

## Article

# Construction and Evaluation of QOL Specialized Dictionary SqolDic Using Vocabulary Meaning and QOL Scale

Satoshi Nakagawa <sup>1,\*</sup> , Huang Minlie <sup>2</sup> and Yasuo Kuniyoshi <sup>3</sup> 

<sup>1</sup> Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan

<sup>2</sup> Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; aihuang@tsinghua.edu.cn

<sup>3</sup> Next Generation AI Research Center & Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan; kuniyosh@isi.imi.i.u-tokyo.ac.jp

\* Correspondence: nakagawa@isi.imi.i.u-tokyo.ac.jp; Tel.: +81-3-5841-6282

**Abstract:** Agents that build interactive relationships with people can provide appropriate support and generate behaviors by accurately grasping the state of the person. This study focuses on the quality of life (QOL), which can be assessed multidimensionally, and aims to estimate QOL scores in the process of human interaction. Although vision-based estimation has been the main method for QOL estimation, we proposed a new text-based estimation method. We created a QOL-specific dictionary called SqolDic, which is based on large-scale Japanese textual data. To evaluate the effectiveness of SqolDic, we implemented a system that outputs the time-series variation of a user's conversation content and the QOL scores based on it. In an experiment for estimating the content of user conversations based on a QOL scale by inputting data from actual human conversations, we achieved a maximum estimation accuracy of 91.2%. Additionally, in an experiment to estimate QOL score variability, we successfully estimated the mental health state and one of the QOL scales with a smaller distribution of error than that in previous studies. The experimental results demonstrated the effectiveness of our system in estimating conversation content and QOL scores as well as the effectiveness of our newly proposed QOL dictionary.

**Keywords:** quality of life; state estimation; human-agent interaction



**Citation:** Nakagawa, S.; Huang, M.; Kuniyoshi, Y. Construction and Evaluation of QOL Specialized Dictionary SqolDic Using Vocabulary Meaning and QOL Scale. *Electronics* **2021**, *10*, 417. <https://doi.org/10.3390/electronics10040417>

Academic Editors: Juan M. Corchado; Stefanos Kollias and Javid Taheri  
Received: 31 December 2020  
Accepted: 4 February 2021  
Published: 8 February 2021

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



**Copyright:** © 2021 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

Building long-term relationships between humans and agents is one of the major subjects in human-agent interaction (HAI). However, it is not as easy to build long-term relationships in HAIs as it is among humans.

In the field of welfare for the elderly, it has been reported that continuous interaction with robots makes the elderly more active and improves users' communication, self-care, and social life [1,2]. Therefore, interactive robots for the elderly are expected to become widely used as one of the effective means to improve the well-being of the elderly. As Japan has already entered a super-aging societal phase, their elderly welfare facilities are no exception to this trend, and they are increasingly promoting the use of ICT. One of the trends in the welfare of the elderly is the introduction of communication robots such as Paro [3] and Palro [4]. However, while the number of adoptions has been increasing, there are still some elderly welfare facilities that have not been able to successfully establish a sustainable relationship with the robots. This is thought to be because the relationship between humans and robots lacks development over time, and humans are entering the burnout phase, where they are not interested in agents [5]. There are also studies that points out other causes such as low intelligence, little variety of reactions in the robot and the lack of established methods to keep people interested in the robot [6,7], and designs robots or agents to keep users interested in them and build relationships with them [8–10]. As a solution to this problem, we intend to create a friendly and continuous relationship

between a person and an agent by letting the agent generate personalized behaviors for the user.

It has been proposed that emotional properties are given to agents' speech, and it has been shown that human beings can correctly perceive the given emotions [11]. However, to achieve affable and empathetic communication for users, it is necessary for the agents to grasp the user's state in real time and to generate appropriate words and actions based on the estimated state. Many studies have implemented estimation systems that deal with human state estimation [12,13]. Research on emotion estimation, in particular, has been ongoing for a long time. However, emotion recognition only shows human emotions at a specific time, and more profound information about human's physical, mental, and social aspects cannot be extracted. The quality of life (QOL) originated in a study that was featured in a clinical evaluation by Karnofsky in the late 1940s [14], is a multidimensional indicator of the human condition, which includes not only physical but also mental and social aspects. In this study, we refer to the QOL index of the SF-36v2 health survey [15]. According to this index, QOL consists of eight subscales: Physical Functioning (PF), Role Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role Emotional (RE), and Mental Health (MH). In the conventional method, scores on these eight QOL scales were calculated through interviews by answering the SF-36v2, which is a selective questionnaire. This indicator has been shown to be effective in estimating depression with respect to mental health QOL scale [16]. However, a problem has been pointed out in that these methods impose a burden on those whose QOL is being measured [5]. Additionally, this question-and-answer format may create a hierarchical relationship between the questioner and the respondents. Furthermore, because only the QOL at the time of measurement can be measured in one examination, repeated examinations are required for the continuous measurement of QOL, which is inefficient and cannot be used to immediately track changes in QOL scores.

In this study, we hypothesized that it is possible to estimate the QOL by extracting information regarding QOL from users' conversations, and proposed a text-based QOL estimation method for human communication processes. By understanding which scales of QOL are related to what users are talking about from textual information and by analyzing the features extracted from the text, it is possible to estimate the score of the relevant QOL scale. Therefore, we created a QOL dictionary called SqolDic, which is based on large-scale Japanese textual data, and implemented a system to output QOL scores based on the dictionary. We demonstrated the effectiveness of the system through verification using data collected through interpersonal experiments.

## 2. Related Works

### 2.1. QOL Estimation in HAI

QOL is different from emotion recognition in terms of that it can evaluate a person's state from multiple perspectives. Due to the growing interest on QOL, many studies in the academic community have dealt with QOL, but many of them consider the impact of their proposed method on QOL [17,18]. In the field of HAI, research studies have aimed at estimating QOL itself through interaction with agents first emerged in 2018 [19]. In this QOL estimation study, an end-to-end estimation model was proposed to estimate QOL scores from facial expressions and prosodic information. In April of 2020, it was suggested that gaze patterns will be effective for QOL estimation [20], and in September of 2020, a multi-feature vector learning-based QOL score estimation system that integrates gaze patterns with visual information on head movements and facial expressions using Open-Pose [21], one of the widely used methods for estimating human posture and recognizing facial expressions, was proposed and shown to be effective [5].

