Toward Application to General Conversation Detection of Dementia Tendency from Conversation Based on Linguistic and Time Features of Speech

Hiroshi Sogabe, Masayuki Numao
Department of Communication Engineering and Informatics
The University of Electro-Communications
h.sogabe123@gmail.com, masayuki.numao@uec.ac.jp

Abstract

Currently, MRI examinations and neuropsychological tests are used to screen for dementia, but they are problematic because they overwhelm medical resources and are highly invasive to patients. If automatic detection of dementia from conversations becomes feasible, it will reduce the burden on medical institutions and realize a less invasive screening method. In this paper, we constructed a machine learning model to identify dementia by extracting linguistic features and time features from the elderly corpus with a control group. Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) were used in the model. We compared the AUC of the single topic model and the general topic model in three cases: (I) All Features, (II) Gini Impurity, and (III) PCA + Gini Impurity. The AUC of the model constructed using RF in (III) for a single topic was 0.91. Furthermore, topic analysis showed that topics with high similarity in utterance content are effective in identifying MCI. In the case of the general topic, the model with AUC of 0.8 showed a high identification performance for unknown topics by cross validation on a topic-by-topic basis, indicating that the general topic model developed in this study can be applied to general conversation.

Introduction

In Japan, the number of elderly people with dementia is increasing due to the aging of the population. As of 2012, the number of elderly people with dementia was 4.62 million, or approximately one in seven persons aged 65 or older (Cabinet Office in Japan 2017). In conjunction with this, the social burden caused by the growing medical costs and the burden on family members due to caregiving are becoming an issue. Since the progression of dementia can be slowed down by treatment, early detection is important to alleviate the problems. Early detection of dementia is also important from the perspective of well-being. Early treatment can maintain the patient's daily functioning. This will allow the patient to lead an independent and personal life, which will enhance their physical and mental health and sense of well-being. In addition, it is possible to respect the patient's right to self-determination, for example, by informing family members of the patient's wishes regarding property management and care while symptoms are mild. Furthermore, by making family members and caregivers aware of dementia in advance, it is possible to prevent the breakdown of family relationships due to sudden mood changes and delusions, which are symptoms of dementia. Thus, early detection of dementia is very important not only for the patients themselves, but also for their families and caregivers, as well as for the well-being of society. Currently, MRI examinations and neuropsychological tests are used as diagnostic methods for dementia, but they are problematic due to a lack of medical resources and are highly invasive to patients. Therefore, it is difficult to provide routine screening tests for dementia to the elderly, and the detection of dementia is delayed. Against this background, noninvasive dementia screening based on the elderly's speech has been attracting attention in recent years.

In this paper, we extract features from elderly people's speech data and identify healthy elderly people from those with mild cognitive impairment (MCI) using machine learning, with the aim of realizing a noninvasive dementia screening test for early detection of dementia.

Related Work

Shibata et al. (2019) pointed out that the lack of widely available speech data of speakers with dementia in Japanese is a reason for the paucity of studies on cognitive function assessment based on language ability in Japan compared to English-speaking countries. They constructed a speech data set of healthy elderly and MCI for Japanese speakers. The

constructed dataset is titled "the elderly corpus with a control group" and They conducted an identification experiment between healthy elderly (n=45) and those with MCI (n=15) using the corpus. Support Vector Machine (SVM) and Logistic Regression (LR) were used for the identification model, and the model was evaluated by calculating the mean value of AUC using 5-fold cross validation. SVM showed the highest identification performance on a single task with an AUC of 0.74, while SVM showed the highest identification performance on multiple tasks with an AUC of 0.85.

Banhong and Okazaki (2019) conducted an experiment to identify between speakers with MCI (n=15) and other speakers including non-elderly people (n=65) by extracting not only linguistic features but also audio features from the elderly corpus with a control group. They compared the accuracy of using only linguistic features, only audio features, and both audio and linguistic features, and reported that the accuracy was highest when only linguistic features were used, at 0.829.

Ishihara, Iribe and Kitaoka (2020) focused on the speech characteristics of speakers with dementia, who are less likely to use complex constructions, and extracted features such as maximum dependency distance as syntactic complexity from the speech and chat dialogues of 24 elderly subjects during HDS-R implementation. Using the features, they identified healthy elderly subjects (n=11) from those with dementia (n=13) with accuracy of 0.93.

