CMPS 115 Better Bookmarks Sprint 2 Summary

7/14/2019

*Note that in depth Machine Learning information can be obtained by looking at the google colab shared on the github

Our project is a Google Chrome bookmark manager that a user can download from the Google Web Store that displays bookmarks, allows users to add new bookmarks, recommends a bookmark title for new bookmarks, and recommend a bookmark category folder to help them stay organized as they accumulate more content.

The inputs to our ML classification model are the title of the webpage, and the description taken from the HTML. The output is a category chosen from 5 of our top performing categories. We used a two layer LSTM with a trainable word embedding layer and softmax output as our NN architecture. Word embeddings represent words in a semantic space (100d in our case) where similar words reside close together in the space. We used word embeddings from the Stanford ML Group trained on twitter. As an example of word vectors the words “Queen” and “King” would be close together in the vector space since they are both royalty. The word embeddings also have the property where for example you can subtract “Queen” - “woman” and yield “King” or maybe a non-binary royalty. The Long Short Term Memory cell is the foundation of our model, it provides a way to incorporate input from previous words and the current word to represent some semantic meaning at the current time-step that is then passed to the next time step. Each time step and specific words can be learned to be remembered or forgotten. This NN architecture seems suitable for our problem because it allows longer term word dependencies to remain, whereas as other ML models like decision trees or SVM’s would require an averaging of the word vectors and would not be able to store any sequence information. Right now we are using a classification model trained on a centralized dataset of user bookmark information. As far as customizability goes, using a clustering method on each persons personal bookmark data would provide a much more personal way to formulate categories, which is a feature we have planned for sprint 3.

The current functionality of our add-on allows users to view their bookmarks, add a new bookmark, and automatically recommend a category for them to put their current webpage into. The automatic category suggestion helps to keep the users bookmarks organized and enables them to find their bookmarks much quicker.