

Natural Language Processing to Assess Communication Dynamics between Cooperating Dyads during Video Gameplay

Jan Watson¹, Adrian Curtin¹, Sukethram Sivakumar¹,
Yigit Topoglu¹, Nicholas DeFlippis¹, Jintao Zhang²,
Rajneesh Suri^{2,3}, and Hasan Ayaz^{1,3,4,5,6}

¹School of Biomedical Engineering, Science and Health Systems,
Drexel University, USA

²LeBow College of Business, Drexel University, USA

³Solutions Institute, Drexel University, USA

⁴Department of Psychological and Brain Sciences, Drexel University, USA

⁵Department of Family and Community Health, University of Pennsylvania, USA

⁶Center for Injury Research and Prevention, Children's Hospital of Philadelphia,
Philadelphia, PA 19104, USA

ABSTRACT

Latent Dirichlet Allocation (LDA) and Sentiment Analysis have become prominent tools in natural language processing applications for both research and industry (Jelodar et al., 2019; Yadav and Vishwakarma, 2020). While LDA is a generative probabilistic modeling methodology that is widely used in Topic Modeling to extract underlying themes and topics from a collection of words (Blei et al., 2003), Sentiment Analysis is defined as identifying the hedonic tone of a corpus of text (Yadav and Vishwakarma, 2020). Here, supervised Sentiment Analysis is used to classify conversations between team gaming dyads in terms of valence. Additionally, LDA is utilized to label segments of cooperative conversation between dyads as topics. Fourteen participants were paired as dyads (7 teams) and instructed to complete thirty-two 150 second gaming scenarios (trials) in the first-person shooter (FPS) video game *Overwatch*. While completing the scenarios, participants were instructed to communicate with their respective teammate via a voice communication headset. The conversations from each scenario were transcribed from recorded voice channels before analysis was performed. Our approach examines the relationship between perceived task difficulty and both conversation sentiment scores and topic frequency in both novice and experienced skill groups. Preliminary results indicate evidence that conversation topic, sentiment and perception dynamics are consistent with an encouragement and frustration sentiment paradigm.

Keywords: Natural language processing, Latent dirichlet allocation, Sentiment analysis, Communication

INTRODUCTION

The objective measurement of group cooperation and communication along with the assessment of team based outcomes has attracted increased attention

in both research and industry (Jones and George, 1998) with team-based video games ideally suited to serve as a platform for teamwork and communication research as players work together to reach a common outcomes (Greitemeyer and Cox, 2013).

Asverbal communication is a primary signal for collaboration and cooperation, natural language processing (NLP) tools have developed to categorize and classify conversational features that can serve as metrics for teamwork with the two prominent methods of Sentiment Analysis and Latent Dirichlet Allocation (LDA) primarily utilized. Through Sentiment Analysis, the feeling (or connotation) of a corpus of text can be observed for both positive and negative valence (Yadav and Vishwakarma, 2020) while LDA utilizes unsupervised machine learning to label segments of text as topics (Blei et al., 2003).

This study seeks to compare the relationship between subjects' speech language and self-perceptions of in-game performance by examining the interplay between self-reported measures, sentiment scores and conversational topic modelling.

METHODS

Participants

A total of 14 healthy adult participants (6 female, mean age = 22) volunteered for this study. All participants fulfilled the requirements of being right-handed with vision correctable to 20/20. Before the study, participants filled and signed the consent form approved by the Institutional Review Board of Drexel University.

Experiment Protocol

Participants completed a single one-hour study where participants sat in front of a computer and played team-based task scenarios in *Overwatch*, a first-person shooter (FPS) game designed by Activision-Blizzard (Santa Monica, CA, www.activisionblizzard.com). Participants filled out background forms to classify whether they were a novice or experienced FPS video game players before playing two 16 trial sessions of the game with an assigned teammate of comparable skill level. Teams played in 'Escort Mode', an *Overwatch* game mode consisting of two teams with the opponent (two artificial intelligence (AI) bots set to a medium difficulty). Within a fixed duration the attacking team was instructed to escort a 'payload' to checkpoint in order to win the game, while the defending team was instructed to prevent the payload from being advanced to its destination. Each trial had a time limit of 150 seconds with 30 second interstitial rest periods. A single trial ended either when the timer expired, or the payload had been successfully escorted to the final checkpoint. At the end of each trial, participants were instructed to rate their perception of gameplay difficulty on a Likert scale (1 to 7). See Fig. 1 below for a visual depiction of the task and protocol diagram.

