

Article

Machine Learning-Based Automatic Utterance Collection Model for Language Development Screening of Children

Jeong-Myeong Choi ^{1,2}, Yoon-Kyoung Lee ³, Jong-Dae Kim ^{1,2}, Chan-Young Park ^{1,2} and Yu-Seop Kim ^{1,2,*} 

¹ Department of Convergence Software, Hallym University, Chuncheon 24252, Gangwon-do, Korea; jeong5905@gmail.com (J.-M.C.); kimjd@hallym.ac.kr (J.-D.K.); cypark@hallym.ac.kr (C.-Y.P.)

² Bio-IT Center, Hallym University, Chuncheon 24252, Gangwon-do, Korea

³ Division of Speech Pathology and Audiology, Hallym University, Chuncheon 24252, Gangwon-do, Korea; ylee@hallym.ac.kr

* Correspondence: yskim01@hallym.ac.kr

Abstract: To assess a child's language development, utterance data are required. The approach of recording and transcribing the conversation between the expert and the child is mostly utilized to obtain utterance data. Because data are obtained through one-on-one interactions, this approach is costly. In addition, depending on the expert, subjective dialogue situations may be incorporated. To acquire speech data, we present a machine learning-based phrase generating model. It has the benefit of being able to cope with several children, which reduces costs and allows for the collection of objectified utterance data through consistent conversation settings. Children's utterances are initially categorized as topic maintenance or topic change, with rule-based replies based on scenarios being formed in the instance of a topic change. When it comes to topic maintenance, it encourages the child to say more by answering with imitative phrases. The strategy we suggest has the potential to reduce the cost of collecting data for evaluating children's language development while maintaining data collection impartiality.

Keywords: utterance data collection; machine learning



Citation: Choi, J.-M.; Lee, Y.-K.; Kim, J.-D.; Park, C.-Y.; Kim, Y.-S. Machine Learning-Based Automatic Utterance Collection Model for Language Development Screening of Children. *Appl. Sci.* **2022**, *12*, 4747. <https://doi.org/10.3390/app12094747>

Academic Editors: Wen-Hsiang Hsieh, Jia-Shing Sheu and Minvydas Ragulskis

Received: 10 April 2022

Accepted: 7 May 2022

Published: 9 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Developmental disabilities affect around 5 to 10% of the overall pediatric population. The majority of a child's development occurs in the areas of movement, language, cognition, and sociality. If there is a delay in language development in infancy, it can result in not just specific language impairment, but also developmental disabilities such as Mental Retardation (MR) and Autism Spectrum Disorder (ASD) [1]. Therefore, it is critical to recognize and cure problems early on by monitoring the degree of language development.

There are two types of tests used to assess children's language development: direct and indirect assessments [2]. The direct test involves the examiner actually examining the child and recording and evaluating the results, while the indirect test involves parental reporting or behavioral observation. However, standardized test tools that measure structural situations are limited when it comes to gauging children's real language skills. As a result, recent, spontaneous speech analysis, which can detect children's use of language in natural contexts, has gained popularity [3].

Initially, it is important to acquire a number of conversation samples for spontaneous speech analysis. Conversation samples are collected through a predefined conversation collection protocol. The expert conducts a conversation with the child using a conversation collection protocol and records the entire process. Then, the recordings are transcribed, converted into text, and then collected. The collected transcriptional data are analyzed on several metrics such as the total number of utterances, average utterance length, number of topics, and topic retention rate to evaluate the level of language development. However, because different biases may be expressed depending on the expert in the conversation,

the objectivity of conversation data may suffer. Furthermore, because conversations are one-on-one, it is difficult for an expert to obtain conversation samples from several children.

In this study, we propose a machine learning-based conversation data collection model to maintain the objectivity of data collection and to reduce costs incurred during the conversation. Similar to spontaneous speech analysis, our proposed model serves to induce children to respond faithfully to a set procedure. To do so, it has to decide whether or not to keep the topic of the current conversation going. When the topic needs to be maintained, a question-like reaction about the child's utterance induces the child's response. When the topic needs to be changed, a new topic should be presented. Therefore, a topic maintenance or change classification model and an imitative sentence generating model are constructed in this study. We assess the applicability of machine learning-based approaches, Light Gradient Boosting Machine (LGBM), Support Vector Machine (SVM), and deep learning-based approaches.

After that, the structure of this paper is as follows. Section 2 introduces studies related to Language Sample Analysis. Section 3 describes the conversation collection protocol, and Section 4 describes the dataset. Section 5 describes the methodology of collecting through the model. Section 6 describes the experiment and evaluation of the model, and Section 7 describes the conclusion of this study and future research.