Shibata et al. (2016) investigated linguistic features specific to dementia speakers with the aim of early detection of dementia. First, based on the results of a neuropsychological test, 18 subjects aged 53-90 years were divided into two groups: healthy subjects (n=9) and dementia subjects (n=9). The difference between the two groups was evaluated by a t-test at a significance level of 0.05 on word frequency for each word category based on the Linguistic Inquiry and Word Count (LIWC) emotional dictionary. The results showed that pronouns were used significantly more frequently in the dementia patients.

Sluis et al. (2020) extracted multiple silence features from the speech of healthy subjects (n=20), mild dementia subjects (n=20), and moderate dementia subjects (n=20) in the image description task in the Pitt Corpus. To clarify the speech features of speakers with dementia, those features were evaluated by statistical tests. The results revealed that silence in speech significantly increased in proportion to the severity of dementia.

Yoshii et al. (2020) conducted a statistical analysis of speech features in speech during neuropsychological tests and daily conversation with a humanoid robot in healthy subjects (n=45) and patients with MCI (n=45). The results showed that there were significant differences in speech features such as speech duration, response time, silence duration, and filler duration in daily conversation.

Dataset

In this paper, we use the elderly corpus with a control group (Shibata et al. 2019). This corpus consists of speech utterances and their transcriptions of elderly aged 65 and over (n=60) and non-elderly (n=20) subjects on 10 topics, consisting of an episode description task, an image description task, and an animation description task. In addition, age, gender, last education, and MMSE scores for the elderly are also included.

The episode description task consists of descriptions of the following eight topics within one to two minutes.

- EP1: A recent event that made you sad
- EP2: A recent event that made you anxious
- EP3: A recent event that made you angry
- EP4: A recent event that made you feel disgusted
- EP5: A recent event that made you surprised
- EP6: A recent event that made you happy
- EP7: People you admire
- EP8: Your recent passions

The image description task (STORY) requires the user to describe the content of the image shown in Figure 1 within one to two minutes.

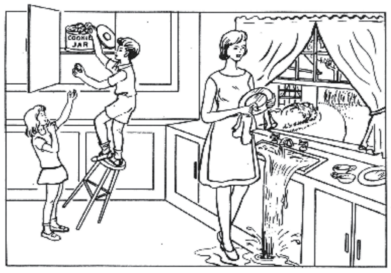

Figure 1: Cookie theft picture (Reprinted from (Shibata et al. 2019))

The animation description task (ANIME) consists of watching a one-minute animation called "NAIST dog story" and explaining what happened within one to two minutes. Figure 2 shows a screenshot of each scene of the NAIST dog story. The contents of each scene are shown below.

(1) Blue dog and red dog taking a walk
(2) Dog comes across a scene where a large unidentified creature is abusing a small unidentified creature
(3) The blue dog surprises the large unidentified creature
(4) The small unidentified creature thanks the dogs for saving and leaves on the back of the blue dog

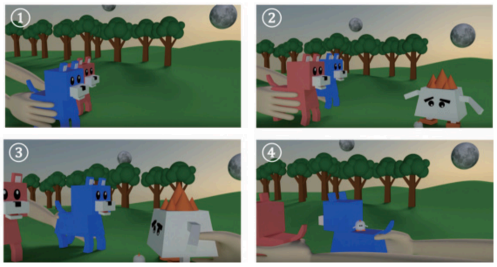

Figure 2: NAIST dog story (Reprinted from (Shibata et al. 2019))

In this paper, as in a previous study (Shibata et al. 2019), only data from 60 elderly subjects were used. 45 elderly subjects with MMSE scores of 28 or higher were defined as healthy control (HC), and 15 elderly subjects with MMSE scores of 23 or higher and 27 or lower were defined as mild cognitive impairment (MCI).

Features

Features were extracted for each subject's utterance on each topic.

Language Features

The following 13 linguistic features were extracted from each utterance.

