UMEED: VR GAME USING NLP MODELS AND LATENT SEMANTIC ANALYSIS FOR CONVERSATION THERAPY FOR PEOPLE WITH SPEECH DISORDERS

Hridyanshu

Department of Mathematics, University of California San Diego, United States

ABSTRACT

UmeedVR aims to create a conversational therapy VR game using natural language processing for patients with Speech Disorders like Autism or Aphasia. This study developed 5 psychological task sets and 3 environments via Maya and Unity. The Topic-Modeling AI, employing 25 live participants' recordings and 980+ TwineAI datasets, generated initial VR grading with a coherence score averaging 6.98 themes in 5-minute conversations across scenarios, forming a foundation for enhancements. Employing latent semantic analysis (gensim-corpus Python) and Term-Frequency-Inverse Document-Frequency (TF-IDF), grammatical errors and user-specific improvements were addressed. Results were visualized via audio-visual plots, highlighting conversation topics based on occurrence and interpretability. UMEED enhances cognitive and intuitive skills, elevating average topics from 6.98 to 13.56 in a 5-minute conversation with a 143.12 coherence score. LSA achieved 98.39% accuracy, topic modeling 100%. Significantly, real-time grammatical correction integration in the game was realized.

KEYWORDS

Virtual Reality, Topic Modeling & Coherence, Latent Semantic Analysis, Speech Disorders, Singular Value Decomposition

1. INTRODUCTION

Language impairment is one of the core features of psychological, mental, or speech disorders. Stroke has the highest disability-adjusted life-years lost in any disease, and approximately one-third of the patients get aphasia. Computers and tablets are innovative and aid in intensive treatments in speech rehabilitation for patients with aphasia. However, mechanical training limits the help of patients.

Post-acquired brain injury (ABI) one acquires aphasia that affects one’s speech, reading, writing, and gestures skills. ABI is a rapidly growing public health problem resulting from traumatic brain injury, stroke, hypoxic-ischemic encephalopathy after cardiac arrest, and brain tumours.

Approximately one-third of stroke patients experience aphasia. Patients with aphasia have a higher risk of not returning to work than those without aphasia. It is likely that an individual's inability to re-enter the workforce post stroke is due to the presence of aphasia. The incidence of stroke in younger patients was considerably lower than that in the older cohorts; however, it remains on the rise, and rehabilitation needs are worthy of attention.

David C. Wyld et al. (Eds): AI&FL, SCM, NLPTT, DSCC -2023
pp. 69-84, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131408

Between 2003–2004 and 2007, the frequency with which parents reported that their child had ever been diagnosed with Autism Spectrum Disorder increased from 5.5 to 11.6 per 1,000. By 2011–2012, it had risen to 20 per 1,000 or 2 percent. Note that the age ranges for children included in these surveys differed over time. These surveys do tell us that the rise of ASD in children is an increasingly important issue we need to combat in a cost-effective way so that the masses can use this.

1.1. Literature Work

Previous work has addressed conversation skills by focusing on different aspects, such as Joint attention which requires the user to attend to his or her virtual nonverbal behaviour to complete an interaction; turn-taking, or reciprocity in the conversation that occurs through collaborative virtual reality systems and with robots; and etiquette practice through a single-user virtual environment [1, 3]. These ideas were pragmatic and executed beautifully but one aspect that needed to be countered was the accessibility and ease of understanding for patients/users.

1.2. Objectives

1. Designing an intuitive Virtual Reality Game set within an imaginative realm, aimed at simulating diverse real-life conversations for individuals dealing with speech or cognitive challenges such as Autism (ASD) or post-stroke Aphasia.

2. Formulating a sophisticated latent semantic analysis model that evaluates the natural speech and mathematical expressions of users, displaying linear and contextual correlations. This model is pivotal for effectively evaluating patient communication.

3. Engineering AI-driven virtual entities within the game that engage in interactive dialogues with users. These AI avatars offer contextually relevant responses, enhancing the overall user experience.

4. Crafting exemplary conversational scenarios and meticulously assessing standard dialogues for integration into the game. This process ensures the provision of meaningful and effective communication practices.

