Detection of dementia from responses to atypical questions asked by embodied conversational agents

Tsuyoki Ujiro1, Hiroki Tanaka1 Hiroyoshi Adachi2 Hiroaki Kazui3 Manabu Ikeda4 Takashi Kudo2 Satoshi Nakamura1

1Graduate School of Information Science, Nara Institute of Science and Technology, Japan
2Health and Counseling Center, Osaka University, Japan
3Department of Neuropsychiatry, Kochi Medical School, Japan
4Graduate School of Medicine, Osaka University, Japan

ujiro.tsuyoki.uq5@is.naist.jp

Abstract

Detection of dementia requires examinations, such as blood tests and functional magnetic resonance imaging (fMRI), that can be very stressful for the patient. Previous studies proposed screenings for easy detection of dementia that utilized acoustic and language information derived from conversations between patients and medical staff. Although these studies demonstrated effectiveness in automatically detecting dementia, the tasks used were created based on neuropsychological tests. The effect of habituation on this limited variety of tasks might have a negative impact on routine dementia screening. We propose a method to detect dementia using responses to more atypical questions asked by embodied conversational agents. Through consultations with neuropsychologists, we created a total of 13 questions. The embodied conversational agent obtained answers to these questions from 24 participants (12 dementia and 12 non-dementia). We recorded their responses and extracted speech and language features. We classified the two groups (dementia/non-dementia) by a machine learning algorithm (support vector machines and logistic regression) using the extracted features. The results showed a 0.95 detection performance in the area under the curve of the receiver operating characteristic (AUROC). This result demonstrates that our system using atypical questions can detect dementia.

Index Terms: Dementia, atypical questions, speech feature, language feature, embodied conversational agents

1. Introduction

According to a survey by the Ministry of Health, Labour and Welfare, Japan’s aging rate is 27.5% and the number of elderly is increasing. Japan’s growing elderly population has led to an increase in the number of dementia patients and is considered a serious social issue. Dementia causes deterioration in memory, thinking, and behavior, thus significantly interfering with the ability to function at work or perform ordinary daily activities [1]. There are various types of dementia such as Alzheimer’s disease (AD), normal pressure hydrocephalus (NPH), and dementia with Lewy Bodies.

Early diagnosis of dementia is obviously critical for patients and families to plan for the future and identify outside sources of assistance. In addition, as potentially useful and proven treatments become available, early diagnosis will become increasingly important [2].

It is challenging to detect dementia in its early stages [3]. Many medical institutions diagnose dementia based on the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV-TR [4] or DSM-V [5]. Early detection combines neuropsychological examinations, blood tests, and brain image scans such as functional magnetic resonance imaging (fMRI) [6]. These examinations are invasive, adding to the patient’s psychological burden by inducing anxiety and stress. In addition, considerable time and cost are required to complete this series of tests.

Therefore, a non-invasive and easy detection method is required. Some methods for the non-invasive and easy detection of dementia have been proposed. These studies use speech and language information to detect or diagnose dementia [1], [7], [8]. Most analyze utterances about photographs spoken during neuropsychological examinations (e.g., in conversations with medical staff).

A detection method during interaction with agents has also been proposed [9]. Previous studies attempted to detect dementia using the patient’s voice as well as language and image information derived from three questions taken from a neuropsychological examination. However, these exam questions might be remembered by elderly people. To detect early stage dementia, it is necessary to monitor the elderly over a long period.

On the other hand, it is well known that dementia is associated with memory disorders [5]. Studies on memory-related to dementia have been conducted [10], [11], [12], [13] showing, for example, that patients with dementia are likely to remember long-term events, but forget more recent occurrences.

In this paper, we propose a new approach to detect dementia using embodied conversational agents and an atypical question set that addresses several degrees of memory.

2. Related Work

Some previous studies effectively used speech and language information to detect dementia, as summarized below.

Roark et al. [8] recorded responses during the Wechsler memory recall task [14] and classified the speech into non-dementia and mild cognitive impairment (MCI) using complexity of the syntactic tree and speech features. The results showed that using multiple complementary measures outperformed the single value of a neuropsychological examination, proving that linguistic and speech information are useful for early detection of dementia.

In addition, Aramaki et al. [7] examined the relationship between cognitive ability and language ability for suspected MCI. The Hierarchic Dementia Scale-Revised (HDS-R) [15] was used as a marker to identify people with MCI. They directed the participants to write and speak, and measured their language ability. Analysis of the writing showed no significant
difference between the non-dementia group and people with MCI. However, the speech analysis did show that the vocabulary size was greater in those with suspected MCI than the non-dementia participants.