- number of words (N_w)
- *number of words (N_wp)
- *number of different words (N_tp)
- *Type Token Ratio (TTR): N_tp / N_wp
- *Part-of-Speech Ratio
  - *noun ratio (R_noun)
  - *pronoun ratio (R_pron)
  - *verb ratio (R_verb)
  - *adjective ratio (R_adj)
  - *adverb ratio (R_adv)
- pronoun ratio to noun (R_p2n)
- number of fillers (N_f)
- filler ratio (R_f): N_f / N_w
- max of dependency distance (DD_max)

The AmiVoice API was used for the extraction of N_f and DD_max. AmiVoice API is an API that provides annotation information about the start and end point time, transcription, and filler for each word in speech. In the extraction of DD_max, since DD_max becomes large in sentences containing fillers regardless of the complexity of the syntax, DD_max was extracted from transcriptions excluding fillers using the AmiVoice API. The Japanese language dependency parser CaboCha (Kudo and Matsumoto 2002) was used to calculate the dependency distance.

Other features were obtained by counting the number of words and the number of occurrences of each part of speech in the corpus using the Japanese morphological analyzer MeCab (Kudo, Yamamoto and Matsumoto 2004). For features marked with a "*" symbol, only nouns, verbs, adjectives, and adverbs in the corpus were used for feature extraction. N_w is the number of words without part-of-speech restriction, and N_wp is the number of words with part-of-speech restriction. The "*Part-of-Speech Ratio" is the ratio of each part-of-speech to N_wp.

Time Features

The following five time features were extracted from each utterance.

- response time (T_res) : time to speech start point
- utterance duration (T_u) : time length from speech start point to speech end point
- filler time ratio (TR_f) : ratio of filler duration in utterance duration (T_f / T_u)
- silence time ratio (TR_s) : ratio of silence duration in utterance duration (T_s / T_u)
- filler silence time ratio (TR_fs) : ratio of filler silence duration in utterance duration (T_fs / T_u)

The indices used in the above time feature extraction are shown below.

- filler duration (T_f) : sum of filler duration
- silence duration (T_s) : sum of silence duration in utterance duration
- filler silence duration (T_fs) : sum of filler duration and silence duration in utterance duration (T_f + T_s)

The time information of AmiVoice API was used to extract time features. Figure 3 shows the correspondence between the time features in the speech waveform and the indices used for feature extraction.

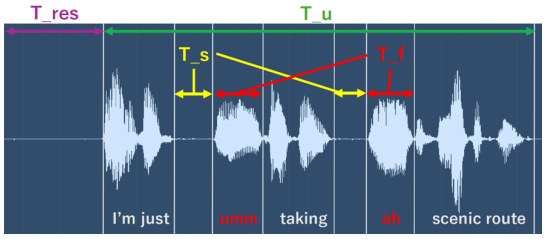

Figure 3: Time features in speech waveforms and indices used for feature extraction

Identification Experiment

As in the previous study (Shibata et al. 2019), we conducted identification experiments on a single topic and compared the AUC.

	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	STORY	ANIME
RF	0.62	0.56	0.59	0.54	0.53	0.63	0.69	0.69	0.77	0.64
SVM	0.68	0.43	0.61	0.33	0.41	0.36	0.72	0.75	0.74	0.36
LR	0.82	0.51	0.79	0.59	0.57	0.75	0.74	0.72	0.76	0.80

Table 1: AUC in (I) of single topic

	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	STORY	ANIME
RF	(0.80, 5)	(0.76, 5)	(0.74, 1)	(0.72, 1)	(0.7, 4)	(0.71, 11)	(0.83, 4)	(0.78, 17)	**(0.88, 11)**	(0.82, 6)
SVM	(0.85, 3)	(0.79, 3)	(0.72, 16)	(0.71, 1)	(0.60, 15)	(0.68, 8)	(0.82, 8)	(0.83, 9)	(0.79, 13)	(0.75, 7)
LR	**(0.86, 5)**	(0.69, 3)	(0.8, 11)	(0.71, 1)	(0.71, 3)	(0.77, 5)	(0.80, 10)	(0.78, 8)	(0.82, 10)	(0.80, 17)