5. Facilitating active user participation and achievement of predefined tasks through dynamic conversations. This approach promotes user engagement and skill enhancement across five distinct conversational dimensions.

2. METHODOLOGY AND EXPERIMENT DETAILS

The following formatting rules must be followed strictly. This (.doc) document may be used as a template for papers prepared using Microsoft Word. Papers not conforming to these requirements may not be published in the conference proceedings.

2.1. Ethics and Code

Firstly, Microsoft’s Responsible AI and Ethics Code were adopted throughout our testing. The guidelines laid down by the NGO for the timings, presence of instructors and overall data recording were also followed thoroughly. We achieved the same through the following ways: -

Inclusiveness: We made sure that our innovation is understandable to any student or patient or instructor who can speak English through the simple language instructions module in the game with interesting and vivid front end graphics work to also ensure a pictorial understanding.

Fairness: No bias was there when the experiments were conducted toward any student. Even a feedback form for the same was circulated for the same where anyone can anonymously express their opinion.

Privacy and Security: Duly signed forms were taken before and after the experimentation from the NGO and each person who participated in the experiment or conducted the experiment with me to allow their experience and work to be noted. No data or media of any sort was exposed to the mass public even in presentations.

Safety: There were always instructors, adults and college students involved in academia and research to supervise the experiments. Everything was duly timed and written on the board for each day.

2.2. Establishing the Criteria and Locations for Speech Therapy

There were five situations analysed that are critical for patients with speech disorders to improve on when discussed with NGO instructors and care takers. Table 1 highlights the tasks and skills achieved through them.

Table 1. Identifying and Deciding the Tasks for the Game

Tasks	Skills Achieved
Oral Expressions	Reading a book, shopping list, repeating list on demand, answering based on intuition
Cognition	Syncing of attention, memory, and reasoning in responses
Naming and Identification	Greeting a person, identifying the articles, identifying proverbs for a person
Arithmetic	Basic addition, subtraction, division and multiplication
Auditory Comprehension	Listening to the other person, a responding mechanism based solely on hearing

Now we proceeded with the creation of 3D environments where the following tasks can be completed. [2] Firstly, they were conceptualized and initial renderings were completed in Autodesk Maya and then exported to Unity for game improvements. Royal free textures for walls and buildings were downloaded from TextureHaven.com. The concept art and Unity zoom-out view can be seen in Figures 1, 2, and 3.

Figure 1. Bakery Shop Render

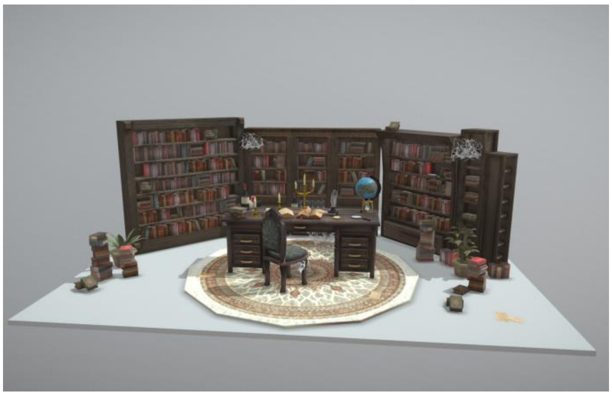

Figure 2. Library Render

Figure 3. Park Render

Three AI chatbots were created namely Ram, Alex, and Lisa, and were placed in three environments namely V Library, V Bakery, and V Park.

In each of these locations, the above five tasks were put to use. These conversations, as said earlier, are necessary for any person to know and survive. In each one of them, we have ensured one degree of the task gets more preference in order to allow the users and their consultants to practice anything specific they want in one environment.

2.3. Creating Sample Conversation Dialogues

With the help of students at JHU (refer to acknowledgment), sample dialogues were created as you can see in Figures 4,5, and 6 based on several headers required in day-to-day lives. These were then tested and trained in our model. We created more than 30 texts to see the limits of our chatbots.