Another prior study attempted to detect dementia through dialogues with a computer agent. Tanaka et al. [9] prepared questions by referring to the Mini-Mental State Examination (MMSE) [16] and other neuropsychological examinations in an attempt to detect dementia. Language features, speech features, and facial expressions were extracted from the responses to these questions to classify non-dementia and dementia. As a result, the area under the curve of the receiver operating characteristic (AUROC) was 0.94 when using the linear kernel in a support vector machine (SVM).

Bahman et al.’s study [17] also prepared questions in the same form as a questionnaire created by a psychiatrist to detect dementia from conversations with an agent. Using speech features and linguistic features, they classified between neurodegenerative dementia and functional memory disorder with about 91% accuracy, thus demonstrating the effectiveness of detecting dementia from responses to agent questions.

In this paper, we propose a method to detect dementia from the responses of elderly people to atypical questions using an agent system of spoken dialogue.

3. Data Collection

3.1. Embodied conversational agents

We used the agent system created by Tanaka et al. [9] for data acquisition using a regular laptop, the Microsoft Surface Pro 3. This system uses the MMDAgent as an embodied conversational agent to decrease the agent’s speech rate and subtitle the content of each utterance, so that the elderly (user) can easily understand it. In addition, six kinds of tasks were prepared to detect dementia (see Figure 1). The summary of these tasks is as follows:

(a) Self-introduction: The agent introduces herself and asks the user’s name and age.
(b) Gaze tracking: Display a small black dot on the screen and move it. The user is instructed to follow the point.
(c) Reading task: Display sentences on the screen and instruct the user to read out loud. The text is derived from the WMS-R [14].
(d) Fixed questions: Ask the user three questions prepared based on a neuropsychological examination such as the MMSE [16].
(e) Random questions: Ask the user five randomly selected questions from 13 prepared questions.
(f) Retelling: The agent reads out a sentence used in the WMS-R [14] and directs the user to remember it. After reading aloud, the user is asked to recall the sentence.

3.2. Atypical question

Our focus is on the random questions provided by an agent system. We prepared 13 types of atypical questions rather than questions derived from a neuropsychological examination. We prepared these questions based on consultations with neuropsychologists. The prepared question set is shown in Table 1. The agent randomly chooses and asks the elderly participants five questions from these 13 questions. During the conversation with the agent, we recorded the user/participant’s voice and video with a microphone and built-in camera. If the user/participant was silent for more than 15 seconds, the system automatically shifted to the next question.

4. Detecting Dementia from Responses to Atypical Questions

4.1. Participants

We recruited a total of 24 participants, whose demographics are shown in Table 2. Each participant agreed to participate in this experiment with sufficient explanation in advance. We obtained informed consent from all participants.

Twelve patients took part in the recording as the dementia group at the Osaka University Medical School Hospital and the remaining 12 served as the non-dementia group at the Nara Institute of Science and Technology. The detailed diagnosis of the dementia patients was nine Alzheimer’s disease (AD), one normal pressure hydrocephalus (NPH), one MCI, and one AD+NPH. Per the existing criteria, the dementia group was diagnosed by the psychiatrists of the Osaka University Medical School affiliated hospital based on DSM-IV-TR [4]. We obtained the age, MMSE score, and educational history of all participants. To use the spoken dialogue agent, we confirmed that all participants completed the task.

4.2. Feature set

We transcribed the speech on the recorded data. From this, we extracted speech features and language features for the responses to five questions and calculated their mean value. Finally, we selected 21 features based on previous studies [9].

Table 1: Prepared random question set.

<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
</tr>
<tr>
<td>Q2</td>
</tr>
<tr>
<td>Q3</td>
</tr>
<tr>
<td>Q4</td>
</tr>
<tr>
<td>Q5</td>
</tr>
<tr>
<td>Q6</td>
</tr>
<tr>
<td>Q7</td>
</tr>
<tr>
<td>Q8</td>
</tr>
<tr>
<td>Q9</td>
</tr>
<tr>
<td>Q10</td>
</tr>
<tr>
<td>Q11</td>
</tr>
<tr>
<td>Q12</td>
</tr>
<tr>
<td>Q13</td>
</tr>
</tbody>
</table>

Table 2: Age of participants, MMSE score, and educational history (education in years) with mean (M) and standard deviation (SD) values.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Age (M, SD)</th>
<th>MMSE (M, SD)</th>
<th>Education (M, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-dementia</td>
<td>12</td>
<td>74.5 (4.3)</td>
<td>27.5 (1.8)</td>
<td>8.8 (2.6)</td>
</tr>
<tr>
<td>Dementia</td>
<td>12</td>
<td>75.9 (7.3)</td>
<td>21.2 (5.1)</td>
<td>13.9 (3.8)</td>
</tr>
</tbody>
</table>

1http://www.mmdagent.jp/
4.2.1. Speech features

We used the Snack Sound Toolkit to extract the speech features related to Gap, Pause, fundamental frequency (F0), power, and answer time. Gap is the time difference from the end of the agent’s question to the start of the participant’s response. Pause is the number of times which exceed one second between the participants’ utterances. We also calculated the longest and mean values of Pause (longest value of Pause and mean value of Pause). Answer time is the total time of participant’s responses.

