Modeling Indoor/Outdoor Images Using Sentence Embeddings

Anisha Babu[1]

[1] University of Oregon

Modeling Indoor/Outdoor Images Using Sentence Embeddings

Reproducibility:

Link to the GitHub repo: https://github.com/anishababu62442/EDLD654-Final

Research problem:

As we navigate the world, we continuously form memories of similar events. However, similarity between memories can lead to interference (Mensink and Raaijmakers (1988); O'Reilly and McClelland (1994)). For example, remembering a password for one account is often complicated by interference from memories of other passwords. Although there have been many studies on memory interference, they have primarily focused on the fact that similarity reduces the probability of successfully remembering an event. Meanwhile, an under-explored question is whether similarity changes memories. On the one hand, similarity may lead to integration of similar memories, such that two memories blend together. On the other hand, similarity may lead to repulsion, wherein information that differentiates similar events becomes prioritized in memory (Hulbert and Norman (2015)) or even exaggerated (Chanales, Tremblay-McGaw, Drascher, and Kuhl (2021)). For example, let's say you visited the same restaurant on two different days: a sunny day in August and a breezy day in September. If memory integration occurs, you may blend details of the events (a sunny day in September). If memory repulsion occurs, your memories may prioritize details that differed across the two events (the weather) or even exaggerate differences (a particularly windy day in September). This example illustrates that similarity between events may shape how they are remembered. Critically, these changes, or distortions, in memory content are potentially systematic and, therefore, predictable.

Ultimately, I will test for similarity-induced changes in memory content using innovative behavioral methods. Specifically, I will use Natural Language Processing (NLP) techniques to determine whether and how similarity induces changes in memory content.

My overarching hypothesis is that similarity between memories will lead to predictable distortions in memory content.

In this paper, I will assess preliminary data that will determine the feasibility of using NLP in a full-scale study. NLP is a powerful tool for transforming natural language into numerical vectors that represent semantic information across hundreds of dimensions. Translating memories into these numerical feature spaces will allow me to mathematically express content similarity between individual memories. However, in order to measure how similar memories change, I must first establish a baseline to which written memory can be compared.

The stimuli that were used in this preliminary experiment will be used in subsequent memory experiments as well. They consist of a set of naturalistic scene images, with 15 exemplars from 30 categories (e.g., 15 beaches, 15 airports, 15 libraries, etc.). Critically, 15 categories were indoor scenes, and 15 categories were outdoor scenes. In this experiment, participants viewed one image from each category, and wrote a description underneath in at least 15 words. This experiment did not involve a memory task and participants were not exposed to any similar images (i.e., no images from the same category). The purpose of this experiment is to generate baseline vectors using NLP that represent the content of each scene image. Subsequent experiments will compare remembered content to these baseline vectors. See below some sample responses:

Image	Description
Ice-skating rink	Ice rink filled with families, children and adults. Two men, and two children hold hands in a row of four. It is a least chilly with many wearing hats, and warm coats.
Library	a library, which is similar to that of trinity college belfast, rows of brown bookshelves filled with old leather bound books. at the end of the corridor a large partly stained glass window. The ceiling is highly decora
Indoor pool	An almost empty, modernist indoor pooL with sliding glass windows around the entire building and multiple skylights. Four beach chairs at the end of the pool, to the right a gie
Train station	Train pulls into a staion, the time is just after 6.10pm, a woman in a mask looks at the train. The station is filled with one other person, sat on a bench. It appears to be a fairly old station metal
Arcade	Arcade, lit only by the various machines, these are mostly ball games for prizes instead of the usual video games often depicted. It has a styrofoam ceiling. The lights are purple, blue, and green in neon.