Things I Can Say When Greeting Someone	Hello!
Good morning!	Good afternoon!
Good evening!	Hi! How are you?
Hey! How are you?	What's new?
How are you doing?	How is everything?

Figure 4. Sample text for AI Greetings

Things I Can Say When I Need Help	Can you help me please?
This is really tricky. Can I get your help?	I need help please.
Can you show me how to do this please?	I'm getting frustrated. I need your help.
Can I get your help with this?	Could you please help me out with this?
Could you please explain this to me?	Can you do me a favor?

Figure 5. Sample text for AI Help

THINGS I CAN SAY TO OTHER KIDS WHILE AT THE PLAYGROUND	Do you want to play with me?
Can you give me a push on the swings?	Do you want to play tag?
Hi! My name is _______. What's your name?	I'm _____ years old. How old are you?
Sure, I would love to play with you!	Can I have a turn?
Would you like to have a turn?	What are you doing?

Figure 6. Sample text for Conversation Choices

2.4. Authors

When the game initiates, the user will get a warm greeting followed by a menu that gives the option of the three environments. In each case, the user shall be greeted by the AI chatbots at the beginning followed by a reply by the user. The game ensures that the chat is kept one-to-one with more time for the user to think and reply. The following flow chart in Figure 7 highlights the flow of conversation same in all the cases.

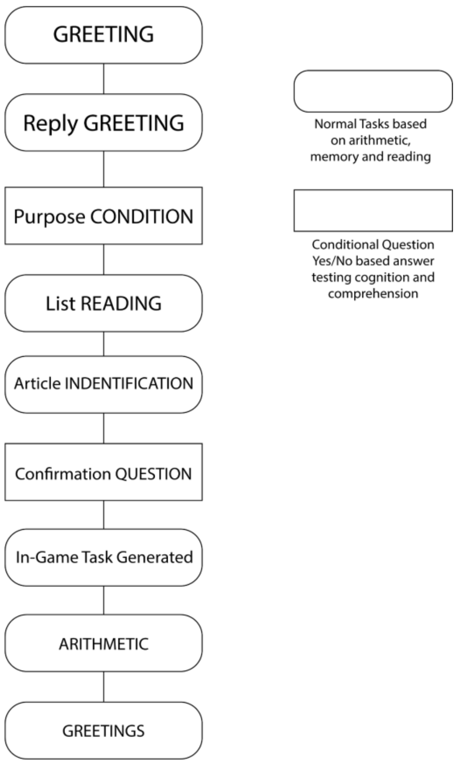

Figure 7. Flow of General Conversation and Tasks

2.5. Materials Needed

In order to play the game we used a standard VR Headset of Samsung (Model - SM-R322NZWATPA) with an external wired microphone.

2.6. Experimentation Procedure

Students (n=25) were selected from the NGO for the VR experiments. Their profile is available in Table 2. The game was, at the same time, seen on the laptop at a separate end of the room. A grading sheet was prepared as seen below where the instructors and I graded them after mutual agreement. Time was also a critical factor to see the response and its effect.

Table 2. Characterization of Participants

Gender		**Participants, n (%)**
	Male	7 (8%)
	Female	18 (72%)
Age (years)		
	<15	14 (56%)
	≥15	11 (44%)
Education		
	Primary Grades	9 (36)
	Middle School Grades	8 (32)
	High School Grades	8 (32)

1st VR attempt with the bakery shop (single subject controlled VR experiment):

This is a psychological method for experimentation, known as single-subject controlled, where we interview each student in a single room, that is, one student at a time. The students were graded out of 5 here. The form and marking are in the Results and Discussion section.

Creating a matrix with a time factor:

The students (n=25) were divided randomly into groups of five to check their status. Each group was named a1, a2, a3, a4, and a5. A matrix was created on the board first using the time calculated in each observation with the video recorded. This was a simple checking mechanism created to check the efficacy of VR. A sample board observation can be seen in Figure 8.

