<td>GPT-based</td>
<td>0.932</td>
</tr>
</tbody>
</table>
According to the results in Table 3, we believe that the text summarized by ChatGPT may not fully capture the semantics in the original content; moreover, the semantics discarded in the summarization process could be crucial for Curie’s stance determination. Based on this assumption, this paper analyzed the number of misjudgments for texts summarized and not summarized by ChatGPT, respectively. Table 3 shows the results of analysis mentioned above. Table 3 shows that the model error rate was 0.045 in utterances without summaries and 0.114 in utterances with summaries (a 2.53 times higher error rate). This significant difference demonstrates that the summarization step decreases the accuracy of stance determination.

<table>
<thead>
<tr>
<th>Total</th>
<th>True</th>
<th>False</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utterances</td>
<td>2240</td>
<td>2089</td>
<td>151</td>
</tr>
<tr>
<td>Utterances do not require summarization</td>
<td>1533</td>
<td>1463</td>
<td>70</td>
</tr>
<tr>
<td>Utterances require summarization</td>
<td>707</td>
<td>626</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure 5 is an example that illustrates the model’s incorrect prediction for an utterance that has been summarized by the GPT. This example is the ChatGPT summaries of the utterance with id 113247-00023 in the training dataset. In contrast to the original utterance, ChatGPT cannot retain the critical content in the utterance that can be used to predict the stance of the utterance, and even experts cannot also identify the stance of the utterance from the summary. This may be the reason why the prediction accuracy of the proposed model decreases for summarized utterances.

5 CONCLUSIONS

This paper demonstrated our approach to tasks related to stance classification. We designed two frameworks for determining politicians’ stances using utterances, with the GPT-based model achieving an accuracy of 0.932. The findings indicate that text processed through summarization still cannot fully replace the original text; moreover, summarization affects the accuracy of the GPT-based model. For text that expresses meanings subtly or indirectly, language models have difficulty comprehending the true intent of the text, leading to the false determination of opposite stances.

Therefore, several works for enhancing the performance of the proposed frameworks could be developed in the future. First, we aim to experiment with other QA models as the judgment model. As Curie is not an open-source model, our training frequency and method were considerably limited. With other new open-source QA models based on different LLMs available, we can explore the performance of judgment models based on alternative QA models. Second, ChatGPT—adopted for generating summaries in this study—was not fine-tuned concurrently in the training process of the judgment model. However, the error analysis revealed a significantly higher error rate in the summarized text compared to that in the un-summarized text. Hence, fine-tuning may potentially improve the performance of the summary generation model and consequently enhance the accuracy of the stance determination model. Third, some ways of expressing opinions in utterances are less common. For instance, because of the serious problems that would arise if the bill was passed, the politician was disapproved of the bill. While these expressions are not common, they are not rare and still constitute a certain proportion of the dataset. Therefore, enabling the model to recognize such utterances may help improve the judgment accuracy of the model.

Acknowledgment

This study was partially supported by the Ministry of Science and Technology, under the grant MOST 110-2511-H-992-003-MY3. It was also financially supported by the "Institute for Research Excellence in Learning Sciences" of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

REFERENCES