Data Acquisition, Processing and Analysis

Each participant dyad produced a total of four audio recordings initially transcribed by Google Cloud's Text-to-Speech API (Bisong, 2019) before

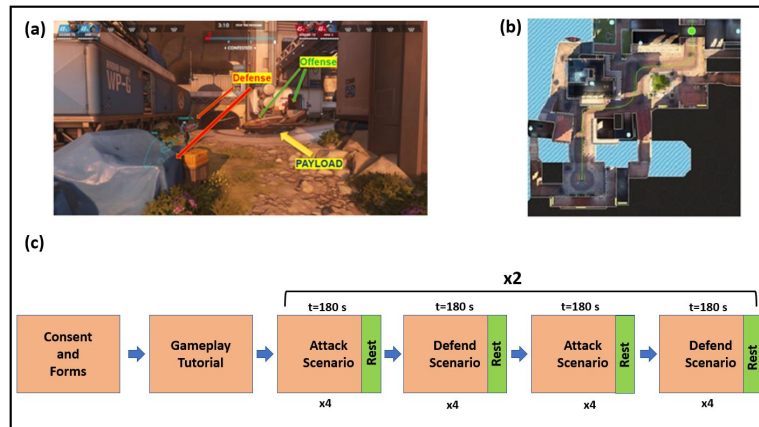


Figure 1: (a) An example snapshot of the Escort Mode, where the attacking team transports the ‘payload’ by standing close to it while defending team tries to prevent the attacking team escort the ‘payload’. (b) The map for the game scenarios. The green line represents the payload’s path, green dot represents the destination. (c) Diagrammatic representation experimental procedure.

recordings were manually corrected to correct machine error when converting conversations to text. All data were preprocessed using MATLAB’s Text Analytics toolbox. Each transcript (document) was separated into 8 bins that corresponded to a single experimental trial (~180 seconds). Prior to LDA analysis sentiment analysis was performed to produce Vader sentiment scores ranging from -1 to 1 , with the lower end reflecting negative sentiment and upper end reflecting positive sentiment.

For LDA modelling, each document bin was tokenized prior to stop word removal, punctuation erasure and lemmatization and converted into a single “bags-of-words” term frequency counter (a string array of unique words and a matrix of word counts) for fitting an LDA model with three topics which allowed for the calculation of the probability that a single topic was present as well as identifying the most frequent topic within a participants document. Additionally, each of three three topics generated were subjectively interpreted prior to statistical analysis producing the three topic labels “Directions”, “Frustration” and “Encouragement”.

Results

LDA results are summarized in Figure 2 comparing both novice and experienced FPS game players’ most frequently discussed topic during a single gaming session in terms of self-reported perception of the sessions difficulty (mean rating across session trials). A main effect is observed for Most Frequent Topic ($F_{2,17.8}=3.8$, $p = 0.04$) and post-hoc analysis reveals a significant interaction between the Directions Topic and Level of Experience ($F_{1,11.1}=8.56$, $p = 0.04$).

Figure 2 shows a comparison of novice and experienced players most frequently discussed session topic and their mean sentiment score (for all session trials) for gameplay during a single session. Here, there is a

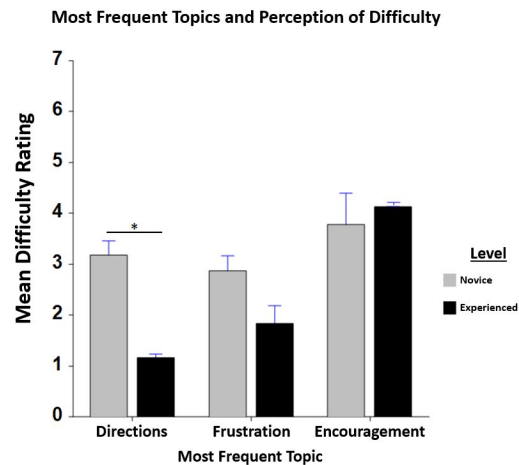


Figure 2: LDA results: Comparison between novice and experienced players mean self-reported trial ratings for game difficulty and the most frequently discussed topic within the session. Whiskers are standard error of the mean (SEM).