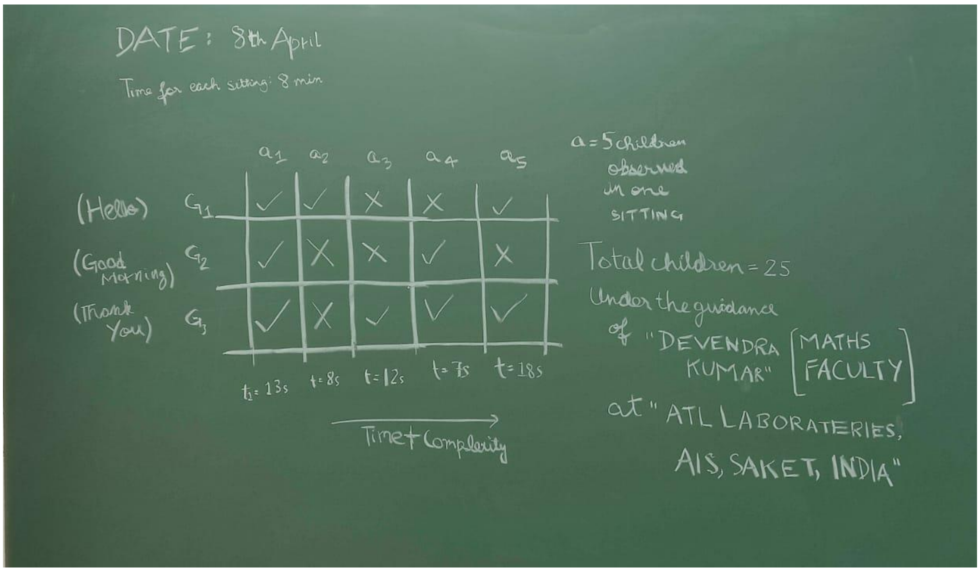

Figure 8. Immediate Board Session after 1st VR Experiment

Accordingly, we conducted two more experiments in group and subject-controlled format respectively.

3. DESIGN OF INVENTION

In this section, the whole model with some snippets of the code is shown to truly understand the dynamics and functioning of the VR game. Core concepts have also been explained with respect to the project.

3.1. Topic Modelling and Coherence

A text analytics algorithm was created to find the group of words from the given file that is the real-time converted voice/input of the user. The words from the file are then separated and paired together to form a topic. These words are then put together under the three environments we created. For example, if "cake" is said, then it gets registered under the bakery environment's document.

This gets further specified in the game using topic coherence to evaluate topic models. It uses the latent variable models. Each generated topic has a list of words. In the topic coherence measure, we found find average/median of pairwise word similarity scores of the words in a topic. The high value of the topic coherence score model will be considered as a good topic model to assess the person.

3.2. Latent Semantic Analysis Mathematical Concept - Singular Value Decomposition

LSA works on the principle of the bag of Word (boW) model, that is, a combination of the term-document matrix, rank lowering, and linguistic analysis used in Natural Language Processing.

Rows represent terms and columns represent documents. LSA learns latent topics by performing a matrix decomposition on the document-term matrix using Singular value decomposition.

Singular Value Decomposition: SVD is a matrix factorization method that represents a matrix in the product of two matrices. Figures 9 and 10 show the working of the matrices.

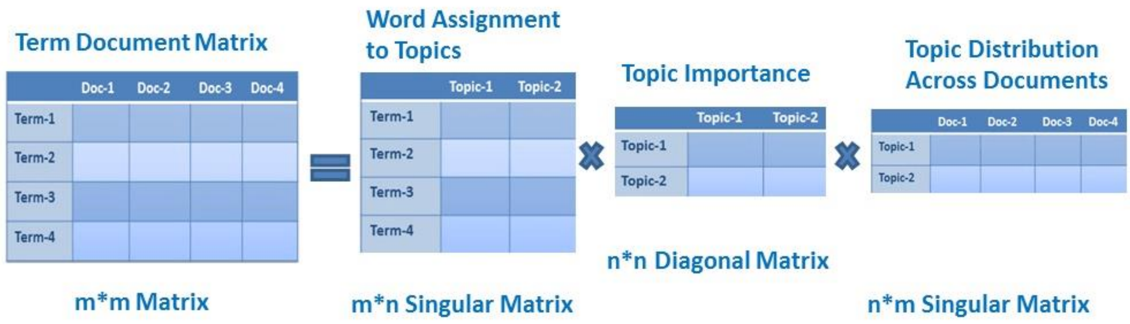

Fig. 9 Matrix Components

$$M = U\Sigma V^*$$

Fig. 10 Singular Value Decomposition

where
M is an m×m matrix
U is an m×n left singular matrix
Σ is an n×n diagonal matrix with non-negative real numbers.
V is an m×n right singular matrix
V* is an n×m matrix, which is the transpose of the V.

