Semantically-related Data Ordering for Neural Network Models

Natalia Wojarnik, Cutter Dalton
Final Project Presentation
Neural Networks and Deep Learning
Fall 2022

University of Colorado Boulder

What and why?
Motivations and Improved Solutions

Recent Developments in machine learning have shown significant improvements in the development of complex and robust models capable of solving multiple NLP. More complicated models demand larger datasets and stronger computational power. Recently, focus has shifted towards data quality in an attempt to eschew more cumbersome models. This project observes the change in performance in models when ordering data by semantic-relatedness during training.

Data-ordering is the intentional grouping of data points when feeding data into the model. In this experiment we observe the efficacy of ordering data by semantic-relatedness where data points (in this case, news pieces) are grouped together by topic, and then fed into a model which predicts the news category.

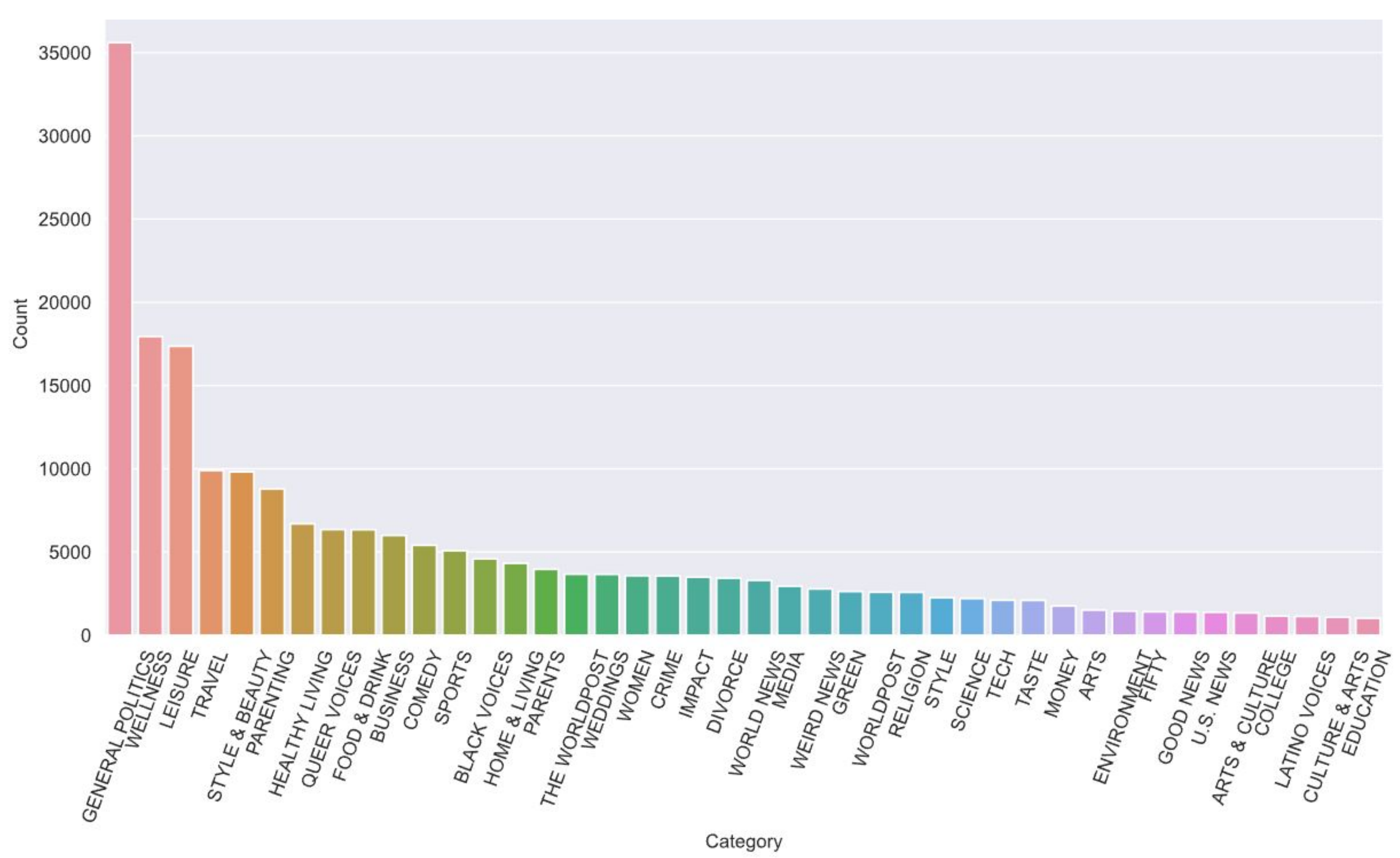
What inspired us?
Related Work

The idea that humans learn faster and more efficient when the examples given are clustered together in some kind of semantic way. The idea of data ordering is not new to the field of NN and DL but it's mostly used in Curriculum Learning approach. We wanted to examine the human-inspired semantic clustering of the training data and its impact on the model's performance.