For F0, we calculated the coefficient of variation, mean, maximum, median, minimum, standard deviation, and range values. For power, we calculated the mean and standard deviation values.

4.2.2. Language features

From the transcription, we performed a Japanese part of speech analysis using the MeCab toolkit. Based on MeCab’s output, we calculated speech rate and the number of tokens (Tokens), fillers (Fillers), nouns (Nouns), verbs (Verbs), adjectives (Adj.), and adverbs (Adv.).

4.3. Classification model

We constructed a classification model using two machine learning algorithms from the extracted features. We built the logistic regression and linear kernel support vector machine (SVM) as classifiers. We normalized each feature to mean value of zero and standard deviation value of one, and trained a model to classify the dementia and non-dementia groups. For evaluation of the classification model, based on Roark et al.’s [8] and Tanaka et al.’s [9] studies, we drew the ROC curves from leave-one-participant-out cross validation and calculated the area under the curve (AUC) of the ROC.

For post-hoc analysis, we also examined the weight of each tuning the model’s parameters and one for the test.

4.4. Results

Figure 2 shows the ROC curve results of the two models. We obtained an AUC of 0.95 (the linear support vector machine) and 0.92 (the logistic regression), respectively. These results show that the detection performance is higher than the previous study [9]. We also performed an experiment with only AD patients, which means that we removed three patients from the 12 people with dementia, and obtained an AUC of 0.89.

For post-hoc analysis, we also examined the weight of each
Table 3: AUC and unweighted accuracies of input variations.

<table>
<thead>
<tr>
<th>Input (model)</th>
<th>AUC</th>
<th>Unweighted accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap only (SVM)</td>
<td>0.69</td>
<td>0.63</td>
</tr>
<tr>
<td>MMSE scores (SVM)</td>
<td>0.85</td>
<td>0.83</td>
</tr>
<tr>
<td>Speech features (SVM)</td>
<td>0.85</td>
<td>0.83</td>
</tr>
<tr>
<td>Speech features (Logistic)</td>
<td>0.90</td>
<td>0.83</td>
</tr>
<tr>
<td>All w/o MMSE (Logistic)</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>All w/o MMSE (SVM)</td>
<td>0.95</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Figure 3: Gap boxplots of non-dementia and dementia patients.

Figure 4: Gap boxplots of non-dementia and dementia patients for each question.

In this paper, we proposed a method to detect dementia from responses to atypical questions using embodied conversational agents. The effect of this method was experimentally tested with 24 dementia patients and non-dementia participants. We obtained a detection accuracy of 0.95 in the AUROC curve. We demonstrated that dementia can be detected even when using atypical questions.

We plan to use this system to automate early dementia detection. However, we need to address several problems in order to put this system to practical use. First, we need to automate the transcription process; we thus plan to use automatic speech recognition (ASR). Second, we did not include features that can be extracted from video. There are a few studies on detecting dementia using facial expressions and eye movements [18]. We will therefore consider extracting a variety of features from video. Third, we raised the possibility of measuring the degree of dementia from range of memory. To confirm this, we need to prepare a question set that considers varied degrees of event terms and measure the relationship between the degree of dementia and terms of the questions. Fourth, we need to consider effects of educational history, part of speech, and difference of five selected questions in the future.

5. Conclusions

In this paper, we proposed a method to detect dementia from responses to atypical questions using embodied conversational agents. The effect of this method was experimentally tested with 24 dementia patients and non-dementia participants. We obtained a detection accuracy of 0.95 in the AUROC curve. We demonstrated that dementia can be detected even when using atypical questions.

We plan to use this system to automate early dementia detection. However, we need to address several problems in order to put this system to practical use. First, we need to automate the transcription process; we thus plan to use automatic speech recognition (ASR). Second, we did not include features that can be extracted from video. There are a few studies on detecting dementia using facial expressions and eye movements [18]. We will therefore consider extracting a variety of features from video. Third, we raised the possibility of measuring the degree of dementia from range of memory. To confirm this, we need to prepare a question set that considers varied degrees of event terms and measure the relationship between the degree of dementia and terms of the questions. Fourth, we need to consider effects of educational history, part of speech, and difference of five selected questions in the future.

6. Acknowledgements

Part of this work was supported by the Collaborative Research Project with Suntory Global Innovation Center Ltd. as well as JSPS KAKENHI Grant Numbers JP17H06101 and JP18K11437.

7. References


