As a researcher in machine learning (ML) and natural language processing (NLP), I build probabilistic and statistical models to simulate and discover latent patterns such as human behaviors in unannotated data such as conversational corpora. Unlike other data such as images, videos, and sound, natural language is mainly created by humans. Accordingly, when analyzing texts, we need to think about who creates it, whereas we can also understand people through texts. Following these principles, I conduct research in NLP by focusing on the relationship between humans and texts. As data, I mainly look at the conversation. A conversation is a typical communication between people, with the context depends on the speaker. For example, a speaker may have different conversations with different conversational partners. This difference is an important factor when attempting to understand the speaker, with the speaker also making the conversation more understandable through the context.

In my research, I aim to 1) develop conversation models that generate human-like responses, and 2) investigate human behaviors from their conversations using novel ML models.

Conversation Models for Generating Human-like Responses

One of the goals in artificial intelligence research is to create an intelligent agent that can acts as a human. Turing presented a test to answer the question, "Can machines think?" in the form of the Turing test. In the test, a human has open-domain conversations that are no restriction with regard to domains or topics with the intelligent agent, communicating with only text. The agent generates human-like responses to the conversations. When the human cannot distinguish whether the conversational partner is a human or an agent, the test is passed.

I conducted research on open-domain conversation modeling to generate conversational responses that match those of a human. Specifically, I developed a neural conversation model that models speakers and generates personalized conversational responses. Additionally, I developed a neural network model that evaluates the generated responses automatically.

Generating Personalized Conversational Responses

In conversation response generation, modeling the speakers in addition to the utterances is important for generating appropriate responses. Knowing information about a speaker, such as her linguistic style or personal information can help predict her response, and knowing more about both speakers from their previous conversations can help predict the contents of their conversation. One more difficult and under-addressed issue in conversation response generation is the cold-start problem, which arises when the training data do not contain one or both of the speakers. It then becomes very difficult to predict the appropriate responses.

Questioner	Answerer	How old are you ?
Speaker A	Speaker B	18 yr old
Speaker A	Speaker C	nothing much :)
Speaker B	Speaker A	i'm not sure , but i am a bit older than you
Speaker C	Speaker A	19 ! ! !

Figure 1: Example of generated responses by VHUCM to speakers' personal information question

To incorporate the speakers into the neural conversation model, I developed the Variational Hierarchical User-based Conversation Model (VHUCM), a variational auto-encoder based open-domain conversation model that models the speakers to generate personalized responses. The main assumption in the model is that the conversational context depends on the speaker. For example, the same speaker pair is likely to have similar conversations. Therefore, I set one latent global variable to represent the context of a conversation, and it is inferred by the conversation speakers' latent variables. VHUCM can generate personalized responses based on the speaker and conversation partner. Table 1 shows examples of personal questions and responses. The answer of A for question from the B is interesting. From the other answers, we know that age of A is 19, and B's age is 18. The answer to the question is correct even it does not generate the exact age number. Additionally, I tested a new user situation in which one of the conversation speakers never appears in the training data. To avoid this cold-start problem, I inferred the new speaker variable from the conversation partners, and it shows better performance than other baselines including the random case. This project has been published at EMNLP 2019 [1].

Evaluating Conversational Responses Automatically

Evaluating responses generated by the machine learning model for open-domain dialogue automatically is a difficult task. There are many possible appropriate responses given a dialogue context, and automatic metrics such as BLEU or ROUGE rate responses that deviate from the ground truth as inappropriate. For example, when the dialogue context is referring to a date plan, the appropriate responses can be accepting, rejecting, or suggesting other plans. However, BLEU and ROUGE fail to evaluate an appropriate response when the response is acceptance but the ground truth response is refusal. ADEM considers the relationship between the context and the generated responses, but it requires human-annotated scores to train the model.

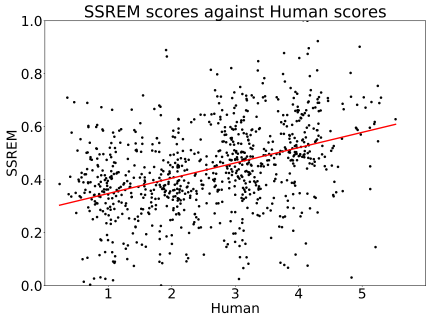


Figure 2: Scatter plot that shows SSREM scores against human scores

To overcome these limitations, I developed the Speaker Sensitive Response Evaluation Model (SSREM), which models the functions to compare the coherence between a given conversational conte xt and a generated response without human-labeled data. The main idea of the model is that utterances from the same speaker are more difficult negative samples to classify the true response than random negative responses. I adopted a self-supervised learning to train the model. More, I built diverse negative samples by considering speakers rather than random negative responses. This was done because random negative samples are easy to be distinguishable from the true response, so it makes the model to poor performance. SSREM shows a higher correlation with human judgments than existing baselines. Figure 2 shows the results. The red line is a linear regression line, and the coefficient of the line has a stronger positive correlation with human judgment compared to those of the other models. SSREM is also applicable when the training corpus and test corpus differ. This project has been published at ACL 2020 [2].