In the field of the state estimation, emotion recognition involves vision-based [22,23], audio-based [24–26], text-based [27–30], electroencephalogram (EEG)-based [31–33], and multimodal approaches [34,35]. However, when it comes to estimation based on QOL scale,

these studies are approaches based on visual or prosodic information, and no text-based state estimation method has yet been proposed.

## 2.2. Text-Based Estimation of Human States

With respect to human state estimation in HAI, SimSensei [36] is another representative example of an interactive agent aimed at interactive state estimation using multimodal information. Although there is a growing trend to integrate interactions with communicative robots and agents into state estimation, the constraint of constant response generation patterns is one of the problems we need to solve. However, these related studies suggest that agents that build interactive relationships with people can also achieve more accurate state estimation by integrating information from multiple sensors. Therefore, we should establish an estimation method for QOL estimation systems based on not only visual information but also any modal information.

Text-based emotion estimation systems have been studied extensively because text information is significantly useful for human state estimation. Shivhare and Khethawat [37] proposed a method for emotion recognition from textual information by analyzing the intensity of each word extracted through morphological analysis, and Shaheen [38] proposed a method for emotion recognition by combining compositional analysis with syntactic and semantic analysis. For a text-based estimation of the mental state, Ren [39] proposed a system for predicting suicidal behavior by estimating changes in the mental state based on textual data from daily blogs. Furthermore, because real-time feelings are converted into text data on Twitter, a prediction model that considers social context and topical context has been constructed, and it is now possible to infer an individual's feelings from the content of tweets [29]. These studies can be combined with audio-based and visual-based depression estimation systems [40,41] to implement more robust estimation systems.

Additionally, Uchida [42] demonstrated that people tend to disclose more negative things to agents than to people. Therefore, it is possible that agents may receive additional information that would not be disclosed to a person. As a result, an agent might understand users by more accurately estimating their state based on their speech. Furthermore, it may be possible to generate personalized behaviors based on a more fully informed estimation. If a text-based estimation method is established, multimodal QOL estimation and a more accurate estimation of the human condition could become possible.

As for the agent's utterances, research on empathetic response generation has been carried out [43], and it has been emphasized in studies on counseling situations [44,45]. One of the contributions of our study is that the integration of a QOL grasping system with such a response generation system will enable agents to generate appropriate behaviors based on an accurate and pluralistic understanding of people's states.

## 3. Methods

To realize the QOL estimation and the generation of appropriate responses from the estimation results, it is necessary for an agent to extract and analyze useful information for QOL estimation from the user's text data during communication with the user. In this study, we created a QOL dictionary called SqolDic, and for its evaluation, we implemented a system for estimating the variability of QOL scores by focusing on the text information in the users' natural dialogues. The flow of our study is shown in Figure 1.

First, we created a dictionary called SqolDic, which is dedicated to QOL. Next, based on the dictionary, we implemented a text content classification system that outputs the strength of association between the user's utterances and eight QOL scales. We then implemented a system that automatically answers the questions in the QOL questionnaire SF-36v2 based on the positive-negative (PN) score, which is one of the features extracted from the text, after which we output eight QOL score variations via an existing scoring algorithm. Figure 2 shows the entire QOL score estimation system. Finally, we evaluated the effectiveness of the proposed system through validation using actual conversation data.

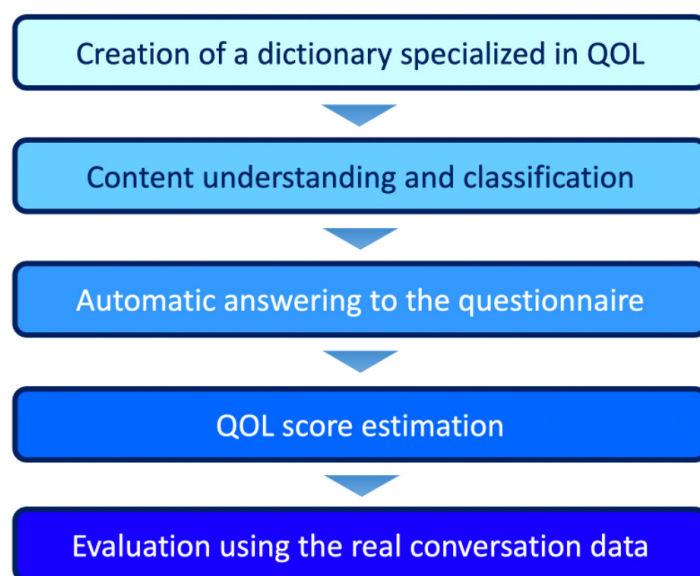


Figure 1. Flowchart of overall frame design.