3.3. Latent Semantic Analysis Procedure

1. The voice is recorded through the microphone and converted to text in real-time in the back end. The following code below in Python helps to achieve the same.

```
importspeech_recognition as sr
r = sr.Recognizer()
mic = sr.Microphone(device_index=0)
with mic as source:
audio = r.listen(source)
result = r.recognize_google(audio)
with open('my_result.txt',mode ='w') as file:
file.write("Recognized text:")
file.write("\n")
file.write(result)
```

Post this, latent semantic analysis is used keeping in mind the average word length, speech rate, average word duration, number of unfilled pauses, number of unfilled crossings, etc. For these criteria to be fulfilled, variables and sounds were classified as lexical, acoustic, and syntactic.

2. Using Gensim for LSA, we first imported the library, particularly importing the Coherence model:

```
importos.path
fromgensim import corpora
fromgensim.models import LsiModel
fromnltk.tokenize import RegexpTokenizer
fromnltk.corpus import stopwords
fromnltk.stem.porter import PorterStemmer
fromgensim.models.coherencemodel import CoherenceModel
importmatplotlib.pyplot as plt
```

3. We created data load function to load articles. [4]

```
defload_data(path,file_name):
    """
Input  : path and file_name
    Purpose: loading text file
Output : list of paragraphs/documents and
title(initial 100 words considred as title of document)
    """
documents_list = []
titles=[]
with open( os.path.join(path, file_name),"r") as fin:
for line in fin.readlines():
text = line.strip()
documents_list.append(text)
print("Total Number of Documents:",len(documents_list))
titles.append( text[0:min(len(text),100)] )
returndocuments_list,titles
```

4. After loading the data, we proprocessed the text by tokenizing the text articles, removing the stop/limit words, and by stemming the final doc. [5, 6,7]

```
defpreprocess_data(doc_set):
    """
Input  :docuemnt list
    Purpose: preprocess text (tokenize, removing stopwords, and stemming)
Output :preprocessed text
    """
    # initialize regex tokenizer
tokenizer = RegexpTokenizer(r'\w+')
    # create English stop words list
en_stop = set(stopwords.words('english'))
    # Createp_stemmer of class PorterStemmer
p_stemmer = PorterStemmer()
    # list for tokenized documents in loop
texts = []
    # loop through document list
fori in doc_set:
        # clean and tokenize document string
raw = i.lower()
tokens = tokenizer.tokenize(raw)
        # remove stop words from tokens
stopped_tokens = [i for i in tokens if not i in en_stop]
        # stem tokens
stemmed_tokens = [p_stemmer.stem(i) for i in stopped_tokens]
```

```
        # add tokens to list
texts.append(stemmed_tokens)
return texts
```

5. Here, we created a term matrix and dictionary of terms based on the document we received after preprocessing. This was our preparation of the corpus - critical for the whole model to succeed.

```
defprepare_corpus(doc_clean):
    """
Input  : clean document
    Purpose: create term dictionary of our courpus and Converting list of documents (corpus) into Document Term Matrix [8, 9]
Output : term dictionary and Document Term Matrix
    """
    # Creating the term dictionary of our courpus, where every unique term is assigned an index.
dictionary = corpora.Dictionary(doc_clean)
dictionary = corpora.Dictionary(doc_clean)
    # Converting list of documents (corpus) into Document Term Matrix using dictionary prepared above.
doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]
    # generate LDA model
returndictionary,doc_term_matrix
```

