MCI project First Milestone Report

Team number: 15
Project Title: Text to Perpetrator: visualising suspect description

Milestone 1	Activities	Planned Outputs	Achieved Outputs
Restate the milestone from your Draft plan.	Restate the key activities from your draft plan.	Restate the planned outputs from your draft work plan.	Outline the actual outputs compared to what was projected (or type "same as planned")
Obtain the training data for machine learning. The training data consist of sentences decrypting human appearance and labels. Build a classified model by utilising pre-existed language models such as RoBERTa in PyTorch. Image generation is done through 3D modelling. Generate more labels and data for more accurate predictions.	Study the semantic structure of the description of human appearances. Then generate the corresponding labelled data with Stacy API in JSON and csv format.	A data generator is made for generating training datasets, which can generate a large amount of data that satisfies our module's conditions in less than 3 seconds.	The current generator can make simple basic datasets and generate more complex datasets, with more sub-labels, sentence patterns, and non-description sentences. These complex datasets can be used for more accurate predictions.
	Appy some data preprocessing techniques in natural language processing. These techniques are punctuation removal, lower casing, tokenization, stemming, and lemmatization.	Text is converted to lowercase and split into smaller units. Words are stemmed or diminished to their root/base form. For example, words like 'programmer', 'programming', 'program' will stem from 'program'.	Lower casing and tokenization and stemming improve the model. However, lemmatization does not improve the model because the words are not diminished to a proper English word.
	Research different language models and be familiar with sequence to sequence and transformer models. Build the data pipeline for training. Employ the RoBERTa in PyTorch and train the model in the Google Cloud Platform.	Expect the accuracy will be archived at around 95%, and the training will be done in less than 2 hours.	The accuracy is 99.6% in the test dataset. The training time takes more than 2 hours. Since we only have $300 free credits on the Google Cloud Platform. This might not be enough to continue our project. We decided to move the training into Google Colaboratory, which is free and comes with a fast enough GPU for our project.
	We tested different learning rates and batch sizes on the models. Hyperparameter tuning and virtualisation of the performance of the model using Tensorboard.	The accuracy varies depending on the learning rate and batch size. We expect a small learning rate and large batch size to optimize performance.	The optimised performance will be a small learning rate and medium batch size.
	Generate more labels and data for more accurate predictions.	Extend categories into more classes and categories. Adding noise by combining the narrator dataset and description dataset.	Same as planned.

	Use the corresponding data to generate an image that closely resembles the description between 10 categories and each might have 1 or 3 classe.	Utilise the categories to generate an image representing the textual description using either 3D models or 2D images. All of which is hosted as a web application that can be accessed only	3D models have proven to be more flexible and easier to manipulate based on the categories. 3D models for all categories have been collected and are ready to be combined. Everything else is as planned.

Team reflection on progress.	Provide some comments below regarding the completion of this milestone, specifically around: 1. How is the project progressing? 2. Are there any differences between projected and actual outputs/outcomes?

Question 1: Zoom meetings, emails and communication apps are our communication methods. We usually meet our client on Monday, at 10 am and then a follow-up team meeting to discuss the feedback and suggestions from our supervisor and exchange our ideas to improve the product that we are developing.

By the end of the first milestones, we had successfully produced the training data composed of 10,000 sentences with labels. The labels consist of 13 categories (e.g. top, bottom, accessory…etc.) and 32 classes (e.g., colour, type and pattern). Then we trained our model with this model, and the model's accuracy achieved 99.7% accuracy in our test data set. Furthemore, the image generation is now able to generate all 2 Gender, 9 Top and 1 Bottom categories with output time of less than 5 seconds.

Question 2: There is no noticeable difference between the projected and the actual outcome.

Team reflection on managing problems	Have you encountered any problems to date? If so, how have you managed them?

We encounter some problems. The first problem we countered is that we do not have the data such as accident reports from the policies due to confidentiality. We decided to generate our dataset using Stacy's framework to overcome this problem. We study the semantic structure of the description of a human and then feed this structure into the frame. We can produce 10,000 sentences from 13 labels sufficient for our machine learning.

The second problem is the machine learning model is not working as intended. The machine model is an end to end process, which means we feed the data in and results come out. However, we have no idea on layers between the input and out, making it hard to debug. To counter this, we take suggestions from our supervisor and implement tensorboard, a tool that allows us to read the transformation layer by layer and step by step. After we fixed the bugs and the model ran successfully, it achieved 99.7% in prediction.

The third problem is the training machine learning model requiring heavy computation. The training process is prolonged for us. If we train on a model on our computer, it takes around 24 hours. Therefore, we adopted Google Colabrary, a cloud computing platform where we can utilize the GPU to conduct our training process. The training time is shortened to 2 hours.

The fourth problem is that we realised that 2D image manipulation is difficult and not flexible. At the same time, the output is not up-to-standard since it is hard to even recognise what the person looks like in the output. Therefore, we changed our approach into using 3D models to generate the image where we can easily manipulate the color, material and even the shape so that it matches with the description.

Supervisor assessment	Please, rate your team (1) effort, (2) project progress and (3) their self-reflection for milestone 1. Rating scale 1-10 as per standard marking scheme, i.e. 5 is a Pass, and 7 is a credit. Add some comments to explain your rating
Effort: Progress: Reflection: yh	