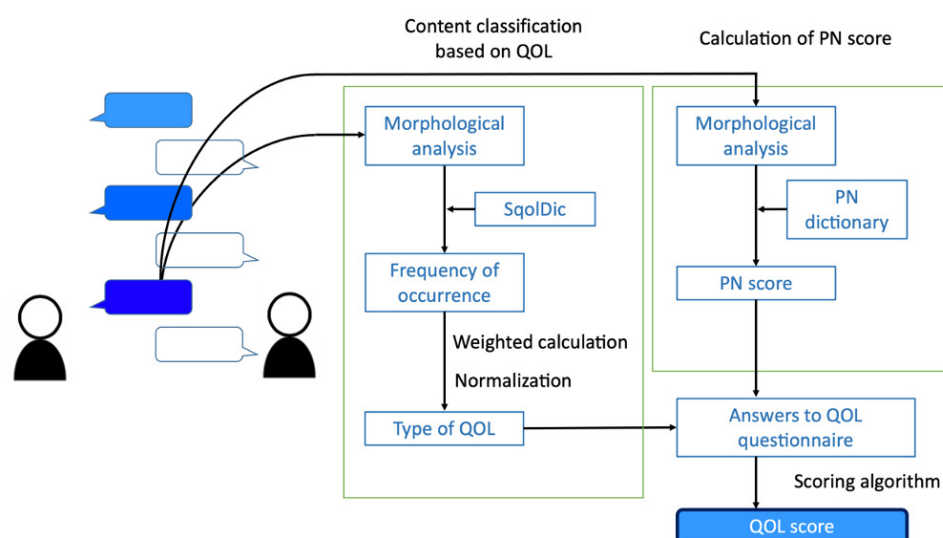


Figure 2. Text-based QOL estimation during communication. SqolDic is a QOL dictionary that we created based on large-scale textual data. PN score is a real number between  $-1$  and  $1$  that indicates whether the content is positive or negative for a given sentence.

### 3.1. QOL Dictionary SqolDic

We developed a specialized dictionary called SqolDic for QOL content classification and score estimation from users' speech. There are no other dictionaries that are specialized for QOL other than the one proposed in this study. In this section, we describe the procedure for creating SqolDic using Word2vec [46] and Mecab [47], which is a morphological analysis system. Word2vec, which was proposed by Tomas Mikolov, is a neural network-based vectorization method for word groups. The Japanese full-text data from Wikipedia were used as a dataset to create SqolDic. In this study, the data downloaded as of October 21, 2019, were used for subsequent operations.

SqolDic was created in three steps:

1. Create a vector model of all the meanings of all the lexical terms
2. Extract the QOL-focused terms
3. Accumulate the words with close vector distance to QOL-focused terms

As a first step, we created a data model using Gensim for the full text of Wikipedia data that were morphologically analyzed using Mecab. With the creation of this data model, a feature vector is assigned to each word. For all the remaining steps, based on the words contained in the SF-36v2 QOL questionnaire and the words contained in the description section regarding SF-36v2 on the SF-36v2 seller's website, other words and phrases strongly related to the word were searched. Table 1 shows a part of the database that was created by extracting words that form the basis of QOL. Each line is in the form of a "phrase: the number of the scale to which the phrase relates." The operations mentioned above resulted in the creation of SqolDic, a dictionary specializing in QOL, which is a collection of words and phrases related to each of the scales of QOL.

**Table 1.** Excerpt from SqolDic.

| Words       | QOL Scale               | Words       | QOL Scale             |
|-------------|-------------------------|-------------|-----------------------|
| activity    | 1: Physical Functioning | energy      | 5: Vitality           |
| walk        | 1: Physical Functioning | tired       | 5: Vitality           |
| ⋮           | ⋮                       | ⋮           | ⋮                     |
| housework   | 2: Role Physical        | family      | 6: Social Functioning |
| participate | 2: Role Physical        | friend      | 6: Social Functioning |
| ⋮           | ⋮                       | ⋮           | ⋮                     |
| backache    | 3: Bodily Pain          | childcare   | 7: Role Emotional     |
| swelling    | 3: Bodily Pain          | help        | 7: Role Emotional     |
| ⋮           | ⋮                       | ⋮           | ⋮                     |
| sick        | 4: General Health       | nervous     | 8: Mental Health      |
| worsen      | 4: General Health       | downhearted | 8: Mental Health      |
| ⋮           | ⋮                       | ⋮           | ⋮                     |

(Total 1416 words. The particles have been removed.)

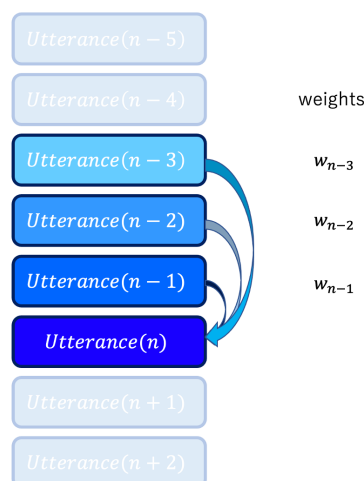
### 3.2. The Strength of the Connection between the Text and the Respective QOL Scale

In this section, we describe a system that classifies the content from the text data based on SqolDic and outputs the closeness of the relationship between the text data and the respective QOL scale. The following sentence is used as an example: "I went shopping with my friends today and had a great time." Because "friend" and "shopping" are words included in the social functioning QOL scale, and "fun" is a word included in the mental health QOL scale, this sentence can be judged as the most related to the social functioning scale, and the second most related to the mental health scale. Therefore, the input sentences are morphologically analyzed, and then, after referring to SqolDic, the strength of each scale is output. However, it is not practical to extract and understand the meaning of each conversation separately because the content of human conversation changes from moment to moment. This is because each sentence we emit is a component of a context, and its content changes continuously. In our experiment, if a person spoke about something related to a particular QOL scale until just before a particular statement, the content of that particular statement was recalled by the content of the previous statement, and there is always a connection between the two. Therefore, in our study, instead of focusing only on the target speech, we sought to output the strength of the connection between the text and the respective QOL scale by taking the previous speech into account.