6. Creating the Model

Now we created our own model using the corpus.

```
defcreate_gensim_lsa_model(doc_clean,number_of_topics,words):
    """
Input  : clean document, number of topics and number of words associated with each topic
    Purpose: create LSA model using gensim
Output : return LSA model
    """
dictionary,doc_term_matrix=prepare_corpus(doc_clean)
    # generate LSA model
lsamodel = LsiModel(doc_term_matrix, num_topics=number_of_topics, id2word = dictionary)  # train model
print(lsamodel.print_topics(num_topics=number_of_topics, num_words=words))
returnlsamodel
```

Moving ahead, to optimize the results, we determined the optimum amount of topics the AI chatbot/avatar needs to cover during the conversation.

```
defcompute_coherence_values(dictionary, doc_term_matrix, doc_clean, stop, start=2, step=3):
    """
    Input   : dictionary :Gensim dictionary
corpus :Gensim corpus
texts : List of input texts
stop : Max num of topics
purpose : Compute c_v coherence for various number of topics
Output  :model_list : List of LSA topic models
coherence_values: Coherence values corresponding to the LDA model with respective number of topics
```

```
    """
coherence_values = []
model_list = []
fornum_topics in range(start, stop, step):
        # generate LSA model
model = LsiModel(doc_term_matrix, num_topics=number_of_topics, id2word = dictionary)  #
train model
model_list.append(model)
coherencemodel = CoherenceModel(model=model, texts=doc_clean, dictionary=dictionary,
coherence='c_v')
coherence_values.append(coherencemodel.get_coherence())
returnmodel_list, coherence_values
```

To study the data, best method is to visually capture it through graphs and their slopes. Thus, we used topic coherence values to be put for the visual analysis where:

X-axis: Presents the number of topics

Y-axis: Presents the coherence score

```
defplot_graph(doc_clean,start, stop, step):
dictionary,doc_term_matrix=prepare_corpus(doc_clean)
model_list, coherence_values = compute_coherence_values(dictionary,
doc_term_matrix,doc_clean,
stop, start, step)
    # Show graph
    x = range(start, stop, step)
plt.plot(x, coherence_values)
plt.xlabel("Number of Topics")
plt.ylabel("Coherence score")
plt.legend(("coherence_values"), loc='best')
plt.show()
```

3.4. Final Latent Semantic Analysis Algorithm Flow

1. Preparation of a Word by Text Rectangular Matrix:
   A co-occurrence matrix that specifies the number of times that word Wi occurs in text Tj. A cell in the matrix is designated as fr(Wi, Tj). The matrix is extracted from all of the words and texts in the entire corpus.

2. Cell Values' Transformation:
   A frequency is determined for each cell by converting it to logarithms: log [fr(Wi, Tj) + 1]. Second, there is a computation that estimates the relative distinctiveness of the word to a particular text, relative to the alternative texts.

3. Singular Value Decomposition:
   SVD decomposes the first matrix {X} into the product of three component matrices {W}, {S}, and {P}. LSA determines a best-fit set of component matrices that approximately reproduces {X}, that is, {X} = {W}{S}{P}.

   The {W} matrix maps the set of words onto the set of K dimensions (i.e., functional features). If there are N-words and K dimensions, an N by K matrix is constructed with each cell showing word dimension combination. {S} is a vector with K values that weights the generic importance of each of the K dimensions. (P} is a K by T matrix that maps the K dimensions onto the set of T texts. Therefore, the Word by Text matrix is reduced to K dimensions that serve as functional features in an easy-to-use data analysis set.

4. RESULTS AND DISCUSSION

The VR game in its current stage is successful at simulating 3D environments with AI Avatars/chatbots. The AI chatbots/avatars were appreciated as well. The coherence score through LSA and document-word matrix, are the two primary computational results needed for the planning and execution of UMEED.

4.1. Coherence Score Through Latent Semantic Analysis

Start, stop, and step values of 25 observations were taken as 2, 12, 1 after observation and median calculation through the model. Figure 11 shows that 6 to 8 topics could be brought up in the conversation with a peak score of 0.60. This score essentially tells that post-VR game, UMEED, students affected with speech disorders could initiate and reply successfully with little to no struggle, as confirmed by the NGO instructors who also assessed the students.