Statistical Models for Analyzing Human Behaviors in Texts

To devise a human-like acting intelligent agent, understanding human behaviors in text is important. Humans are social animals. They want to have social relationships between others. The activity of using text is related to social interaction. For example, people talk, write, call, or text friends, families, neighbors, or colleagues. These communications increase the closeness between people and allow people to collaborate with others.

I researched modeling to identify human behaviors and to discover the important factors of behaviors from conversations. Specifically, I developed a Bayesian model to identify self-disclosure in casual conversations. I also developed a neural network model to identify decision making process in a discussion corpus.

Modeling Self-Disclosure Behavior

Self-disclosure is important and pervasive social behavior. People disclose personal information about themselves to improve and maintain relationships. A common instance of self-disclosure is the start of a conversation with an exchange of names and additional self-introductory information. Another example of self-disclosure is where the information disclosed is about a family member's serious illness, which can be much more personal than a simple exchange of names. However, most researchers use surveys or human annotations to determine self-disclosure. These methodologies are biased by the participants' memories and cannot be applied to large datasets.

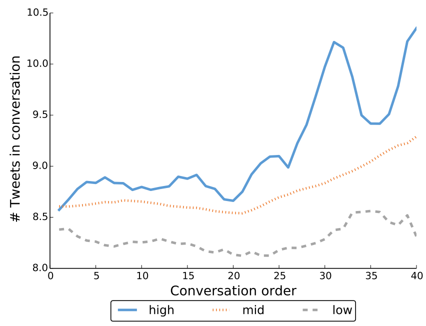


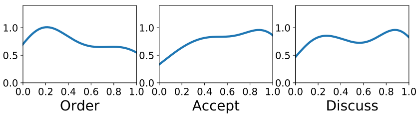
Figure 3: Changes in conversation length over time

To investigate the self-disclosure behavior in a large conversation corpus, I developed the Self-Disclosure Topic Model (SDTM), a Bayesian topic model with external prior knowledge. It classifies the self-disclosure level into three categories: low, mid, and high disclosure. Self-disclosure behavior can be modeled using a combination of simple linguistic features (e.g., pronouns) with automatically discovered semantic themes (i.e., topics). To identify high disclosure themes, I built the external prior knowledge by extracting seed words and phrases from an external corpus of anonymously posted secrets. SDTM outperforms baseline models in the identification of self-disclosure behavior. Additionally, it shows a correlation between self-disclosure and the closeness of relationships, thus supporting earlier results in social psychology research. Figure 3 shows an example of outcome. The conversation length increases noticeably over time in the medium and high self-disclosure level groups but only slightly in the low self-disclosure level group. This project has been published at ACL 2012 [3] and at EMNLP 2014 [4].

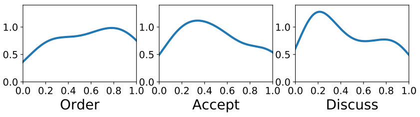
Modeling Decision-Making Process Behavior

Decision making in groups refers to the process of making choices to resolve issues by discussing the issues with group members. Social psychologists note that decision making affects the group performance and the satisfaction of its members, and that leadership also plays a role. However, most research about the decision-making process use surveys or a lab environment, causing bias in participants.

To investigate decision-making process behavior, I built discussion records from the annals of the Joseon dynasty (AJD). The AJD consists of the historical records of kings who governed the Korean peninsula from 1392 to 1910. In the AJD, the kings discuss national issues with government officials and decide upon a course of action such as ordering or accepting. To identify the leaders' decisions from a discussion, I developed Conversational Decision-Making Model (CDMM) which is a hierarchical RNN model with attentions. The model incorporates words in an utterance with the speaker and predicts the leaders' final decisions. CDMM outperforms baseline models and clarifies the keywords and key members to identify the king's final decisions. Figure 4 shows an example outcome. CDMM assigns a high attention score to the word "Okay" for an 'Accept' decision as compared to the other decisions when the speaker is the king. However, when officials use this word, CDMM assigns a high attention value to the word in the 'Order' decision. This project has been published at LaTeCH-CLfL 2015 [5] and at EMNLP 2018 [6].



(a) Word "Okay" from kings



(b) Word "Okay" from officials

Figure 4: Attention weight distribution of word for each decision from kings and officials

Future Research Directions

Open-domain conversation modeling has many challenges: the problem is excessively and lacks clear boundaries of the domain. Recently, Google and Facebook researchers suggested methodologies that use various and large conversation corpora with complex structural models [7, 8]. These models cannot easily identify the reasoning behind the generation of non-appropriate responses because they function as a black-box.