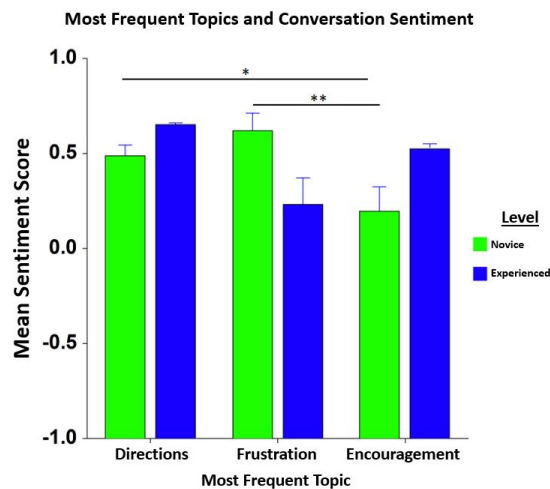


Figure 3: LDA and sentiment analysis results: Comparison between novice and experienced players mean trial sentiment score and the most frequently discussed topic within the session. Whiskers are SEM.

main interaction between Level of Experience and Most Frequent Topic ($F_{2,17.9}=4.51$, $p = 0.03$) and post-hoc analysis indicates significant interactions for the Novice Group between Direction and Encouragement Topics ($F_{1,18.5}=8.47$, $p = 0.05$) as well as for Frustration and Encouragement Topics ($F_{1,19.9}=10.24$, $p = 0.03$).

DISCUSSION

Just as the perception of task difficulty is closely tied to an individual's skill level, both the nuances of spoken words and the tone surrounding them may very well be subject to the effect of one's perception of expertise (or lack

of expertise). In this study, we examine the relationship between perceived task difficulty and both conversation sentiment scores and topic frequency in both novice and experienced skill groups. Preliminary results indicate that when perceived difficulty is high, both novices and experts tend towards words of encouragement when collaborating to achieve a common goal. However, experienced game players exhibit greater frequencies of frustrated language when task difficulty is low, highlighting that a lack of challenge catalyzes an intensified criticism of otherwise negatable behaviors. More interesting though, is an observable relationship between the prominent topics of conversation and sentiment scores related to these topics in the context of competitive tasking. It is apparent that for novices, despite being frustrated more often, these participants possessed an overall higher sentiment in these conversations, suggesting that the relative personal unimportance of a task lends to feelings of, “just being happy to be there”. Alternatively, experienced and seasoned players who favored topics of directives and encouragement maintained an overall heightened positivity over novices while experienced players favoring topics of frustration were significantly more negative than novices in sentiment score. These trends suggest a subtle paradigm at work, where internalized perception of skill and expertise lead to a higher contrast in sentimental valence between differing topics of conversation.

In conclusion, this study highlights the use of Natural Language Processing techniques, specifically Latent Dirichlet Allocation and Sentiment Analysis to elucidate both topic and hedonistic tone of conversation in a cooperative team-based competitive video gaming environment. These preliminary findings warrant further investigation into the relationship between self-perceptions of skill and performance and its effect on the language of teamwork.

REFERENCES

- Bisong E., (2019) “An overview of Google cloud platform services,” *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, pp. 7–10.
- Blei D. M., Ng A. Y., and Jordan M. I., (2003) “Latent Dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022.
- Greitemeyer T. and Cox C., (2013) “There’s no “I” in team: Effects of cooperative video games on cooperative behavior,” *European Journal of Social Psychology*, vol. 43, no. 3, pp. 224–228.
- Jelodar H. et al., (2019) “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211.
- Jones G. R. and George J. M., (1998) “The experience and evolution of trust: Implications for cooperation and teamwork,” *Academy of management review*, vol. 23, no. 3, pp. 531–546.
- Yadav A. and Vishwakarma D. K., (2020) “Sentiment analysis using deep learning architectures: a review,” *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385.