Table 2: AUC in (II) of single topic

	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	STORY	ANIME
RF	**(0.91, 3)**	(0.84, 8)	(0.80, 4)	(0.69, 9)	(0.71, 4)	(0.84, 3)	(0.84, 3)	(0.74, 10)	**(0.88, 12)**	(0.72, 5)
SVM	(0.85, 2)	(0.8, 4)	(0.72, 1)	(0.57, 1)	(0.86, 2)	(0.80, 4)	(0.8, 6)	(0.76, 15)	**(0.88, 7)**	(0.68, 6)
LR	(0.88, 2)	(0.83, 1)	(0.85, 11)	(0.74, 8)	(0.80, 4)	(0.8, 3)	(0.81, 2)	(0.84, 6)	(0.8, 7)	(0.80, 16)

Table 3: AUC in (III) of single topic

	(I)	(II)	(III)
RF	0.77	**(0.80, 11)**	(0.77, 9)
SVM	0.77	**(0.80, 11)**	(0.77, 8)
LR	0.78	(0.79, 16)	(0.79, 10)

Table 4: AUC in general topic

Furthermore, we constructed a general topic model for application to general conversation. To evaluate the identification performance of the general topic model for unknown topics, a 10-fold cross validation splitting the data by topic was conducted using the speech data of all topics.
Then, based on the feature importances obtained from the construction of the general topic model, we identified features that are effective in identifying dementia in a wide range of topics.

Single Topic

Models and Evaluation Methods

Identification experiments were conducted on HC(n=45) and MCI(n=15). All features were standardized to have a mean of 0 and variance of 1 for each topic. Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) were used for the model. The models were evaluated by calculating the mean value of AUC using 5-fold cross validation, as in the previous study (Shibata et al. 2019). The implementation and evaluation of the model in this experiment were performed using Scikit-learn (Pedregosa et al. 2011). The default values set by Scikit-learn were used for the RF, SVM, and LR parameters in this experiment. In addition, we compared the AUC of feature selection for training in the following cases.

(I) All Features
(II) Gini Impurity
(III) PCA + Gini Impurity

(I) All Features

Models were built for each topic using all 18 features described above, and their AUCs were calculated. Table 1 shows the AUC of each model for each topic.

(II) Gini Impurity

The feature selection and model training procedures are shown below.

(1) Select a topic (EP1, EP2, ..., EP8, STORY, ANIME)
(2) Select a model (RF, SVM, LR)
(3) Using all the features extracted from the selected topics, train with RF to obtain the importance of the features based on Gini impurity
(4) Using only the N features with high importance, train on the selected model and calculate AUC (repeat for N=1, 2, ..., 18)
(5) The maximum value of the AUC obtained in (4) is used as the AUC in the selected model for the selected topic
(6) Perform (1)-(5) for all topic-model pairs

Table 2 shows the AUC for each model for each topic obtained by the above procedure. The value of each cell in the table 2 is (AUC, number of features when AUC is achieved). Table 2 shows that the AUC of RF in STORY was the largest at 0.88, exceeding the AUC of 0.85 in the previous study (Shibata et al. 2019). The AUC of LR in EP1 was 0.86, which exceeded 0.85.

(III) PCA + Gini Impurity

The feature selection and model training procedures are shown below.

(1) Select a topic (EP1, EP2, ..., EP8, STORY, ANIME)
(2) Select a model (RF, SVM, LR)
(3) 18 features extracted from the selected topic are transformed into 18 principal components by principal component analysis (PCA)
(4) Using all the transformed principal components, train with RF to obtain the importance of the principal components based on Gini impurity
(5) Using only the N principal components with high importance, train on the selected model and calculate the AUC (repeat for N=1, 2, ..., 18)
(6) The maximum value of the AUC obtained in (5) is used as the AUC in the selected model for the selected topic
(7) Perform (1) - (6) for all topic-model pairs

Table 3 shows the AUC for each model for each topic. The value of each cell in the table 3 is (AUC, number of principal components when AUC is achieved). Table 3 shows that in (III), the AUC of the RF in EP1 was the largest at 0.91, exceeding the value of 0.85 in the previous study (Shibata et al. 2019). It also exceeded 0.85 for EP5 and STORY.