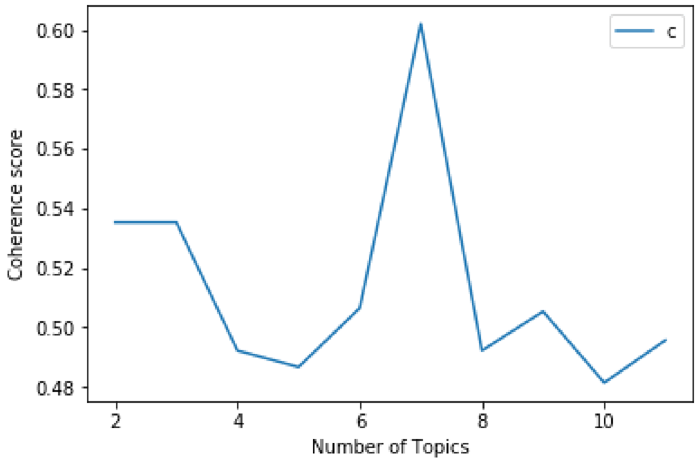

Figure 11. Graphical Analysis of the Topics Covered with respect to Coherence Score

4.2. Document Word Matrix

Figure 12 shows the number of topics repeated in the matrix generated, where the topics are rows and the number of groups is columns. This graph was approved by the instructors at the NGO and the undergraduate students I worked with to make this the industry level. We also ensured to follow a colour scheme for the matrix that makes it understandable for the autistic population. The darker cells represent the topics covered more in-depth compared to the light-shaded ones at a particular instant of time.

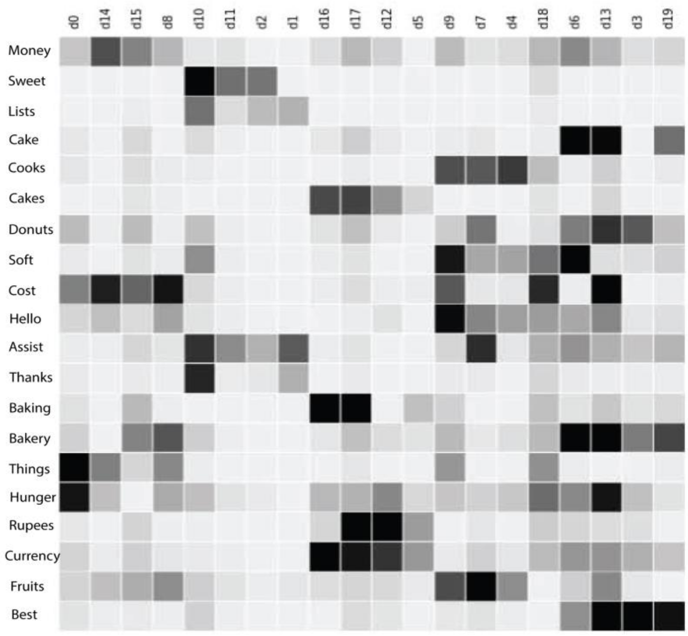

Figure 12. Document Word Matrix for Topic Modelling

4.3. Initial Virtual Reality Reading

In the three conversations and interviews (combined), we graded the participants for their conversation in the presence of the instructors, and me. The following table 3 below highlights the grading that was done in the finals.

Table 3. Grading on Increasing Difficulty and Traits based on Response by Participants

Criterion	Influence degree		
	Great	Medium	Small
Theory analysis	3.5	2.4	3.3
Working experience	2.3	4.2	1.1
Referring to literature	4.15	1.1	3.05
Self-intuition	3.05	2.1	4.85

4.4. Scope for the Future

For the following year, the plan is to go ahead and collaborate with researchers in behavioural and medical sciences to conduct studies on autism spectrum disorder and aphasia with respect to future technologies. This shall help in optimizing the game further.

Collaboration with the industry's 3D artists is also in a plan to create more interactive environments in Unity.