To overcome this limitation, we need to investigate conversations by answering the following question first: "What, Who, Where, When, Why, and How do people have conversations?" From these characteristics of conversational behaviors, we can understand the struggles in the conversation modeling and devise ideas to solve them. My research will continue to investigate human behaviors in a huge number of texts. The analysis will help me and other researchers gain a better understanding of conversations and will serve to generate as insightful ideas for conversation modeling.

Topics of Conversations

In the near future, I plan to investigate what people are talking about in conversations. People have casual conversations about various topics, such as politics, economics, and sports. As part of this potential direction, I now investigate the degree of difficulty in generating responses based on conversation topics. A person who participates in the conversation finds it easy or difficult to answer depending on the topic of the conversation. For example, it is easy to talk about yesterday's sports event, but responding to political issues is difficult. I aim to build ML models that identify the level of difficulty of answering based on the topic. The results will define the relationship between response difficulty and the topic, possibly leading to the development of conversation models that can respond to specific topics fluently.

Intentions of Conversations

There are various reasons why people enter into conversations with each other such as to manage relationships, exchange information, and solve problems. When a person participates in a conversation, he/she expects responses from the conversation partner. To generate appropriate responses, the conversation model should identify the intentions of the conversation. I aim to build ML models that identify the intentions of conversations, and incorporate them into the conversation model. I plan to utilize the definition of intentions from discourse analysis research and build computational approaches to refine it.

Multimodality of Conversations

A conversation does not only consist of texts. It has various forms of metadata, such as speakers, times, locations, and voices. Specifically, AI speakers such as Amazon Alexa and Google Assistant engage in conversations with humans using a voice. In my work, I have found that combining voice and text together can be important when seeking to identify emotions [9]. I aim to build ML models that use additional metadata to understand texts and human behaviors more accurately.

Beyond Conversations

People communicate with others by texts, and turn-taking conversation one of many ways they do this. There are many types of texts, such as articles, letters, and comments, among others. Though I focus on turn-taking conversations, I can investigate NLP research problems with the same principles. In particular, I aim to solve the text summarization problem, as it has a structure similar to that of conversation modeling. Other goals are to understand existing texts, consider authors/readers, and generate appropriate sentences.

I am also interested in multilingual NLP research. Most NLP research work focuses on specific languages, such as English or Chinese. Despite the fact that ML models are independent of language, considering language-dependent linguistic and cultural characteristics can be important. I devised what can be termed a 'human-in-the-loop system' to identify and expand keywords in Indonesian tweets [10]. With this experience, I aim to investigate NLP research problems, especially in low-resource languages.

These research directions are inherently interdisciplinary. Social scientists investigate conversational behaviors qualitatively, while machine learning researchers analyze them by developing computational models and applying large corpora. These qualitative and quantitative study results can instill ideas in both types of researchers. I have collaborated with political scientists [11, 12]. I am ready to collaborate with people from various fields, and I am certain that I can contribute to both research fields.

References

[1] **JinYeong Bak** and Alice Oh. Variational hierarchical user-based conversation model. In *Proceedings of the EMNLP*, 2019.

[2] **JinYeong Bak** and Alice Oh. Speaker sensitive response evaluation model. In *Proceedings of the ACL*, 2020.

[3] **JinYeong Bak**, Suin Kim, and Alice Oh. Self-disclosure and relationship strength in Twitter conversations. In *Proceedings of the ACL*, 2012.

[4] **JinYeong Bak**, Chin-Yew Lin, and Alice Oh. Self-disclosure topic model for classifying and analyzing Twitter conversations. In *Proceedings of the EMNLP*, 2014.

[5] **JinYeong Bak** and Alice Oh. Five centuries of monarchy in Korea: Mining the text of the annals of the Joseon dynasty. In *Proceedings of the SIGHUM LaTeCH*, 2015.

[6] **JinYeong Bak** and Alice Oh. Conversational decision-making model for predicting the king's decision in the annals of the Joseon dynasty. In *Proceedings of the EMNLP*, 2018.

[7] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. Towards a human-like open-domain chatbot, 2020. [link].

[8] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot, 2020. [link].

[9] Gabriel Lima and **JinYeong Bak**. Speech emotion classification using raw audio input and transcriptions. In *Proceedings of the SPML*, 2018.

[10] **JinYeong Bak**, Imaduddin Amin, Jong Gun Lee, and Alice Oh. Keyword expansion for understanding crisis events in Indonesian tweets. In *ICML Workshop on Interactive Machine Learning and Semantic Information Retrieval*, 2017.

[11] **JinYeong Bak** and Oul Han. "Meaning Structures" in political discourse: Measuring institutional dynamics of a hybrid democracy via the topic modeling from contested concepts in newspaper articles. GESIS CSS Symposium, 2014. [link].

[12] **JinYeong Bak**, Christopher Paik, and Alice Oh. At the king's command: rules and natural disasters in the annals of dynasty. POLTEXT, 2019. [link].