Problem Statement

SELF-AWARE COMEDY ROBOT

CAPSTONE PROJECT

FALL 2019

Timothy Bui

Tony Chen

Brian Ozarowicz

Trevor Webster

Abstract

This document defines the problem statement for the Capstone project Building More Self-Aware Everyday Robots and details the proposed solution to be implemented by the Capstone team under the direction of the established research team, as well as the performance metrics that will be used to gauge successes of the project's tasks in relation to its goals.

Dr. Naomi Fitter desires improvements to a robot comedian's ability to adapt in real-time to audience cues about its performance. The Capstone team will use machine learning on data about joke successes in past performances to improve the robot's ability to determine whether a joke is a hit or a bomb based on the audience's response with at least 85% accuracy in its evaluation.

CONTENTS

1 PROBLEM

The primary problem for this project is to improve the ability of a NAO robot to read audience cues during a stand-up comedy performance and accurately gauge whether a joke was a hit, a bomb, or somewhere in-between. The motivation of this project is to increase the capability of robots to perceive more subtle cues of human interaction and become better able to adapt to them in order to experience more successful interpersonal connections.

The robot performs by telling prewritten jokes and listening for audience laughter, using that feedback to determine whether a joke was well received or poorly received. The robot can then change its follow-up response to the joke based on this feedback. Compared to data collected from past performances which were categorized by hand, the goal is to autonomously determine - with consistent accuracy - which of the three categories of response each joke told falls into. Challenges present in this part of the project include recognizing the difference between human laughter and other background noises, adapting to different acoustic environments, and dealing with different audiences in potentially unpredictable circumstances.

The primary goal of this research project is to produce a robot capable of giving convincing stand-up comedy performances in front of a live audience while adapting to the audience's response in appropriate ways. In a broader sense this work has implications for improving human-robot interactive experiences in many contexts, such as improving the capabilities of AI assistants and improving the ability of autonomous AI systems to react to their environment.

2 PROPOSED SOLUTION

To achieve the desired improvement in the robot's self-awareness the Capstone team will aid the current research team in analyzing a pre-constructed database of audio recordings from past comedy performances.

The initial task of this project will be to divide the overall problem into smaller actionable problems. The first is to extract data that can be analyzed from the audio. In the past work by the existing research team intensity and pitch have been extracted from the mp3 files without any pre-processing but with normalization. Keeping in mind the limitation that the data must be processed, normalized, and sorted in real time by the robot during performances, the group will look into additional ways to effectively extract and process data from the audio. Possible pre-processing techniques such as noise reduction, spectrograms, or logarithmic scaling of magnitude could be added to improve the results of the data. Further possibilities for data extraction might be inspired by research into current music classification techniques.

The second task will be to find the most appropriate machine learning or sorting algorithm to categorize the data. With guidance from Dr. Fitter, our primary focus will be looking into time series approaches such as a Hidden Markov Model or Dynamic Bayesian Network. The group will also be looking into simpler machine learning algorithms such as K Nearest Neighbor, Decision Trees, or Statistical Modelling algorithms such as Logistic Regression. The group will utilize Python utilities such as SciKit and look into the possibility of further utilities such as TensorFlow or PyTorch if they are operable under the hardware limitations of the NAO robot.

For data validation, leave-one-out validation will be our primary focus since it best represents the situation the robot will be in during performances and works well with small datasets. However, other methods such as K-Fold or random subsampling could also be tried for the sake of thoroughness.

The results of the machine learning analysis will be used to update the real time adaptive models of the robot, improving its ability to determine if the audience found a joke funny or not and thereby increasing its ability to adjust its comedy routine in real time and creating a better connection with its audience.

3 PERFORMANCE METRICS

The deliverables for this project are machine learning analysis on the audio data from past comedy performances and application of the learning results into better real time adaptation in the robot's software.

Successful learning on the database will be represented by the production of results that determine a joke to be either a hit - meaning it had a near-unanimous positive reaction from the audience, a bomb - meaning it had negative or no reaction, or a mixed-reaction. This determination will be made by evaluating the audience's laughter, which is isolated from other noise using audio frequency analysis. The accuracy rate which the machine learning evaluation of the jokes should achieve in determining the audience response must be 85% or greater. This is to meet the accuracy level observed by humans manually rating the audience's response to jokes.

There are two sets of audio data, one from the V5 version of the robot and one from the V6. The V5 data has been reviewed by human coders and labeled so it is easier to judge the success of a machine learning approach that is applied to it. The machine learning will be performed on the V5 data first to identify what approaches have higher levels of a successful evaluation. When a consistently successful method is found which performs at the desired accuracy level the learning will be expanded over the full dataset to begin analysis of every recorded joke.

Currently, the robot only begins listening for audience feedback after the joke delivery concludes. An additional potential goal of the project is to add the ability to listen for reactions mid-joke and pause the delivery accordingly if the response is loud or prolonged. Introducing this ability would further improve the interpersonal interaction with the robot as it would behave more like a human comedian, as well as improve the quality of the comedy performance itself by making sure the joke is heard.

4 SECONDARY GOALS

An additional problem we could help address if enough time is available would be to improve the robot's sense of timing by having it wait until the laughter from the previous joke has almost entirely died down before telling its next joke. One method for this would be to assign an arbitrary value below a certain threshold to wait for. Another potential method would be after the robot has finished telling a joke have it wait until the x second interval of volume maintains a steady pitch before proceeding to the next joke. A third potential solution could be to take audio clips from successful

comedians and use a machine learning algorithm such as a neural network to learn the wait time after a joke before telling the next joke based on audience laughter. We could then have the robot use the results of the neural network's training to determine how long to wait after laughter in its own performance before telling the next joke.

Another potential problem we could solve if enough time is available would be to improve the robot's ability to change which jokes it will tell and in which order based on the audience's response to prior jokes during a performance. One potential way of doing this would be to run all jokes through a naive bayesian algorithm with every word in the joke acting as an item in the library. The algorithm would then analyze every joke based on a value assigned to each word in the joke. As each subsequent joke is told it would update the full list of jokes based on how effective the previous joke was, improving the value of words in a joke that was successful while decreasing the value of words in a joke that fell flat. It would also be possible to do this using the same method, but rather than using the words in the joke instead assign each joke a category based on its contents such as 'nerdy' or 'rowdy'. Either way, this would guide the robot toward making jokes that connect with the current audience in some way while steering it away from jokes that tend to fall flat.