General Topic

Models and Evaluation Methods

A 10-fold cross validation was conducted using all elderly speech data(n=600) of 10 topics, splitting the data by topic. As with single topic, RF, SVM, and LR were used to evaluate the model by calculating the mean value of AUC in cases (I), (II), and (III). The standardization of the features and the transformation by PCA were performed not on a topic-by-topic basis but on a whole-data basis.

Identification Performance

The case-specific AUC for (I), (II), and (III) in the general topic model are shown in Table 4. Table 4 shows that the AUC of RF and SVM for method (II) is the largest at 0.8.

Discussion

Single Topic

(I) All Features

Table 1 shows that the average AUC for method (I) is low at 0.62. The reason for the low AUC may be that the model was overfitted due to the large number of features relative to the number of data.

(II) Gini Impurity

Table 2 shows that the average AUC for method (II) is 0.77, which is an improvement compared to method (I). The reason for the improved AUC can be attributed to the suppression of model overfitting by feature selection using method (II).

Table 2 shows that STORY has the largest AUC of 0.88 in (II), suggesting that STORY is the most effective topic for identification. Hypothesis for the reasons for the higher identification performance of STORY compared to episode descriptions are discussed below. In the case of episode description, subjects can select "speech content" with low cognitive load, whereas in the case of STORY, it is difficult to reduce the cognitive load by selecting speech content due to the nature of all subjects' explanations of the same content. Therefore, it is thought that differences in cognitive function are more likely to be reflected in speech in STORY. Furthermore, since the content of speech is common and many subjects recall the same words during speech, memory loss, a typical symptom of dementia, may be reflected in speech in the form of increased silence and pronoun.

EP1 is the next highest AUC at 0.86, and EP1 is also an effective topic for identification.

(III) PCA + Gini Impurity

Table 3 shows that the average AUC for method (III) is 0.79, which is an improvement compared to methods (I) and (II). The reason for the improved AUC can be attributed to the suppression of model overfitting by feature selection using method (III). In fact, the average number of features used to achieve the maximum AUC was 7.3 for (II) and 5.6 for (III), indicating that the number of features used for training was reduced.

From Table 3, the AUC of EP1 was the highest in (III) at 0.91, followed by the AUC of STORY at 0.88, suggesting that EP1 and STORY are effective topics for identification, as in (II). Below, we discuss effective features in EP1 and STORY, which showed high AUC.

Table 5 shows the principal component loadings for EP1. In Table 5, pc_N means the N-th principal component loadings. pc_N is ordered from top to bottom by importance based on Gini impurity. Some columns with small values are omitted. In the following, we discuss the features that contributed to the identification, focusing on large absolute values of pc_N. For convenience of explanation, the features obtained by PCA are referred to as principal components.

In EP1, the values corresponding to R_pron and R_p2n of pc_3 are large. This indicates that pronoun ratio is effective in the identification of MCI. These features are related to memory loss, a symptom of dementia. Dementia speakers fail to recall nouns or take longer to recall nouns due to memory loss. Thus, it is thought that the noun that was supposed to be referred to is replaced by a pronoun in the utterance. Pronoun ratio reflect this characteristic. In the pc_5, many features such as nouns and pronouns contribute. Nouns are related to memory. Dementia speakers tend to

rank	pc	R_noun	R_pron	R_verb	R_adj	R_adv	R_p2n	R_f	T_res	T_u	TR_f	TR_s	TR_fs
1	pc_3	0.26	**0.59**	-0.24	-0.08	-0.11	**0.56**	-0.21	0.14	0.05	-0.22	0.08	-0.05
2	pc_5	**-0.29**	**0.27**	**0.31**	**-0.27**	**0.38**	**0.34**	**0.32**	-0.17	-0.06	**0.28**	0.03	0.22
3	pc_4	-0.004	0.14	-0.22	0.48	-0.07	0.13	0.19	-0.30	-0.22	0.21	**-0.50**	**-0.39**

Table 5: Principal component loadings in EP1

rank	pc	N_w	N_wp	N_tp	R_noun	R_pron	R_p2n	R_f	T_res	T_u	TR_f	TR_s	TR_fs
1	pc_1	**0.38**	**0.37**	**0.36**	0.07	0.10	0.09	0.16	-0.08	**0.31**	0.21	-0.18	-0.07
2	pc_7	0.05	0.02	0.11	-0.17	-0.005	-0.01	-0.02	**0.79**	0.003	0.03	-0.15	-0.13
3	pc_5	0.08	0.08	0.07	0.15	**-0.50**	**-0.55**	0.06	0.25	0.22	-0.004	**0.29**	**0.30**