2. Related Work: Language Sample Analysis

Several studies on Language Sample Analysis (LSA) are in progress. Ref. [4] analyzes Correct Information Unit (CIU) word-to-word ratios using CIU in spontaneous speech in normal elderly, mild cognitive impairment, and mild Alzheimer's patients. As a result of analyzing the number of nouns, verbs, adjectives, and adverbs per utterance in spontaneous utterances, there were significant differences between the three groups, normal elderly, mild cognitive impairment, mild Alzheimer's. Ref. [5] is utterance data collected from the daily lives of children aged 2 to 5 years, and verb types and verb frequencies were evaluated by age. The number and frequency of verb types according to age increased with age, and the frequency of verbs also increased. Ref. [6] compares the frequency of use of expression types by age using transcribed data from children aged 3 to 5 years of age. The frequency and type of expressiveness utilized by children increased with age.

There are also various studies on the analysis of speech patterns according to the types of speech collection. Ref. [7] analyzed the grammatical expression ability of first, third, and fifth graders of school-age according to the type of speech collection. There was a difference between groups according to the two types of speech collection through conversation and speech collection through pictures and stories. It was confirmed that the higher the grade, the greater the difference according to the utterance collection type. Ref. [8] evaluated the speech patterns of preschool-age children with speech impairment according to the three speech sample collection methods: conversation, free play, and story. The conversation was shown to be a method that reliably guarantees an appropriate amount of speech regardless of the language ability or communication style of each child with language impairment. The story method showed that the quality of speech length and syntax structure can include a relatively high level of speech.

3. Conversation Collection Protocol

The spontaneous analysis must be conducted under the same conditions and environment to ensure the reliability of the analysis. The rules created for this purpose are conversational protocols. In this study, the conversation collection protocol uses the conversation collection method established by the Division of Speech Pathology and Audiology of Hallym University [9]. The conversation is one-on-one and takes about 10 to 15 min. As shown in Figure 1, the conversation collection protocol proceeds in the order of introduction, procedure explanation, conversation initiation, topic maintenance after conversation initiation, topic change, and dialogue ending, and the description of each step is as follows:

- **Conversation Introduction Stage:** It is a process of greeting and introducing each other, and it is a process of talking naturally with the child. The purpose of the experiment is not revealed to the child in order to maintain the objectivity of the test.
- **Procedure Explanation Stage:** This is the stage where you explain how the conversation will proceed. The conversation procedure is as follows. First, the expert shows the prepared photographs of 3–5 sheets, and the child expresses what comes to mind after looking at the pictures. The topic of the picture includes scenes of conversations with friends, scenes of school grounds, scenes of gathering with family, and more. All pictures are in color and are the same size, A4. Each picture indicates a topic for the child to talk about. At this time, experts should inform the child that they should be talking about their experiences rather than describing the situations described in the picture.
- **Conversation Initiation Stage:** It is very important because it is the stage in which conversation data necessary for spontaneous speech analysis are collected in earnest. When a child looks at a picture and tells an experience about it, the expert must either preserve the conversation topic or transition to a different picture topic, depending on the utterance. For example, if the child says “I went on a trip with my family” the expert maintains the topic to elicit various utterances from the child. Therefore, the expert must respond as if imitating a child’s utterance so as not to change the topic the child is talking about. The expert must have a response that encourages the child to answer, such as “Did you go on a trip?” or “And then?”. If the child directly states that there is nothing to talk about the topic, or if they talk about a topic for a long time, the expert uses their judgement to change the topic. In case of changing the topic, the expert shows the rest of the pictures that the child has not selected and repeats the procedure of maintaining the topic and changing the topic while continuing the conversation. When enough utterances have been collected or the talk regarding the prepared picture has been completed, the conversation ends.

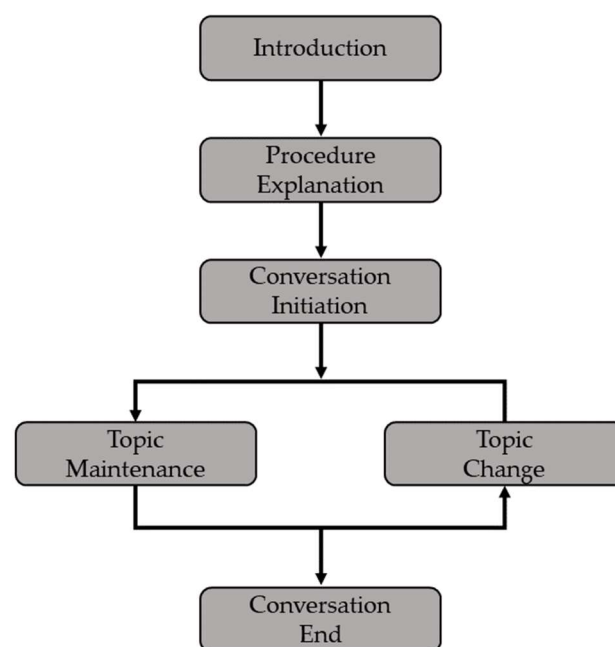


Figure 1. Conversation collection protocol process.