As shown in Figure 3, we calculate the QOL relevance of the  $n$ th utterance  $Utterance(n)(U(n))$  through weighted calculation using the past three utterances. Each weight

was set to decrease as the distance from the target utterance increased. With the above settings, the QOL relevance of the  $n$ th utterance  $QSL(n)$  is expressed as Equation (1).

$$QSL(n) = \sum_{k=0}^3 w_{n-k} \times U(n-k). \quad (1)$$



**Figure 3.** Relationship between the  $n$ th utterance and the previous ones.

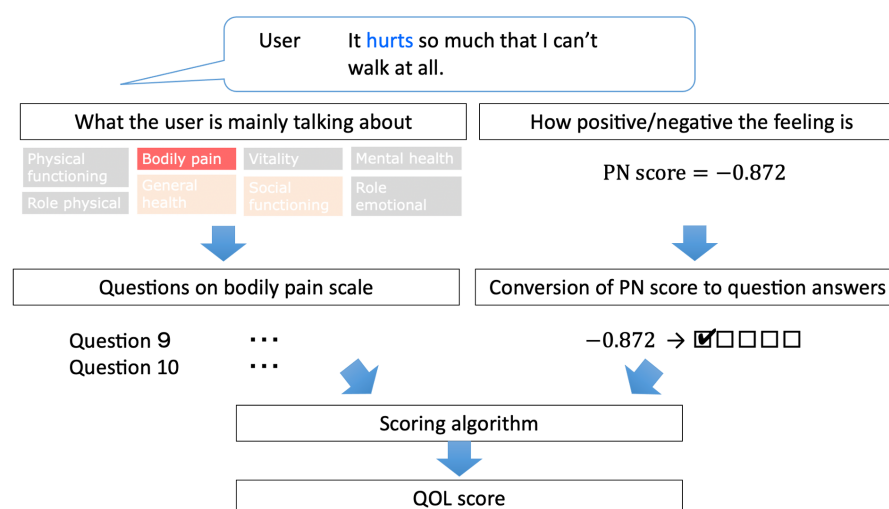
### 3.3. Estimation of Real-Time QOL Score Fluctuations

We implemented a system for answering the QOL questionnaire automatically by referring to the contents of the text and outputting the score. As an example, after determining that the target text is most relevant to the mental health QOL scale, we can assume that if the text is positive, the person's mental health is high, and if the content is negative, the person's mental health is considered to be low. Based on the hypothesis that this is the case, we used the PN score of the text to output QOL scores. The PN score is a real number between  $-1$  and  $1$  that indicates whether the content is positive or negative for a given sentence. The closer the PN score was to  $1$ , the more positive the content was, and vice versa. The PN values are generally used in the form of morphological analysis of the document to be analyzed for sentiment analysis by referring to the lists of words and their semantic orientation. This time, we used a list of semantic orientations of words created using a spin model [48].

According to the system described above, when a single sentence is entered along with its context, one of the eight QOL scales most relevant to the content is identified and output. At the same time, the PN score, which is a feature of the text, is calculated as shown in Figure 4. From the operations mentioned above, we implemented a system that automatically answers the questions on the QOL scale. The QOL questionnaire SF-36v2 was used in this study, and most of the questions had a format for selecting one option from  $1$  to  $5$ . Therefore, in converting PN scores (from  $-1$  to  $1$ ) into questionnaire options (from  $1$  to  $5$ ), we referred to the correspondence listed in Table 2.

The system we have described so far realizes automatic responses to the questionnaire on the corresponding QOL scale. We update the QOL score for each statement by outputting the QOL scores via the QOL scoring algorithm.





**Figure 4.** Overall flowchart of QOL estimation from utterances.

**Table 2.** Correspondence between PN score and answer number.

| PN Score              | Answer |
|-----------------------|--------|
| $-1 \leq x < -0.6$    | 1      |
| $-0.6 \leq x < -0.2$  | 2      |
| $-0.2 \leq x < 0.2$   | 3      |
| $0.2 \leq x < 0.6$    | 4      |
| $0.6 \leq x \leq 1.0$ | 5      |

#### 4. Interpersonal Experiment

To evaluate the implemented system, we conducted an interpersonal experiment to collect conversational text data to be input into the system. Four university students (age range: 20–23) participated in this experiment. Participants first received an explanation of the experiment, after which they reflected on the past week and answered the SF-36v2 QOL questionnaire. They then had a dialogue with two people regarding the events of the past week, such as what impressed them, and what they were thinking about. The subjects were not informed in advance about the estimation of QOL from their speech. After the dialogue, the verbatim transcripts were distributed to the participants and they labeled this verbatim data. The participants labeled each utterance with a QOL scale that they considered most relevant to it. We asked the participants to write all the scales in the same speech if they believed that the subject was speaking about related to more than one QOL scale. Conversely, they were allowed to write nothing if there was no QOL scale corresponding to the content of the speech.

#### 5. Experimental Results

##### 5.1. The Strength of the Connection between the Text and the Respective QOL Scale

Figure 5 shows the trend of the relevance of the QOL scale output from one subject's speech data. The horizontal axis corresponds to the utterances, the number represents the number of the utterances from the beginning, and the vertical axis represents the relevance of each of the QOL scale to the content of the utterance. To calculate the accuracy of the output QOL relevance, we checked the accuracy between the QOL scale output by the system as the most strongly related and the QOL scale that the subjects labeled on each sentence. For one of the subjects, 91.2% of the subjects' speeches matched the subject's opinion with the QOL scale that our proposed system judged most relevant to them. The accuracy calculated by analyzing the textual data of all subjects is listed in Table 3.

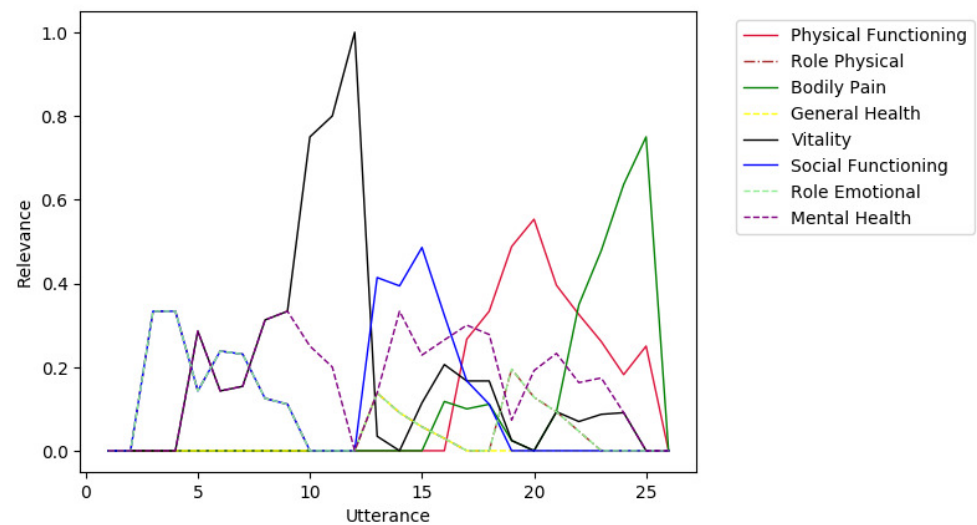


Figure 5. Trend of the relevance of the QOL scale.