Table 6: Principal component loadings in STORY

	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	STORY	ANIME
Similarity	**0.21**	0.18	0.14	0.13	**0.11**	0.20	0.19	0.13	**0.28**	**0.17**

Table 7: Similarity of utterance content of each topic

	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	STORY	ANIME
(II)	0.84	0.75	0.75	0.71	0.67	0.72	0.82	0.80	0.83	0.79
(III)	0.88	0.82	0.79	0.67	0.79	0.81	0.81	0.78	0.85	0.73

Table 8: Mean AUC of each topic

omit nouns when they fail to recall nouns due to memory loss. Noun ratio reflects this feature. From pc_4, we can see that the speech hesitation (TR_s, TR_fs) is effective. These features are thought to reflect the fact that it takes time to think about utterance content due to cognitive decline and that it takes time to recall words due to memory loss.

In STORY, Table 6 shows the principal component loadings. As in EP1, some columns are omitted. The pc_1 indicate that the utterance lengths (N_w, N_wp, N_tp, T_u) are effective. Utterance length reflects the tendency to cut off utterances early due to decreased motivation, which is a symptom of dementia. In addition, the dementia speaker does not notice the detailed description of the image due to a lack of attention, and the number of events mentioned is less. the utterance lengths also reflect this feature. The pc_7 indicate that response time (T_res) is effective. Response time is thought to reflect the time required to think about utterance content and comprehend image content due to cognitive decline. The pc_5 indicate that pronouns (R_pron, R_p2n) and speech hesitation (TR_s, TR_fs) are effective.

For a single topic, the AUC for method (III) was 0.91, which was higher than AUC of the previous work (Shibata et al. 2019). Furthermore, we discussed the effective features and found that the speech hesitation (TR_s, TR_fs) and pronouns (R_pron, R_p2n) were particularly effective.

Topic Analysis

In the following, we will test the previously mentioned hypothesis regarding the reason for the effectiveness of STORY in identification. To test the hypothesis, we first quantified the similarity of the utterance content in each topic. Assuming that the utterance content is reflected in nouns, each utterance was converted into a vector by word frequency for each topic for nouns excluding pronouns. Each vector was normalized to have a norm of 1 so that the inner product between the vectors is equal to the cosine similarity. Let $\{v_1, v_2, \dots, v_{60}\}$ be the utterances vectorized by the above method. Please note that the number of speech samples for each topic used in this study is 60. The sum of "the inner product between the different vectors ($v_i \cdot v_j$ $(i \neq j \wedge i, j \in \{1,2, \dots ,60\})$)" for each topic was then calculated, and the average value was defined as the similarity of the utterance content for each topic. Equation (1) shows the definition formula for the similarity of the utterance content of each topic.

$$\frac{1}{S_{59}} \sum_{i=1}^{59} \sum_{j=i+1}^{60} (v_i \cdot v_j) \qquad (1)$$

where

$$v_i,\ v_j \in \{v_1, v_2, \dots, v_{60}\},$$

$$S_{59} = \sum_{k=1}^{59} k$$

Table 7 shows the similarity of the utterance content for each topic. The high similarity indicates that many subjects are speaking the same content.

The mean AUCs for each topic in (II) and (III) are shown in Table 8. The correlation between the mean AUC in Table 8 and the similarity in Table 7 is 0.64 for (II) and 0.62 for (III), indicating that the higher the similarity, the higher the AUC. This result confirms the previously mentioned hypothesis for the reasons for the higher identification performance of STORY compared to episode descriptions. In fact, STORY with the highest similarity of utterance content among all topics has the high AUC. Furthermore, EP1, which had the highest similarity among the EPs, had the highest AUC. The reason for the high AUC of EP1 may be that many subjects selected similar utterance contents, and the cognitive load such as recall of the same word was reflected in their utterances. EP5, which has the lowest similarity within EPs, has the lowest maximum AUC in (II).