Among the spontaneous speech collection procedures, the introduction and procedure explanation are the parts where the expert explains the set contents. However, since the child and the expert interact and proceed with the conversation after the conversation starts, the expert must respond according to the situation. The utterance collecting model

suggested in this work intends to be able to respond with relevant utterances based on the utterances of children at the stage of conversation beginning.

4. Dataset

The raw transcription data collected by the conversation collection protocol is shown in Table 1. We collected conversations from 67 children aged 5 to 6 years. A total of 1800 utterances were chosen from among them. Of the 1800 selected utterances, 250 are utterances that should be answered by changing the topic, and the remaining 1550 are utterances that should be answered by maintaining the topic.

Table 1. Transcribed examples of data collected by the conversation collection protocol. The same utterance is expressed in Korean and English.

Speaker	Utterance (Kor)	Utterance (Eng)	Category
Expert	daehwaga mueos-inji al-ayo?	Do you know what a conversation is?	Procedure Explanation
Child	aniyo. mollayo.	No. I don't know.	Procedure Explanation
Expert	daehwaneun yeoleosalam-i gat-i iss-eul ttae seolo mal-eul jugobadneun geos-eul daehwalago haeyo. jigeum ulido daehwaleul hago issneun geoyeyo.	The conversation is about talking to each other. When we talk like this, it calls the conversation.	Procedure Explanation
Child	ne	Yes.	Procedure Explanation
Expert	yeogie issneun geulim kadeu ne gaeleul bogo daehwaleul hal geoyeyo. geulim-e daehae seolmyeonghaneun ge anila gieognaneun geoleul yaegihameon dwae. hana gollabolkkayo?	We're going to look at the four picture cards here and talk to them. You don't have to explain the picture, you just have to talk about what you remember. Would you like to pick one?	Conversation initiation (topic selection)
Child	ne. gollass-eoyo. (family picture)	Yes. I picked (family picture)	Conversation initiation (topic selection)
Expert	geuleom yaegileul haejuseyo.	Then please talk about family	Conversation initiation
Child	gajog-ilang yeohaeng-eul gass-eoyo.	I went on a trip with my family.	Conversation initiation (topic maintenance)
Expert	yeohaeng-eul gass-eoyo?	Did you go on a trip?	Conversation initiation (topic maintenance)
Child	gajog-ilang imjingag gaseo lolleokoseuteo tass-eoyo.	I went to Imjingak with my family and rode the roller coaster.	Conversation initiation (topic maintenance)
Expert	lolleokoseuteo tass-eoyo?	Did you ride the roller coaster?	Conversation initiation (topic maintenance)
Child	ne ollagassda naelyeogassda haneun geosdo tassgo	Yes, I rode it up and down	Conversation initiation (topic maintenance)
Expert	ollagassda naelyeogassda haneun geosdo tassgo geuligo?	You also rode up and down rides. And then?	Conversation initiation (topic maintenance)
mediated syncope			
Child	deo eobs-eoyo.	No more.	Conversation initiation (topic change)
Expert	ibeon-eneun jangnangam yaegileul haebolokka? joh-ahaneun jangnangam iss-eoyo?	Would you talk about toys? Do you have a favorite toy?	Conversation initiation (topic change)

Table 1. *Cont.*

Speaker	Utterance (Kor)	Utterance (Eng)	Category
Child	mimiyo.	It's MIMI	Conversation initiation (topic maintenance)
Expert	Mimiyo?	MIMI?	Conversation initiation (topic maintenance)
Child	mimi jangnangam malgo tto iss-eoyo.	There are more than MIMI toys.	Conversation initiation (topic maintenance)
Expert	tto iss-eoyo?	Do you have more?	Conversation initiation (topic maintenance)
mediated syncope			

Table 2 shows a data sample with structured data for training the model. The label column indicates whether the topic should be maintained or changed in response to the child's utterance. The label is that the expert directly listened to the recorded conversation, and the responses to what the child said were classified as maintaining the topic or changing the topic. The topic change utterances are typically obvious, such as "deo eobs-eoyo." ("No more."), and "geuman hallaeyo." ("I want to quit."), as seen in Table 2.