Table 3. Accuracy of content classification based on QOL scale.

| User No. | 1     | 2     | 3     | 4     |
|----------|-------|-------|-------|-------|
| Accuracy | 91.3% | 73.0% | 63.8% | 78.0% |

### 5.2. Estimation of Real-Time QOL Score Fluctuations

Figure 6 shows the results of each of the eight QOL scale scores for each utterance from one of the subjects, to observe the real-time trends of each QOL scale score. For the analysis, we determined the average of the scores on each QOL scale derived from all the subjects' speech data. To evaluate the accuracy of the estimation of QOL scores from the text data, a graph showing the distribution of the error of each estimated score relative to the actual score is shown in Figure 7.

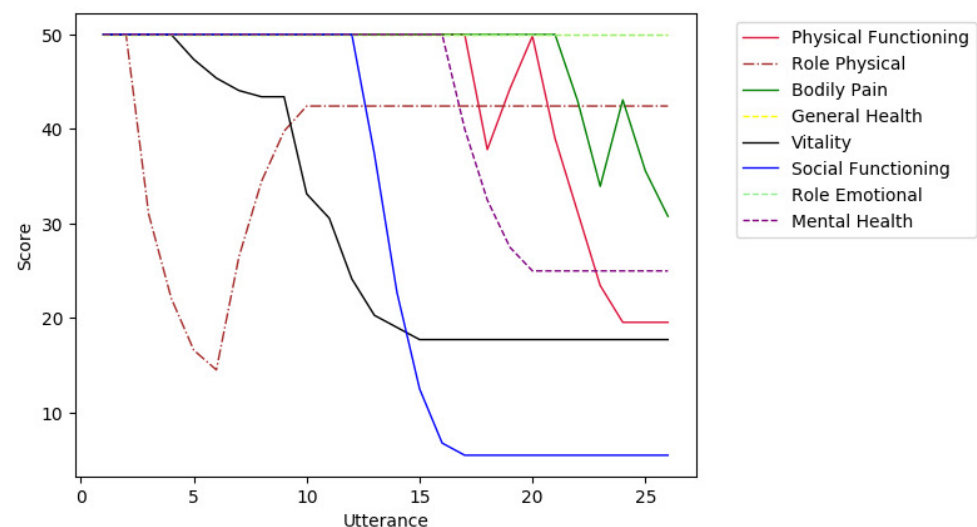
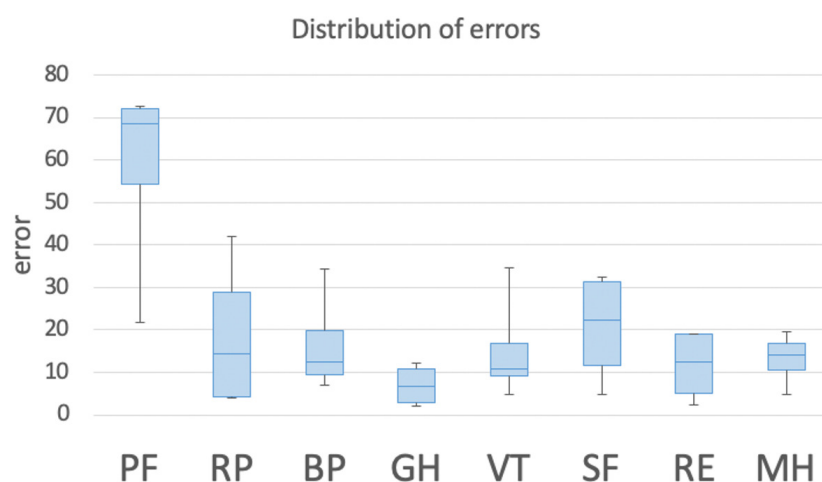


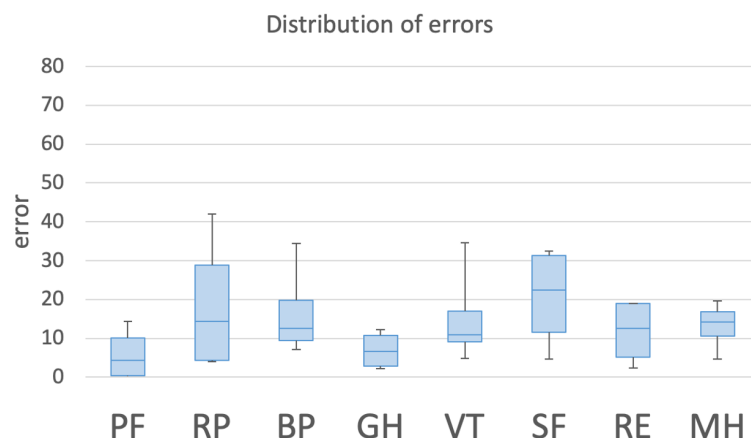
Figure 6. Trend of the QOL scores.





**Figure 7.** Distribution of errors between the estimated scores and the actual scores.

The horizontal axis represents the eight QOL scales, and the vertical axis represents the distribution of the error (absolute value) between the estimated and actual scores. The distribution of error in the estimation of the score of the physical functioning scale, which is on the leftmost side of the graph, is remarkably large. A possible solution to this problem is changing the mapping between the PN score, which is a feature obtained from the text data, and the questionnaire response options. As a result of updating the mapping, the distribution of the estimation error is shown using a graph in Figure 8.