If the fact that "all subjects give explanations for the same content" in STORY suppresses subject's selection of utterance content and contributes to identification, then the same AUC should be shown in ANIME, which has the same characteristics as STORY. However, the AUC of ANIME is lower than that of STORY. The content used in ANIME lacks concreteness, such as the appearance of unidentified creatures, and is more abstract than STORY. As a result, as shown in Table 7, the similarity of utterance content was low and the differences in cognitive function were not reflected in the utterances. In fact, STORY with a high similarity has a high concreteness of the content to be described (the characters are likely to be a mother and her children, and the events include "children trying to take things from a cupboard" and "water overflowing from a sink"). In other words, the abstract nature of the content to be described in ANIME resulted in a low similarity of utterances, and thus a high AUC could not be achieved. These results indicate that for content description tasks such as STORY and ANIME, it is important for MCI identification to use highly concrete content that does not contain ambiguous events or characters with ambiguous naming.

The results of the topic analysis showed that there is a significant relationship between the similarity of utterance content and the identification of MCI. Topics with high similarity were shown to be effective in the identification of MCI because the cognitive load reduction caused by the selection of utterance content did not occur, and the recall of the same word reflected memory in the utterance. From this perspective, STORY and EP1 were shown to be effective topics for identification. Furthermore, it was found that the use of highly concrete content is effective for MCI identification in content description tasks such as STORY and ANIME.

General Topic

Identification Performance Evaluation

Table 4 shows that the AUC for method (II) is 0.8, which is the maximum for identification by general topic model. To evaluate the identification performance for unknown topics not included in the training data, the general topic model calculates the AUC by cross validation on a topic-by-topic basis. The results showed high identification performance with an AUC of 0.8 for method (II). The high identification performance of the general topic model for unknown topics indicates the possibility of applying the general topic model to general conversation.

Feature Evaluation

The feature importances of method (II) in the general topic model are shown in Figure 4. The features located above the red line are the features that were used for training when the maximum AUC was achieved. Based on feature importances, we discuss the features that are effective for identification of MCI on a wide range of topics.

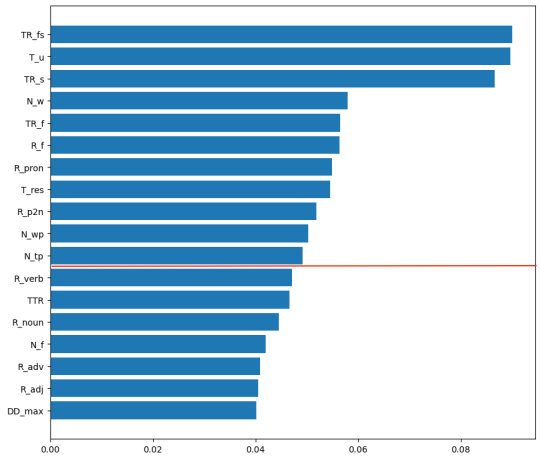

Figure 4: Feature importances of method (II) in the general topic model

From Figure 4, the top 11 features in terms of importance were: speech hesitation (TR_s, TR_fs, TR_f, R_f), utterance length (T_u, N_w, N_wp, N_tp), pronoun ratio (R_pron, R_p2n), and response time (T_res). The speech hesitation (TR_s, TR_fs, TR_f, R_f) may reflect the fact that it takes longer to think about utterance content due to cognitive decline or to recall vocabulary due to memory decline. As for the utterance length (T_u, N_w, N_wp, N_tp), it is thought to reflect the tendency to cut off the conversation early due to decreased motivation, which is a symptom of dementia. The pronoun ratio (R_pron, R_p2n) is thought to reflect the decline in memory, a symptom of dementia. MCI speakers may fail to recall nouns or take a long time to recall them due to poor memory, so they may omit nouns that should

have been mentioned or convert them to pronouns. The response time (T_res) is thought to reflect the time required to think about utterance content due to cognitive decline.

As shown above, the features effective for identification of MCI in the wide range of topics were occupied by features that were strongly related to the typical symptoms of dementia, such as memory loss and decreased motivation.