What did we work on?
Data and Modality

News Category Prediction

Around 210k news headlines from 2012 to 2022 from HuffPost.



Classes to predict in news category dataset. Broken down further into 9 superclasses.

How did we approach this?
Data Orderings

Alphabetic Clustering

- ordering by superclass alphabetically - EXPERIMENTAL ORDER 1
- ordering by class alphabetically - EXPERIMENTAL ORDER 2

Semantic Clustering

- ordering by class hierarchically; descending number of texts assigned to one superclass - EXPERIMENTAL ORDER 3 (the opposite of Curriculum Learning)
- ordering by class hierarchically; ascending number of texts assigned to one superclass - EXPERIMENTAL ORDER 4 (Curriculum Learning approach)
- ordering by class hierarchically; shuffled within one superclass- EXPERIMENTAL ORDER 5 (the opposite of Curriculum Learning)

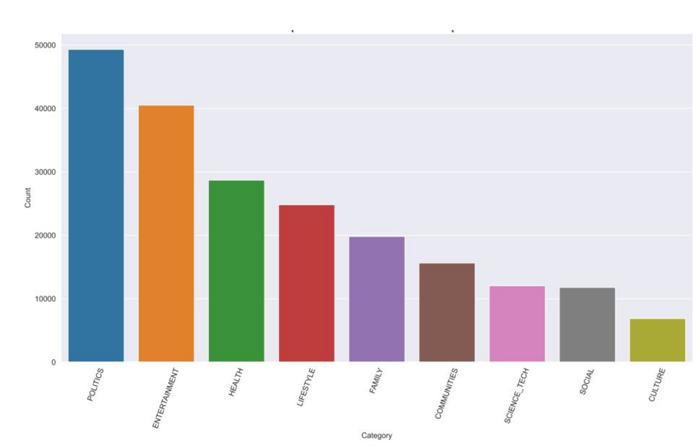
How did we build the models?
Methods and Architecture

Experiments were conducted on 2 different models (below) and with the use of 5 different Data Ordering and Clustering techniques (above). The architecture of the models were chosen as the ones used to establish the performance on the randomly shuffled data. We treat the randomly shuffled data as a baseline for Model1 and Model2. Model architecture was not the main focus of our project, although Model2 performed significantly better.

Model 1- Simple RNN

Embedding (70)
Bidirectional Simple RNN (64, dropout = 0.1, recurrent_drpout = 0.20, activation = Tanh)
Bidirectional Simple RNN (64, dropout = 0.1, recurrent_drpout = 0.20, activation = Tanh)
Simple RNN (32, activation = Tanh)
Dropout (0.2)
Dense (num_classes, activation = Softmax)

Model 1 architecture.



Superclasses manually created.

