**Intelligent Multi- Domain Chat-Bot**

**Introduction**

Before deep-learning stepped into the field of chat-bots, there were traditional techniques using tf-idf, information retrieval etc. to output appropriate responses for the given context. These bots handled different domains intelligently using techniques like LDA, multi-class classifier etc. Deep Learning based methods, however, is treated as a black box to train model on conversations (usually from social media). Thus, it is very inefficient in detecting different domains a person is conversing. Domain knowledge is very important as a similar context can be present in two different domains. And the responses to these domains will be completely different. Therefore, our project focuses on how we can use DL based seq2seq model and make it smarter by responding to different domains appropriately. Later in future, we might want to track the history of domains a person is interested in and can improve our model iteratively as a user starts interacting with it.

**DataSet**

We started with building a model with 3 domains. The three domain are : Games, Movies and Social. For games and movies domain, we took dataset from Reddit forums. These data were downloaded from Google’s BigQuery. We labelled title/forum-query as a context and their comments as the responses. For Social domain, we downloaded twitter conversation used in our previous homework.

The number of utterances for each domain is as follows:

Games :

Movies :

Social :

**Model (Still in experimental phase)**

We trained three separate domains on the data extracted. The model used is seq2seq. We did early stopping and checked if it answers perfectly. One thing we observed was that in the movie or gaming domain, we usually got single worded answer: E.g. “What is your favorite movie ?” --> “Star wars”. This came from the fact that people in reddit do answer very shortly to the queries. Therefore, we are now thinking to expand our dataset to comment-reply level in the hope to get long responses from chat bot. Some more examples are:

*> PS4 or XBox what should I buy*

*A computer*

*>What can i play in PC games nowadays*

*http : //store . steampowered . com/app/000000/*

*> which film is better : X or Y*

*Both are both good .*

Second example is really good, where the chatbot responds by giving an appropriate link to play games when asked. This way we think it highly depends upon the dataset we use it to train and thus we should try training with long comment-reply pairs.

Apart from this, we have built a tf-idf based classifier to distinguish between different domains. Currently it is set to hard classifier but we are discussing the ideas for changing it to soft classifier.

Our claim is that once we are successful in building a robust model for just 3 mentioned domains, it would be quite simpler to expand it for handling other domains as well.