Furthermore, human relations are critical for anyone to survive and ensure peace of mind. Thus, in the future, using sentiment analysis and a bi-directional recurrent neural network, UMEED can host complex emotional programs such as making a friend, dealing with parents, conversing with teachers, etc.

5. CONCLUSIONS

This VR game has successfully empowered the students involved in the trials to feel confident. UMEED is a computationally efficient, and user-friendly solution for combatting speech disorders such as autism spectrum disorder, aphasia post-stroke, etc.

There is a thorough backend analysis of the dialogues (input) given to the user and the response (output) generated by the AI chatbots using latent semantic analysis with natural language processing. From the results and discussions, it was clear that the front-end work, that is, the 3D environments was simple to access and did not overwhelm them, which has historically posed a challenge.

The point collection system was additionally loved by the primary grade students. Our game will not only cater to the ones with speech disorders but also raise awareness about these societal issues. With more data to be acquired in the future for further analysis, it is clear that UMEED can go global and attract talents from various fields.

With the rise of COVID-19 and psychological disorders, it is imperative that bold steps packed with innovation need to be taken to face the battles with hope, precisely the name of our brand, UMEED.

ACKNOWLEDGEMENTS

I would sincerely like to thank the Atal Tinkering Labs (an initiative by the Department of Science and Technology for the Government of India) for opening an ATL lab where most of the work was done for this research.

I would like to further express my thanks to ReshamTalwar, an undergraduate researcher at Johns Hopkins University, has been instrumental in getting UMEED the recognition it needed in the scientific community in the United States through the JHU's DREAMS Symposium (Spring).

I would like to thank Pranav Talreja, the first collaborator on UMEED, for partnering with me for the Moonshot Pirates challenge where UMEED won the People's Choice Award as the most-voted online venture and scientific research. His role in spreading UMEED's work has been pivotal to its launch and research I did post the Moonshot Pirates event.

REFERENCES

[1] Brady M, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. Cochrane DB Syst Rev 2020 Jun 01:2016.

[2] Intille, S. S., "Designing a Home of the Future," IEEE Pervasive Computing, pp. 80-86, 2002.

[3] Lavoie M, Routhier S, Légaré A, Macoir J. Treatment of verb anomia in aphasia: efficacy of self-administered therapy using a smart tablet. Neurocase 2016 May 26;22(1):109-118.

[4] Miikkulainen, R., "Subsymbolic Case-Role Analysis of Sentences with Embedded Clauses," Cognitive Science, vol. 20, pp. 47-74, 1996.

[5] Katz, B., "From Sentence Processing to Information Access on the World Wide Web," presented at AAAI Spring Symposium on Natural Language Processing for the World Wide Web, Stanford University, 1997

[6] Graham JR, Pereira S, Teasell R. Aphasia and return to work in younger stroke survivors. Aphasiology 2011 May 25;25(8):952-960.

[7] Rehder, B., Schreiner, M. E., Wolfe, M. B. W., Laham, D., Landauer, T. K., and Kintsch, W., "Using Latent Semantic Analysis to Assess Knowledge: Some Technical Considerations," Discourse Processes, vol. 25, pp. 337-354, 1998

[8] Marshall J, Booth T, Devane N, Galliers J, Greenwood H, Hilari K, et al. Evaluating the benefits of aphasia intervention delivered in virtual reality: results of a quasi-randomised study. PLoS One 2016 Aug 12;11(8):e0160381

[9] Chen Y, Li Z, Peng J, Shen L, Shi J. Development of a care bundle for stroke survivors with psychological symptoms: evidence summary and Delphi study. Evid Based Complement Alternat Med 2020 Jun 30;2020:7836024

AUTHORS

Hridyanshu is an undergraduate student researcher at the University of California San Diego, United States who has extensive experience in scientific research and product development. From presenting award-winning work at NASA Johnson Space Center, Houston to collaborating with individuals at the New York Academy of Sciences, Hridyanshu possesses a vision for futuristic technologies with a focus on the integration of computer science, computational linguistics, cognitive sciences, and designing in various research fields.