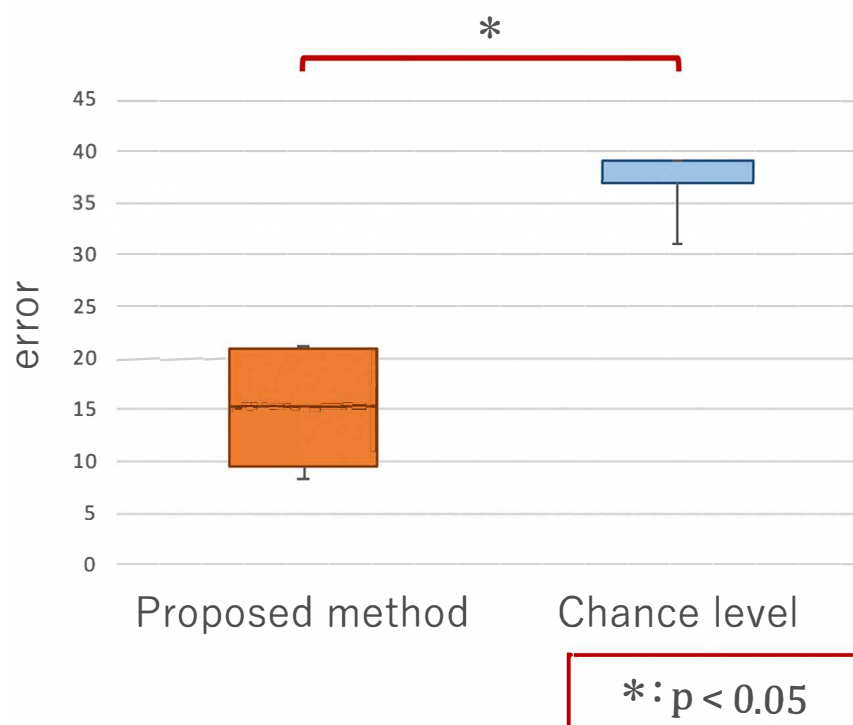


**Figure 8.** Distribution of errors between the estimated scores and the actual scores.

### 5.3. Evaluation of QOL Estimation Based on SqolDic

In this study, we applied parameter adjustments to the physical function QOL scale to reduce estimation errors. We compared it to the chance level to verify that parameter adjustments can improve the estimation function. Figure 9 compares the physical functioning scale score's error between the actual scores and the results of the proposed method, as well as the distribution of the error between the actual scores and the scores that would be output if all the responses to the questionnaire were in the middle choices as in the chance level. The results of the correspondence *t*-test showed a significant difference in the reduction of the error, indicating that the estimation accuracy of the proposed method was improved.

Furthermore, although previous studies have shown that estimating mental health QOL scores is the most difficult, the comparison of the results of the previous study (median: 11, first quartile: 5.6, third quartile: 26) to the QOL estimation system based on SqolDic (median: 14, first quartile: 11, third quartile: 17) showed that our proposed method reduced the distribution of error.



**Figure 9.** Comparison of errors between the proposed method and the chance level.

## 6. Discussion

In this study, we created SqolDic, which is a dictionary specific to QOL. To evaluate its effectiveness, we conducted two experiments; QOL content classification and QOL score estimation, using textual data collected from an interpersonal experiment.

In this study, a maximum accuracy of 91.2% was obtained for the content classification based on QOL. The realization of content classification means that we can understand the intentions of our conversation partners. This may help in estimating a person's state in the process of communication. Although we developed a system for the content classification of text information specifically for QOL estimation, the mapping between text contents and the eight scales that constitute QOL can be used for purposes other than QOL estimation. For example, we can estimate what topics a person is interested in and what topics he or she wants to talk about based on the content of the conversation, and this can be used as a response to help the dialogue agent decide what to talk about next.

The estimation of QOL scores based on textual data was also realized by using the correspondence between the PN scores and the questionnaire answers. In particular, the distribution of error in mental health score estimation was smaller than that in previous studies [19], which used only the time-series data of facial expressions extracted from video data. Therefore, text information may also be as useful as facial expressions for QOL estimation. With respect to the physical health scale, we also updated the mapping of PN scores to reduce the error. Additionally, a comparison with the case where the score was output without using the proposed method showed that the estimation error can be reduced by adjusting the parameters. However, we believe that we can further reduce the error by updating this correspondence using a trainer.

These results demonstrated the effectiveness of the proposed system in both QOL content classification and QOL score estimation. Therefore, it was shown that the proposed system can be used to extract high-dimensional information regarding QOL from textual data.

Because the effectiveness of the QOL content classification and the QOL score estimation was shown, we can say that the validity of the underlying QOL dictionary called SqolDic was also shown. The reason for this result is that we used a large amount of textual

data, and we were able to record contextually relevant words and phrases, which allows us to cover a wide range of topics. From the experimental results, we conclude that SqolDic can be applied to QOL estimation based on text information.

One of the limitations of our study is the number of subjects. In this study, real data were used to verify the effectiveness of SqolDic, the QOL content classification system, and the QOL score estimation system. In our study, the number of subjects did not affect the QOL content classification system because the system was based on SqolDic, which we created. As for the QOL score estimation system, we improved the accuracy of the estimation of the physical functioning scale in a form suitable for the participant data. Therefore, the model for estimating scores on the physical functioning scale is specific to the textual data of these subjects. Additionally, although the speakers in the collected dialogue database did not explicitly answer the questions in the QOL questionnaire during the dialogue, they were pre-selected as users who were aware that they were participating in a study on QOL through their completion of a QOL questionnaire. Therefore, it is necessary to validate the results in everyday conversations where people are not aware of any tasks. Based on the above discussion, the next task is to verify the generalization performance of the proposed method. To achieve this, it is necessary to collect a wide range of textual data and train appropriate parameters based on that data. If we can obtain appropriate correspondence for each scale, we can further improve the accuracy. Additionally, the accuracy of the system can be further improved by combining it with the QOL estimation system based on prosodic and visual information used in previous studies.