At the end of the experiment, I had a set of five descriptions for each of the 450 images (30 categories x 15 exemplars). Each of these descriptions was transformed into numerical vectors using the NLP algorithm MPNet (Song, Tan, Qin, Lu, and Liu (2020)). In this paper, I will assess the specificity of these vectors using machine learning techniques. Specifically, I seek to classify descriptions as corresponding to either an indoor or outdoor image. This analysis will reveal the ability to extrapolate semantic information from MPNet output. Rather than classifying descriptions into specific categories (which

are often explicitly written in descriptions), classifying indoor/outdoor images will allow for an interesting assessment of NLP techniques more generally. If successful, this would indicate NLP is a feasible method for quantifying written memory in future experiments. It would be especially interesting to see if a model trained on this baseline data is successful in classifying data from a memory experiment.

Description of the data:

As described before, there were a total of 30 scene image categories, each with 15 images. Critically, half of these categories were outdoor scenes, and half of these categories were indoor scenes.

Using separate Python code, written descriptions were transformed into 768-dimension numerical vectors using MPNet. Each description also includes a marker for whether the description corresponds to either an indoor or outdoor image. See first seven columns of sample data:

Location	V1	V2	V3	V4	V5	V6
Indoor	-0.0630010	-0.0864007	0.0056232	0.2349402	0.0210362	0.0293891
Indoor	0.0555179	0.0459704	0.0075291	0.0934091	-0.2211835	0.1114800
Indoor	-0.1413176	-0.2072940	0.1051668	-0.0065399	-0.0954954	-0.0198655
Outdoor	-0.1735294	0.0211764	0.0929864	0.1799997	0.0699336	-0.0089860
Indoor	0.0804916	-0.0159084	0.0089862	0.2014005	-0.1113927	0.0549167

See below for count of indoor/outdoor images:

Var1	Freq
Indoor	1125
Outdoor	1125

Description of the models:

The first modeling approach I used is logistic regression. I chose this model rather than linear regression because I have a binary outcome (indoor/outdoor), and the model assumptions require a binary outcome. We can also assume the 768 variables outputted from MPNet are independent of one another. To select a model version, I assessed a model without penalty, a model with ridge penalty, and a model with lasso penalty. The cutoff for predictions is: probability > 0.5. For the ridge penalty and lasso penalty versions, multiple lambda values were assessed to find the best fit. The best performing model will be chosen based on measures of: area under the curve, accuracy, and precision.

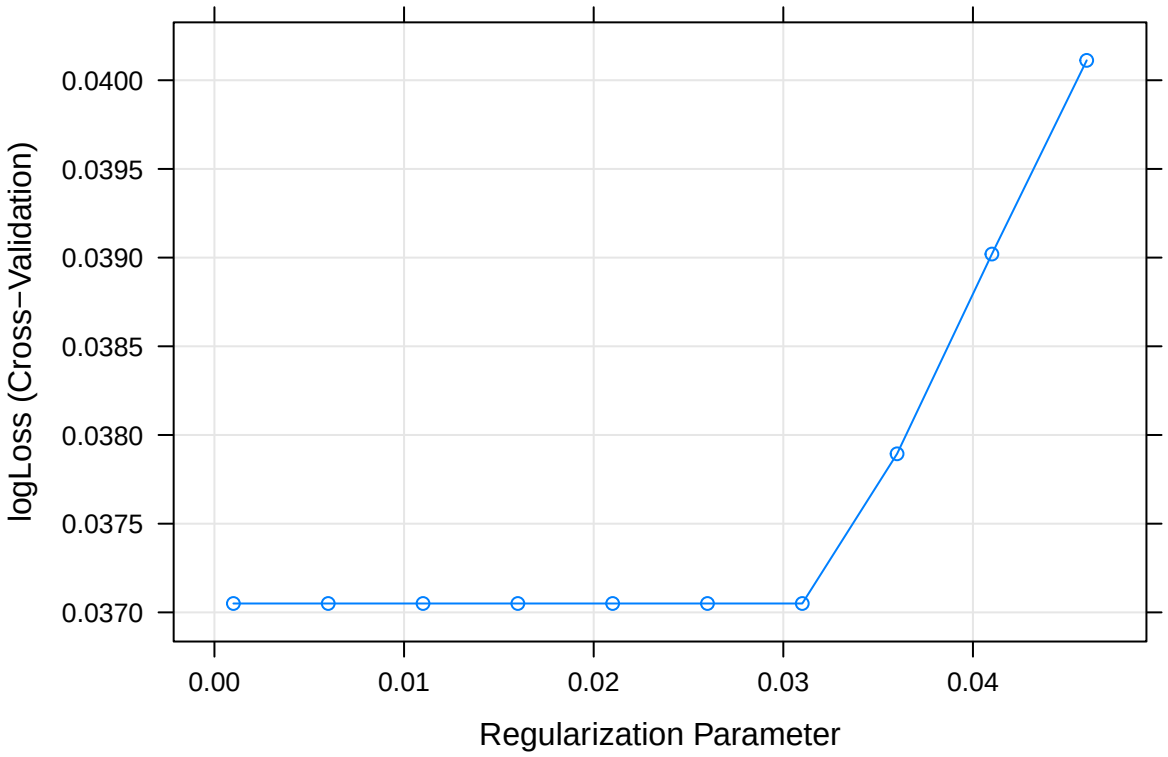
The second modeling approach I used is a decision tree. NLP output includes hundreds of dimensions representing semantic information. However, it is unlikely that all dimensions are relevant in determining if an image was indoor or outdoor. As such, a decision tree is useful in that it selects the most important variable dimensions, and classifies samples based on values along those important dimensions. The cutoff probability for predictions is: probability > 0.5. To find the best complexity parameter, multiple values were assessed. The model will be assessed based on measures of: area under the curve, accuracy, and precision.