Conclusion

In this paper, we construct a machine learning model to identify dementia by extracting linguistic and time features from the elderly corpus, with the aim of realizing noninvasive screening for early detection of dementia. The identification performance of the three methods, (I), (II), and (III), was compared by AUC on the single topic model and the general topic model. The AUC of the model constructed using RF in (III) for the single topic was 0.91, showing a higher AUC compared to previous study (Shibata et al. 2019). Topic analysis showed that topics with high similarity in utterance content were effective in identifying MCI. In addition, it was found that for content description tasks such as STORY and ANIME, the more concrete the content, the more effective the identification of MCI. In general topic, the identification performance for unknown topics was verified, and the results showed a high AUC of 0.8, indicating that the general topic model developed in this study can be applied to general conversation.

For future issue, it is necessary to construct a conversation system equipped with a general topic model and to verify whether it can identify MCI from daily conversations with actual elderly people.

Impact of GenAI on Social and Individual Well-being

For the conversation system to detect dementia mentioned as a future issue, it is possible to use GenAI such as LLM as a daily conversation module to collect daily conversations. Then, the collected conversation data could be used to build an identification model for dementia, which could lead to the early detection of dementia. Considering that the general topic model showed high identification performance with 600 data points, high identification performance can be expected if the system becomes widely used and the number of data points available for model construction increases. Early detection by such a system is very important not only for the patients themselves, but also for their families and caregivers, and for the well-being of society.

On the other hand, there are privacy concerns in implementing the system. As observed in the episode description task used in this study, everyday conversations often contain personal information. There is a risk of personal information being leaked by being entered into LLM. In addition, collecting and analyzing conversation data may lead to an invasion of privacy.

In summary, a GenAI-based dementia detection system from conversations has the potential to have a positive impact on social and individual well-being, but its implementation requires careful consideration and appropriate ethical considerations.

References

Banhong, S.; and Okazaki, N. 2019. Katari ni motozuku nintichisho hantei(in Japanese). In Proceedings of the 25 Annual Meeting of the Association for Natural Language Processing, 501–504.

Cabinet Office in Japan. 2017. Annual Report on the Aging Society: 2017. NIKKEI PRINTING, Inc.

Ishihara, S.; Iribe, Y.; and Kitaoka, N. 2020. Dementia Detection from Chat Dialogue using Vocabulary and Dependency Structure. In Proceedings of the 82th National Convention of IPSJ, 459–460.

Kudo, T.; and Matsumoto, Y. 2002. Japanese Dependency Analysis Using Cascaded Chunking. IPSJ Journal, 43(6): 1834–1842.

Kudo, T.; Yamamoto, K.; and Matsumoto, Y. 2004. Applying conditional random fields to Japanese morphological analysis. In Proceedings of the 2004 conference on empirical methods in natural language processing, 230–237.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12: 2825–2830.

Shibata, D.; Wakamiya, S.; Kinoshita, A.; and Aramaki, E. 2016. Detecting Japanese patients with Alzheimer 's disease based on word category frequencies. In Proceedings of the Clinical Natural Language Processing Workshop (Clinical-NLP), 78–85.

Shibata, D.; Ito, K.; Shoko, W.; and Aramaki, E. 2019. Detecting Early Stage Demantia based on Natural Language Processing. Transactions of the Japanese Society for Artificial Intelligence, 34(4): B–J11_1.

Sluis, R. A.; Angus, D.; Wiles, J.; Back, A.; Gibson, T.; Liddle, J.; Worthy, P.; Copland, D.; and Angwin, A. J. 2020. An automated approach to examining pausing in the speech of people with dementia. American Journal of Alzheimer's Disease & Other Dementias®, 35: 1533317520939773.

Yoshii, K.; Kimura, D.; Kosugi, Y.; Arakawa, K.; Takase, T.; Kobayashi, M.; Yamada, Y.; Nemoto, M.; Watanabe, R.; Tsukada, E.; Ohta, S.; Higashi, S.; Nemoto, K.; Arai, T.; and Nishimura, M. 2020. Hitogata robot tono nitijou kaiwa onsei wo motiita nintishou kani screening no tameno kisoteki kentou(in Japanese). The Special Interest Group Technical Reports of IPSJ SIG-SLP, 2020(7): 1–4.