## 7. Conclusions

In this study, we created SqolDic as a QOL dictionary that is based on large-scale Japanese textual data. We also proposed a system that outputs the conversation content and QOL scores based on the QOL by extracting features from the user's conversation using SqolDic and demonstrated its effectiveness. Our future studies will involve improving the accuracy of QOL estimation by implementing a multimodal learning estimation model that integrates all the visual, auditory, and textual information that can be collected during the interaction with the user. This method solves the problems of the inefficiency of conventional QOL measurement methods and the difficulty of establishing a continuous relationship between humans and robots due to the uniformity of the relationship. It also enables robots to use their own functions to communicate more successfully with users. Estimating the user's QOL during daily interactions with the robot is expected to be a new application of robots for dynamic and efficient QOL measurement. In addition, HAI based on QOL estimation will enable intelligent robots to estimate the state and generate user-specific behaviors, and provide a clue to the question of how to improve the quality of interaction and build dynamic, better-informed relationships between humans and agents.

**Author Contributions:** Conceptualization, S.N. and Y.K.; methodology, S.N.; software, S.N.; validation, S.N.; formal analysis, S.N.; investigation, S.N.; resources, S.N. and Y.K.; data curation, S.N.; writing—original draft preparation, S.N.; writing—review and editing, S.N., H.M. and Y.K.; visualization, S.N., H.M. and Y.K.; supervision, H.M. and Y.K.; project administration, S.N.; funding acquisition, S.N. and Y.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the KDDI Foundation Research Grant Program 2019, the Graduate Program for Social ICT Global Creative Leaders (GCL) of The University of Tokyo by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Chair for Frontier AI Education in School of Information Science and Technology, and Next Generation AI Research Center, The University of Tokyo.

**Institutional Review Board Statement:** The study was approved by the Research Ethics Committee of the University of Tokyo (protocol code UT-IST-RE-171031-1 and date of approval Nov. 27th, 2017).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We thank Kana Naruse for useful discussions and for improving the content of the paper. We are also grateful to the reviewers for their comments and suggestions to improve the quality of the paper.

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

## References

- Wada, K.; Shibata, T.; Saito, T.; Sakamoto, K.; Tanie, K. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 2785–2790.
- Obayashi, K.; Kodate, N.; Masuyama, S. Socially assistive robots and their potential in enhancing older people's activity and social participation. *J. Am. Med. Dir. Assoc.* **2018**, *19*, 462–463. [\[CrossRef\]](#)
- Shibata, T. Research on interaction between human and seal robot, Paro. *J. Robot. Soc. Jpn.* **2011**, *29*, 31–34. [\[CrossRef\]](#)
- Ninomiya, H. Introduction of the communication robot Palro and efforts in robot town Sagami. *J. Robot. Soc. Jpn.* **2015**, *33*, 607–610. [\[CrossRef\]](#)
- Nakagawa, S.; Yonekura, S.; Kanazawa, H.; Nishikawa, S.; Kuniyoshi, Y. Estimation of Mental Health Quality of Life using Visual Information during Interaction with a Communication Agent. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 1321–1327.
- Kuwamura, K.; Nishio, S.; Sato, S. Can we talk through a robot as if face-to-face? Long-term fieldwork using teleoperated robot for seniors with Alzheimer's disease. *Front. Psychol.* **2006**, *7*, 1066. [\[CrossRef\]](#)
- Sabelli, A.M.; Kanda, T.; Hagita, N. A conversational robot in an elderly care center: An ethnographic study. In Proceedings of the 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Lausanne, Switzerland, 8–11 March 2011; pp. 37–44.
- Ren, F.; Bao, Y. A review on human-computer interaction and intelligent robots. *Int. J. Inf. Technol. Decis. Mak.* **2020**, *19*, 5–47. [\[CrossRef\]](#)
- Tanaka, F.; Cicourel, A.; Movellan, J.R. Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17954–17958. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jimenez, F.; Yoshikawa, T.; Furuhashi, T.; Kanoh, M. Effects of a novel sympathy-expression method on collaborative learning among junior high school students and robots. *J. Robot. Mechatron.* **2018**, *30*, 282–291. [\[CrossRef\]](#)
- Hu, T.; Xu, A.; Liu, Z.; You, Q.; Guo, Y.; Sinha, V.; Luo, J.; Akkiraju, R. Touch your heart: A tone-aware chatbot for customer care on social media. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.
- Li, S.; Deng, W. Deep facial expression recognition: A survey. *arXiv* **2018**, arXiv:1804.08348.
- El Ayadi, M.; Kamel, M.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [\[CrossRef\]](#)
- Karnofsky, D.A. The clinical evaluation of chemotherapeutic agents in cancer. In *Evaluation of Chemotherapeutic Agents*; Macleod, C.M., Ed.; Columbia University Press: New York, NY, USA, 1949; pp. 191–205.
- Fukuhara, S.; Bito, S.; Green, J.; Hsiao, A.; Kurokawa, K. Translation, adaptation, and validation of the sf-36 health survey for use in Japan. *J. Clin. Epidemiol.* **1998**, *51*, 1037–1044. [\[CrossRef\]](#)
- Ware, J.E., Jr.; Gandek, B. Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *J. Clin. Epidemiol.* **1998**, *51*, 903–912. [\[CrossRef\]](#)
- Wallerstedt, A. Quality of life after open radical prostatectomy compared with robot-assisted radical prostatectomy. *Eur. Urol. Focus* **2019**, *5*, 389–398. [\[CrossRef\]](#)
- Karefard, A.; Nordgren, L. Effects of dog assisted intervention on quality of life in nursing home residents with dementia. *Scand. J. Occup. Ther.* **2019**, *26*, 433–440. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nakagawa, S.; Enomoto, D.; Yonekura, S.; Kanazawa, H.; Kuniyoshi, Y. A telecare system that estimates quality of life through communication. In Proceedings of the 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, 23–25 November 2018; pp. 325–330; doi:10.1109/CCIS.2018.8691360. [\[CrossRef\]](#)
- Nakagawa, S.; Enomoto, D.; Yonekura, S.; Kanazawa, H.; Kuniyoshi, Y. New telecare approach based on 3D convolutional neural network for estimating quality of life. *Neurocomputing* **2020**, *397*, 464–476. [\[CrossRef\]](#)
- Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.
- Chang, Y.; Vieira, M.; Turk, M.; Velho, L. Automatic 3D facial expression analysis in videos. In *International Workshop on Analysis and Modeling of Faces and Gestures*; Zhao, W., Gong, S., Tang, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 293–307.
- Wang, J.; Yin, L.; Wei, X.; Sun, Y. 3D facial expression recognition based on primitive surface feature distribution. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 1399–1406; doi:10.1109/CVPR.2006.14. [\[CrossRef\]](#)