Table 2. Data sample with structured data for training the model.

Children Utterance	Expert Utterance	Label
gajog-ilang yeohaeng-eul gass-eoyo.	gajog-ilang <S> yeohaeng-eul gass-eoyo.	Topic maintenance
gajog-ilang imjingag gagajigo lolleokoseuteo tass-eoyo.	gajog-ilang imjingag gagajigo <S> lolleokoseuteo tass-eoyo.	Topic maintenance
ne ollagassda naelyeogassda haneun geosdo tassgo	ne <S> ollagassda naelyeogassda haneun geosdo tassgo	Topic maintenance
deo eobs-eoyo.	-	Topic change
geuman hallaeyo.	-	Topic change

In order to maintain the topic, it is necessary to respond with a questioning-like response so that the topic is maintained through the child's utterance. We discovered that segmenting some sentences of a child's utterance could produce a questioning-like response. Because Korean has a fairly free sentence structure, meaning can be conveyed even if the subject is removed or the word order is modified. Therefore, it is possible to create a natural questioning-like response even by segmenting a sentence based on the segmentation point in an utterance. We constructed the data by marking the segmentation point (<S>) in children's utterances, and it is shown in Table 2.

5. Methodology

To effectively use the model proposed in the conversation initiation stage, we devised a two-stage strategy, as shown in Figure 2. Stage 1 is the stage of classifying whether to maintain or change the topic through the model by analyzing the child's utterances. In the case of a topic change, another predetermined topic is returned by the conversation collection protocol. In case of topic maintenance, Stage 2 is executed. Stage 2 finds a segmentation point in the sentence and generates a questioning-like sentence. For example, if the utterance "I want to quit." is input in the Stage 1 model, it is classified as a topic change and the next topic is returned based on the conversation collection protocol. If an utterance "We make volcanoes and dinosaurs" is input in the Stage 1 model, it is classified as topic maintenance and the utterance is transmitted to Stage 2. In Stage 2, it generates a questioning-like sentence by finding the segmentation point and returns the sentence "Do you make dinosaurs?".

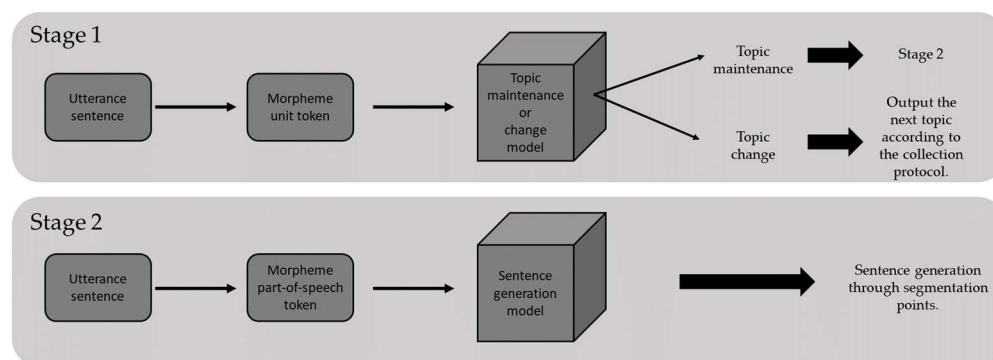


Figure 2. A flow with the two stages of the proposed method.

We compared machine learning approaches (LightGBM and SVM) and deep learning approaches to determine an appropriate analytical model when developing a topic maintenance or change model and a sentence generation model.

5.1. Pre-Processing

At each stage, the input uses tokens stemming from the utterance. In this case, we used Mecab in the KoNLPy library as a Korean morphological analyzer [10]. In Stage 1, morphological tokens are used, and in Stage 2, part-of-speech tokens in units of morphemes are used. For example, “hwasando mandeulgo gonglyongdo mandeul-eoyo.” (“I make volcanoes and dinosaurs.”) is analyzed as nine morphological tokens as (hwasan, do, mandeul, go, gonglyong, do, mandeul, woyo.). The part-of-speech token in the morphological unit can be analyzed as (NNG, JX, VV, EC, NNG, JX, VV, EF, SF) through a predefined part-of-speech tags. All morphological tokens and part-of-speech tokens are each predefined as a unique integer and encoded accordingly. To make the encoded sequence of the same length, zero padding is used.