Model 2 - LSTM, GRU

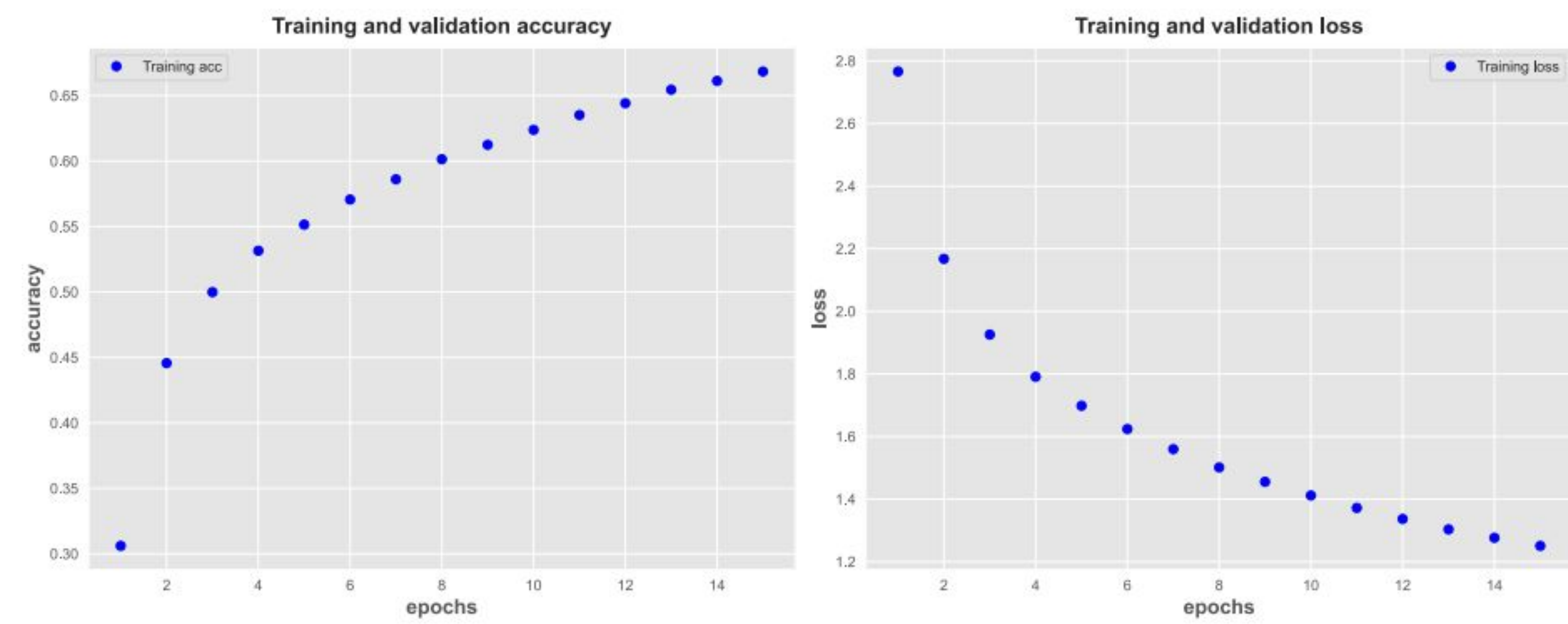
Embedding (100)
Bidirectional Simple LSTM (64, dropout = 0.1, recurrent_drpout = 0.10, activation = Tanh)
Bidirectional Simple LSTM (64, dropout = 0.2, recurrent_drpout = 0.20, activation = Tanh)
Bidirectional Simple RNN (64, dropout = 0.2, recurrent_drpout = 0.20, activation = Tanh)
Conv1D (72, 3, activation = ReLU)
MaxPooling1D (2)
Simple RNN (64, activation = Tanh, dropout = 0.2, recurrent_dropout = 0.20)
GRU (64, recurrent_dropout = 0.20, recurrent_regularizer = L1_L2)
Dropout (0.2)
Dense (num_classes, activation = Softmax)

Model 2 architecture.

What did we notice?
Results

	Baseline	ORDER 1	ORDER 2	ORDER 3	ORDER 4	ORDER 5
Model 1	49 %	66.31%	66.77 %	66.69 %	**67.37 %**	66.92 %
Model 2	50 %	75.60 %	76.39 %	77.08 %	**78.61 %**	76.89 %

Evaluation of the models trained on the news headlines with experimental data odredinds.



Accuracy and loss of the best performing data ordering - ORDER 4, Model1.

What did we learn?
Conclusions

- All the models trained on the experimental data orderings did better than the baseline models with randomly shuffled training data examples.
- The best performing data ordering (ORDER 4) was semantically clustered within one superclass and also ordered in the ascending order where the first superclass had the most number of headlines and the last one - the least number of headlines.
- Second best model was also clustered within one superclass but surprisingly, the data examples were shuffled within one cluster.
- ORDER 1 had the poorest performance. This data ordering was not clustered semantically but only alphabetically to group together same news headlines within one superclass. Semantically clustered data did better than alphabetically ordered.
- Curriculum Learning seems to be the key component to data ordering techniques bringing best results to the models.
- LSTM and GRU layers improved the performance and helped to process long sequences and preserve the features of previous input at each time step.
- Further experiments with data ordering and clustering combined with the Curriculum Learning approach are planned to be conducted to explore the presented phenomenon.

References

Chang, E., Yeh, H.-S., and Demberg, V, Does the order of training samples matter?, improving neural data-to-text generation with curriculum learning, 2021

Avikumart (2022) Word Embeddings & RNN [Source Code] https://github.com/avikumart/Natural-language-processing.

Heo, H.-S., Jung, J.-w., Kang, J., Kwon, Y., Kim, Y. J., Lee, B.-J., and Chung, J. S. Self-supervised curriculum learning for speaker verification, 2022