24. Scherer, K.R. Vocal communication of emotion: A review of research paradigms. *Speech Commun.* **2003**, *40*, 227–256. [\[CrossRef\]](#)
25. Kwon, O.W.; Chan, K.; Hao, J.; Lee, T.W. Emotion recognition by speech signals. In Proceedings of the Eighth European Conference on Speech Communication and Technology, Geneva, Switzerland, 1–4 September 2003; International Speech Communication Association: Baixas, France, 2003; pp. 125–128.
26. Vidrascu, L.; Devillers, L. Detection of real-life emotions in call centers. In Proceedings of the Ninth European Conference on Speech Communication and Technology, Lisbon, Portugal, 4–8 September 2005; International Speech Communication Association: Baixas, France, 2005; pp. 1841–1844.
27. Ren, F.; Deng, J. Background knowledge-based multi-stream neural network for text classification. *Appl. Sci.* **2018**, *8*, 2472. [\[CrossRef\]](#)
28. Ren, F.; Matsumoto, K. Semi-automatic creation of youth slang corpus and its application to affective computing. *IEEE Trans. Affect. Comput.* **2016**, *7*, 176–189. [\[CrossRef\]](#)
29. Ren, F.; Wu, Y. Predicting user-topic opinions in Twitter with social and topical context. *IEEE Trans. Affect. Comput.* **2013**, *4*, 412–424. [\[CrossRef\]](#)
30. Quan, C.; Ren, F. Sentence emotion analysis and recognition based on emotion words using Ren-CECps. *Int. J. Adv. Intell.* **2010**, *2*, 105–117.
31. Lin, Y.P.; Wang, C.H.; Jung, T.P.; Wu, T.L.; Jeng, S.K.; Duann, J.R.; Chen, J.H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1798–1806. [\[CrossRef\]](#)
32. Jenke, R.; Peer, A.; Buss, M. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2014**, *5*, 327–339. [\[CrossRef\]](#)
33. Ren, F.; Dong, Y.; Wang, W. Emotion recognition based on physiological signals using brain asymmetry index and echo state network. *Neural Comput. Appl.* **2019**, *31*, 4491–4501. [\[CrossRef\]](#)
34. Banziger, T.; Grandjean, D.; Scherer, K.R. Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion* **2009**, *9*, 691–704. [\[CrossRef\]](#)
35. Fan, Y.; Lu, X.; Li, D.; Liu, Y. Video-based emotion recognition using cnn-rnn and c3d hybrid networks. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; pp. 445–450.
36. DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; et al. SimsenSei Kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, Paris, France, 5–9 May 2014; International Foundation for Autonomous Agents and Multiagent Systems; pp. 1061–1068.
37. Shivhare, S.N.; Khethawat, S. Emotion detection from text. *arXiv* **2012**, arXiv:1205.4944.
38. Shaheen, S.; El-Hajji, W.; Hajj, H.; Elbassuoni, S. Emotion recognition from text based on automatically generated rules. In Proceedings of the 2014 IEEE International Conference on Data Mining Workshop, Shenzhen, China, 14 December 2014; pp. 383–392.
39. Ren, F.; Kang, X.; Quan, C. Examining accumulated emotional traits in suicide blogs with an emotion topic model. *IEEE J. Biomed. Health Inform.* **2015**, *20*, 1384–1396. [\[CrossRef\]](#)
40. Jain, V.; Crowley, J.L.; Dey, A.K.; Lux, A. Depression estimation using audiovisual features and fisher vector encoding. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, Orlando, FL, USA, 7 November 2014; pp. 87–91.
41. Valstar, M.; Gratch, J.; Schuller, B.; Ringeval, F.; Lalanne, D.; Torres, M.; Scherer, S.; Stratou, G.; Cowie, R.; Pantic, M. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In Proceedings of the 6th International Workshop on Audio/visual Emotion Challenge, Amsterdam, The Netherlands, 16 October 2016; pp. 3–10.
42. Uchida, T.; Takahashi, H.; Ban, M.; Shimaya, J.; Yoshikawa, Y.; Ishiguro, H. A robot counseling system—What kinds of topics do we prefer to disclose to robots? In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication, Lisbon, Portugal, 28–31 August 2017; pp. 207–212.
43. Rashkin, H.; Smith, E.M.; Li, M.; Boureau, Y.L. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv* **2008**, arXiv:1811.00207.
44. Perez-Rosas, V.; Mihalcea, R.; Resnicow, K.; Singh, S.; An, L. Understanding and predicting empathic behavior in counseling therapy. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 1426–1435.
45. Perez-Rosas, V.; Sun, X.; Li, C.; Wang, Y.; Resnicow, K.; Mihalcea, R. Analyzing the quality of counseling conversations: The tell-tale signs of high-quality counseling. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018.
46. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *arXiv* **2013**, arXiv:1310.4546.
47. Kudo, T.; Yamamoto, K.; Matsumoto, Y. Applying Conditional Random Fields to Japanese Morphological Analysis. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), Barcelona, Spain, 25–26 July 2004; pp. 230–237.
48. Takamura, H.; Inui, T.; Okumura, M. Extracting Semantic Orientations of Words using Spin Model. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005), Ann Arbor, MI, USA, 25–30 June 2005; pp. 133–140.