The morphological token is the smallest unit of speech that has meaning and is used in the topic maintenance or change classification model of Stage 1 because it can effectively capture the pattern of topic maintenance and change: for example, “deo isang eobs-eoyo” (“No more”), and “geuman mal hallaeyo” (“Stop”). It can capture the appearance of words such as “eobs-eoyo” (“no”) and “geuman” (“Stop”) well. The sentence generation model of Stage 2 uses part-of-speech tokens as input because the point where the sentence should be segmented and the point where it is not can be captured well with the part-of-speech order pattern.

5.2. Machine Learning-Based Approach

There are various machine learning algorithms that can be applied to text data: Random Forest [11], LGBM, SVM, Naïve Bayes, modified Naïve Bayes [12]. In this study, we used LGBM and SVM among various machine learning-based approaches. Although LGBM and SVM may take a sequence as input, it is difficult to output it in sequence format. In the topic maintenance or change model of Stage 1, a sequence is an input and the result of binary classification is output. However, in the case of topic maintenance, it is necessary to reconstruct the data because it is difficult to output the sequence in the sentence generation model of Stage 2. For data reconstruction, we first slice tri-gram into morphemes. For example, (NNG, JX, VV, EC, NNG, JX, VV, EF, SF), there is a sequence consisting of nine part-of-speech tokens where the “EC” token is the segmentation point (<S>). For this, we use tri-gram slicing to reconstruct the data as in (<padding>, JX, VV), (NNG, JX, VV), (JX, VV, EC), (VV, EC, NNG), (EC, NNG, JX), (NNG, JX, VV), (JX, VV, EF), (VV, EF, SF), (EF, SF, <padding>). Then, we label only (VV, EC, NNG) with a value of 1, meaning the segmentation point, and label the rest with a value of 0. In this way, we grouped and separated tokens in tri-gram units and reconstructed them to replace the output of the sequence format by labeling whether each part is a segmentation point or not.

5.2.1. Light Gradient Boosting Machine (LightGBM)

LGBM is one of the models that use boosting ensemble method as a tree-based algorithm [13]. In the existing tree-based gradient boosting algorithm, the tree is level-wise pruned as shown in A in Figure 3. eXtreme Gradient Boosting (XGB) is a representative model for level-wise pruning [14]. On the other hand, LGBM is known to build a well-balanced tree that is strong against overfitting because the tree is pruned in a leaf-wise method as shown in Figure 3b. Recently, various task and studies have shown good results using LGBM [15–17]. In this study, we also tested whether LGBM is a suitable model for sentence maintenance or change classification and sentence generation tasks.

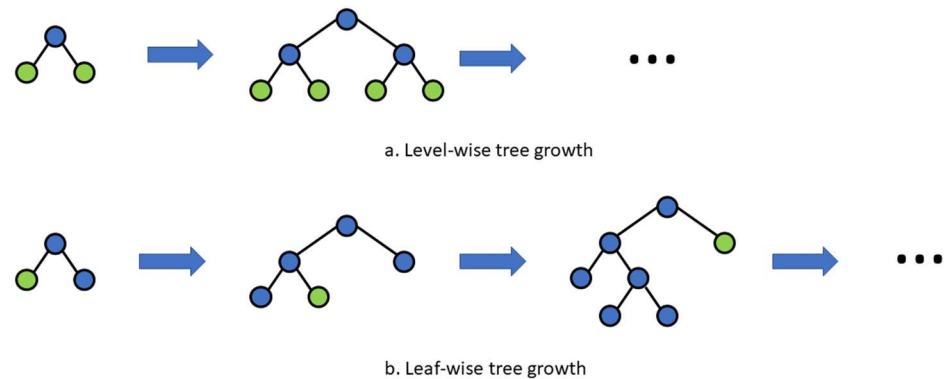


Figure 3. A pruning method of models using the gradient boosting algorithm. Each circle represents a node, and green is a leaf node.

5.2.2. Support Vector Machine

SVM is an algorithm that defines a decision hyperplane that can distinguish each sample and aims to maximize the margin of a given sample and decision hyperplane [18]. The hyperplane means an $(N - 1)$ -dimensional subspace in an N -dimensional space. Figure 4 shows that data exist in two-dimensions. The dimension of the hyperplane that separates the class boundary of the support vector is a one-dimensional line. SVM has the advantage of being able to accommodate nonlinear boundaries by extending the variance space using the kernel.

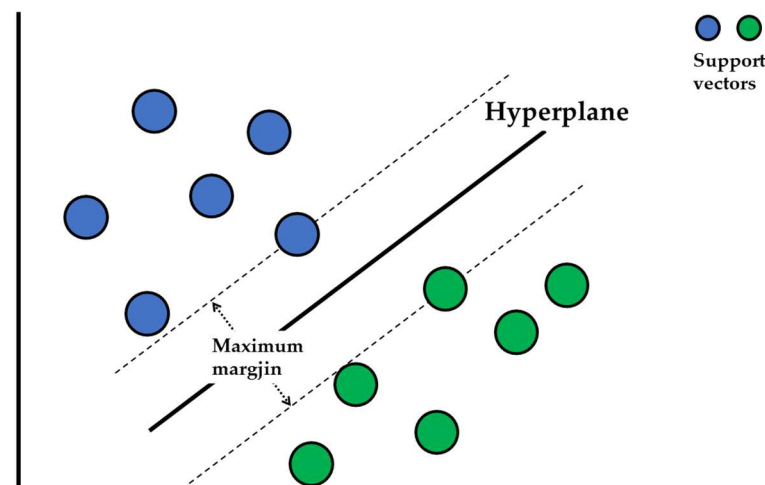


Figure 4. Examples of data and SVM that exist in 2-dimensions.

5.3. Deep Learning-Based Approach

Deep learning is one of the machine learning methods that train by forming an artificial neural network with a shape like the human brain. By deeply stacking multiple hidden layers, it is possible to build high-level abstractions through non-linear relational modeling [19]. When building a deep learning structure, there are layers such as Recurrent

Neural Network (RNN), Convolutional Neural Network (CNN), Fully Connected Layer (FCN), etc., and the model is built by stacking several layers according to the task to be solved [20–22].

In this study, we built the architecture of a deep learning-based model to enable input and output in a sequence format, as shown in Figure 5. In other words, the deep learning-based model does not use the data reconstruction method used to find the segmentation point of a sentence in the machine learning-based approach. For example, the correct answer for a sequence in which the “EC” token is a segmentation point (<S>), such as (NNG, JX, VV, EC, NNG, JX, VV, EF, SF), consists of a sequence such as (0, 0, 0, 0, 1, 0, 0, 0, 0, 0). In the topic maintenance or change classification model, sequence output is not required, so only the output layer is changed so that it can be output as a binary classification.

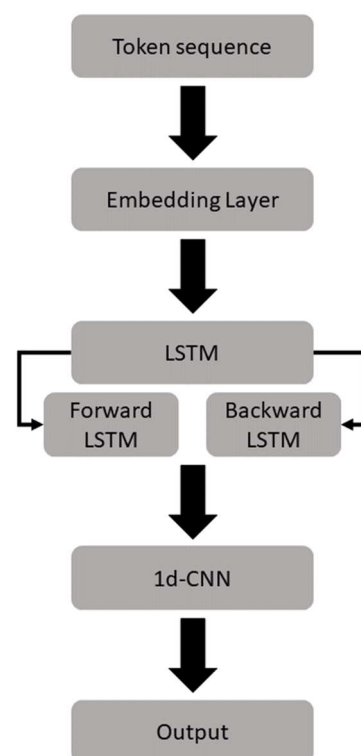


Figure 5. Deep neural network structures used in deep learning-based approach.

The embedding layer plays a lookup table role representing integer-encoded tokens as vectors. First, words are initialized as D -dimensional dense vectors having random values, and then the weights are updated through learning. When learning for a specific task, the weights of the embedding layer are optimized for the task to represent the relationship between tokens. That is, the tokens having similar characteristics are arranged closer to each other in the vector space. The embedding layer converts sequences into vectors, which are then fed to the Long Short-Term Memory (LSTM).

LSTM is one of the RNN models that complement the long-term dependency problem. LSTM extracts features from the sequence by remembering the feature to be remembered from the input sequence and forgetting the unnecessary feature. The output of the LSTM is fed to a 1-dimensional Convolutional Neural Network (1d-CNN).

1d-CNN extracts a feature map by capturing local features in a sequence. Depending on the kernel size of 1d-CNN, it can be expressed as a bi-gram when it is two and a trigram when it is three. In this study, the kernel size of 1d-CNN was set to three. The feature map extracted through 1d-CNN outputs the result according to the problem through the output layer.

The topic maintenance or change classification model flattens the output of 1d-CNN in one dimension and then outputs a probability of zero to one using the sigmoid function.

The sentence generation model outputs the probability of a segmentation point for each time-step in the output of 1d-CNN.

6. Experiment and Result

6.1. Data Split

We grouped training, validation, and testing data according to children to collect generalized assessment results. We assigned all the utterances of child A as training data, all of the utterances of child B as validation data, and all the utterances of child C as test data. Of the 1800 data of utterances collected from a total of 67 children, the data used for training and validation data are 1675 collected from 63 children, and the test data to be evaluated are 125 sentences collected from four children. Of the 125 test data, 25 are topic change classes, and the remaining 100 are topic maintenance classes.

Some of the training data were used as validation data. The performance was monitored through the validation data and the test data were evaluated with the best performing model.

6.2. Parameters

The LGBM uses Gradient Boosting Decision Tree as a boosting parameter. The default value of num_leaves parameter is 31, and the objective is binary, so that binary classification is targeted. The weights of the model are saved whenever there is an improvement during the 1000 training processes. If there is no improvement during the 50 learning processes, it is stopped early. Other parameters are set as default values.

The SVM uses a radial basis function (RBF) as a kernel parameter. The C value is 1.0, which gave a small value. Other parameters are set as default values.

For DNN, the dimension of the embedding layer is set to 32, and the number of nodes for 1d-CNN is set to 64. The number of nodes of LSTM is set to 32, and it is set to learn bidirectional feature. The optimizer uses Adam optimizer [23], and the loss function uses binary cross entropy. The batch size is 16, and 100 epochs is running. If there is no improvement during the 10 epochs, it is stopped early.

6.3. Evaluation Metrics

The topic maintenance or change classification model measures the accuracy of binary classification to use quantitative evaluation indicators. The sentence generation model can be quantitatively evaluated based on the directly defined segmentation point. However, even if the wrong segmentation point is selected, a natural questioning-like sentence can be generated. Therefore, we directly evaluate by humans whether the sentence is natural or not. In other words, when the imitative sentence generated based on the segmentation point predicted by each model is used as a questioning-like sentence during a conversation, humans vote for a sentence that is not awkward. This allows voting on multiple models when evaluating a sentence. For example, when the sentence “geulaeseo jeodo geulim daehoe nagass-eoyo.” (“I also participated in a painting contest.”) is input to the model, LGBM, SVM, and DNN can predict difference segmentation point. SVM is “jeodo geulim daehoe nagass-eoyo?” (“Did I participate in the painting contest too?”), LGBM is “geulim daehoe nagass-eoyo?” (“Did you participate in the painting contest?”), and DNN is “daehoe nagass-eoyo?” (“Did you participate in the contest?”). Among them, according to the flow of conversation, humans vote for sentences that are not awkward as a questioning-like sentence and use them as qualitative evaluation indicators.

6.4. Experiment Results

Table 3 shows the performance of each model. Performance measures were evaluated with sentences collected from four children who were not used in the train learning process. The topic maintenance or change classification model shows the binary classification accuracy. The sentence generation model is the result of eight evaluators allowing overlapping votes for the imitation sentences generated by each model and averaging the votes of the eight evaluators.

Table 3. Performance evaluation results of each model.

Model	Topic Maintenance or Change Classification (Accuracy)	Imitation Sentence Generation (Votes)
SVM	79.41	60.375
LGBM	89.70	76
DNN	94.17	36.75

In the topic maintenance or change classification task, the model of Stage 1 using the DNN showed better performance than the rest of the models with 94.17%. However, in the model of Stage 2 evaluation of the imitation sentence generation task after predicting the segmentation point, the evaluation result of the DNN is the worst with an average of 36.75 votes, and LGBM showed the best performance with an average of 76 votes. SVM showed poor overall performance in classification task and imitation sentence generation task.

For the task of maintaining or changing the topic, the words indicating topic change, such as “deo eobs-eoyo.” (“No more.”) or “geuman hallaeyo.” (“I want to quit.”), are clear, so there is no significant difference in performance between each analysis model, and the overall result is excellent. On the other hand, the model for generating imitation sentences differed significantly by analysis models. This is because when reconstructing sequence data to use SVM and LGBM models, a lot of data are created from one sequence datum, so the model can learn many cases through various samples. Thus, it seems that SVM and LGBM performed better than DNN.

7. Conclusions

In this study, a machine learning-based model is proposed to automate the process of collecting transcriptional data used to evaluate children’s language development level. The utterance collection model is performed in the steps of maintaining the topic and changing the topic after initiating the conversation during the conversation collection protocol. The main role of the proposed collection model is to analyze children’s utterances, classify whether to maintain the topic or change it, and then maintaining the topic, generating a questioning-like sentence according to the segmentation point of the sentence. In order to perform the above role, we used previously collected data. For the analysis model, we tested which of the three models, LGBM, SVM, and DNN, is suitable. As a result, in the topic maintenance or change classification tasks, DNN showed the best performance with 94.17%, and in the task of generating imitation sentences through the segmentation point, humans directly evaluated the naturalness of the sentence. LGBM showed the best performance with an average of 76 votes. As a result of confirming the performance results of the two tasks covered in this study, LGBM showed excellent results on average in both tasks, so LGBM is judged to be a suitable analysis model.

Based on the above conclusions, it is clearly difficult to apply the proposed method immediately in the process of the collection protocol. In the next study, we will structure the utterance data of more children and construct a model that generates questioning-like sentences more naturally by analyzing them.

Author Contributions: Conceptualization, J.-M.C. and Y.-S.K.; data curation, Y.-K.L., J.-D.K. and C.-Y.P.; formal analysis, J.-M.C., Y.-K.L. and C.-Y.P.; methodology, J.-M.C., J.-D.K. and C.-Y.P.; validation, J.-D.K. and C.-Y.P.; resources, Y.-K.L. and Y.-S.K.; visualization, J.-M.C. and J.-D.K.; writing—original draft preparation, J.-M.C.; writing—review and editing, Y.-S.K.; supervision, Y.-S.K.; project administration, Y.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF) (NRF-2019S1A5A2A03052093), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (NO.2021-0-02068, Artificial Intelligence Innovation Hub (Seoul National University)), and ‘R&D Program for Forest Science Technology (Project No. 2021390A00-2123-0105)’ funded by the Korea Forest Service (KFPI).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, S.W.; Shin, J.B.; You, S.; Yang, E.J.; Lee, S.K.; Chung, H.J.; Song, D.H. Diagnosis and Clinical Features of Children with Language Delay. *J. Korean Acad. Rehabil. Med.* **2005**, *29*, 584–590. (In Korean)
2. Kim, Y.T. Content and Reliability Analyses of the Sequenced Language Scale for Infants (SELSI). *Commun. Sci. Disord.* **2002**, *7*, 1–23. (In Korean)
3. Lee, H.J.; Kim, Y.T. Measures of Utterance Length of Normal and Language—Delayed Children. *Commun. Sci. Disord.* **1999**, *4*, 1–14. (In Korean)
4. Jin, C.; Choi, H.J.; Lee, J.T. Usefulness of Spontaneous Language Analysis Scale in Patients with Mild Cognitive Impairment and Mild Dementia of Alzheimer’s Type. *Commun. Sci. Disord.* **2016**, *21*, 284–294. (In Korean) [[CrossRef](#)]
5. Oh, K.A.; Nam, K.W.; Kim, S.J. The Development of Verbs through Semantic Categories in the Spontaneous Speech of Preschoolers. *J. Speech-Lang. Hear. Disord.* **2014**, *23*, 63–72. (In Korean)
6. Park, C.H.; Lee, S.H. *A Study on Use of Expression Pattern in Spontaneous Speech from 3 to 5-Year-Old Children*; Korean Speech-Language & Hearing Association: Busan, Korea, 2017; pp. 187–190. (In Korean)
7. Jin, Y.S.; Pae, S.Y. Grammatical Ability of School-aged Korean Children. *J. Speech Hear. Disord.* **2008**, *17*, 1–16. (In Korean) [[CrossRef](#)]
8. Kim, S.S.; Lee, S.K. The Comparison of Conversation, Freeplay, and Story as Methods of Spontaneous Language Sample Elicitation. *KASA* **2008**, *13*, 44–62. (In Korean)
9. Park, Y.J.; Choi, J.; Lee, Y. Development of Topic Management Skills in Conversation of School-Aged Children. *Commun. Sci. Disord.* **2017**, *22*, 25–34. (In Korean) [[CrossRef](#)]
10. Park, E.J.; Cho, S.Z. KoNLPy: Korean natural language processing in Python. In Proceedings of the Annual Conference on Human and Language Technology, Kangwon, Korea, 10–11 October 2014; pp. 133–136.
11. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
12. Rana, S.; Garg, R. Slow learner prediction using multi-variate naïve Bayes classification algorithm. *Int. J. Eng. Technol. Innov.* **2017**, *7*, 11–23.
13. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
14. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 785–794.
15. Pokhrel, P.; Ioup, E.; Hoque, M.T.; Abdelguerfi, M.; Simeonov, J. A LightGBM based Forecasting of Dominant Wave Periods in Oceanic Waters. *arXiv* **2021**, arXiv:2105.08721.
16. Li, F.; Zhang, L.; Chen, B.; Gao, D.; Cheng, Y.; Zhang, X.; Peng, J. A light gradient boosting machine for remaining useful life estimation of aircraft engines. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3562–3567.
17. Taha, A.A.; Malebary, S.J. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access* **2020**, *8*, 25579–25587. [[CrossRef](#)]
18. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
21. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural. Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
23. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.