The third modeling approach I used is a bagged tree model. This model randomly selects from the 2250 samples and makes aggregate predictions for multiple different trees. The cutoff probability for predictions is: probability > 0.5. To select the best number of trees, I assessed different numbers from 1-200. The model will be assessed based on measures of: area under the curve, accuracy, and precision.

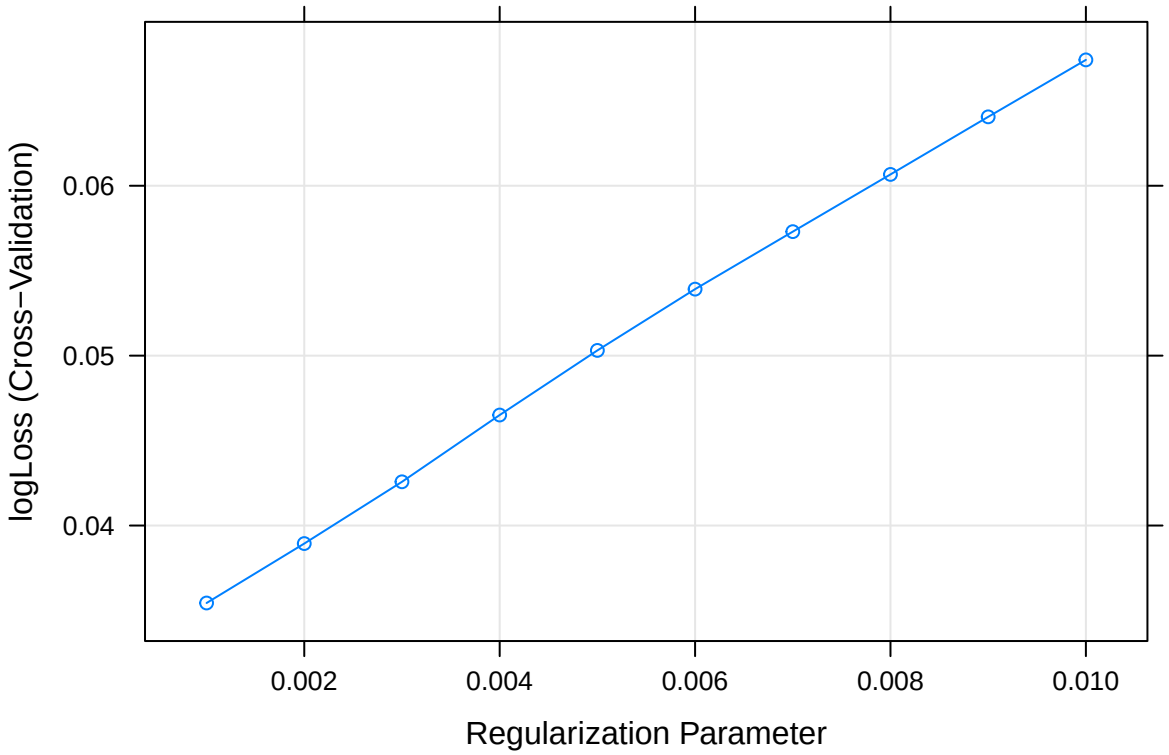
Model fit:

For the logistic regression model, I tried three different versions: without penalty, with ridge penalty, and with lasso penalty. For ridge penalty, I tested multiple lambda

values as seen in the figure below.



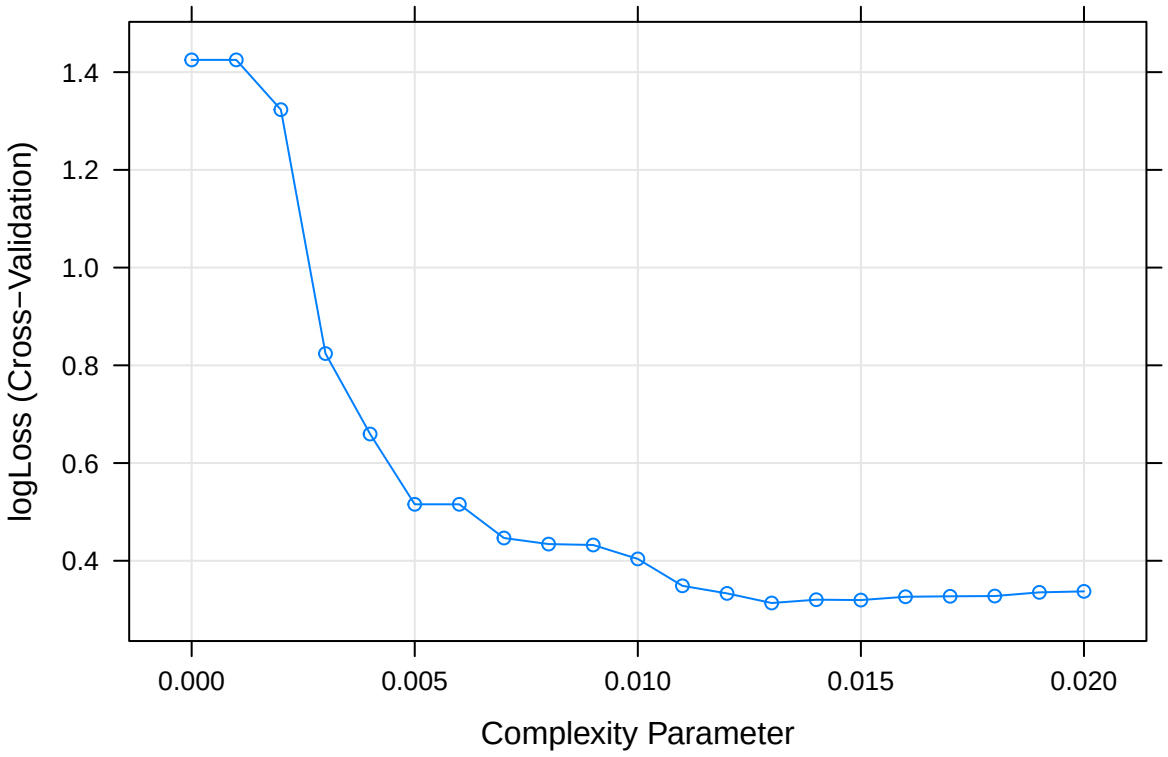
For lasso penalty, I tested multiple lambda values as seen in the figure below.



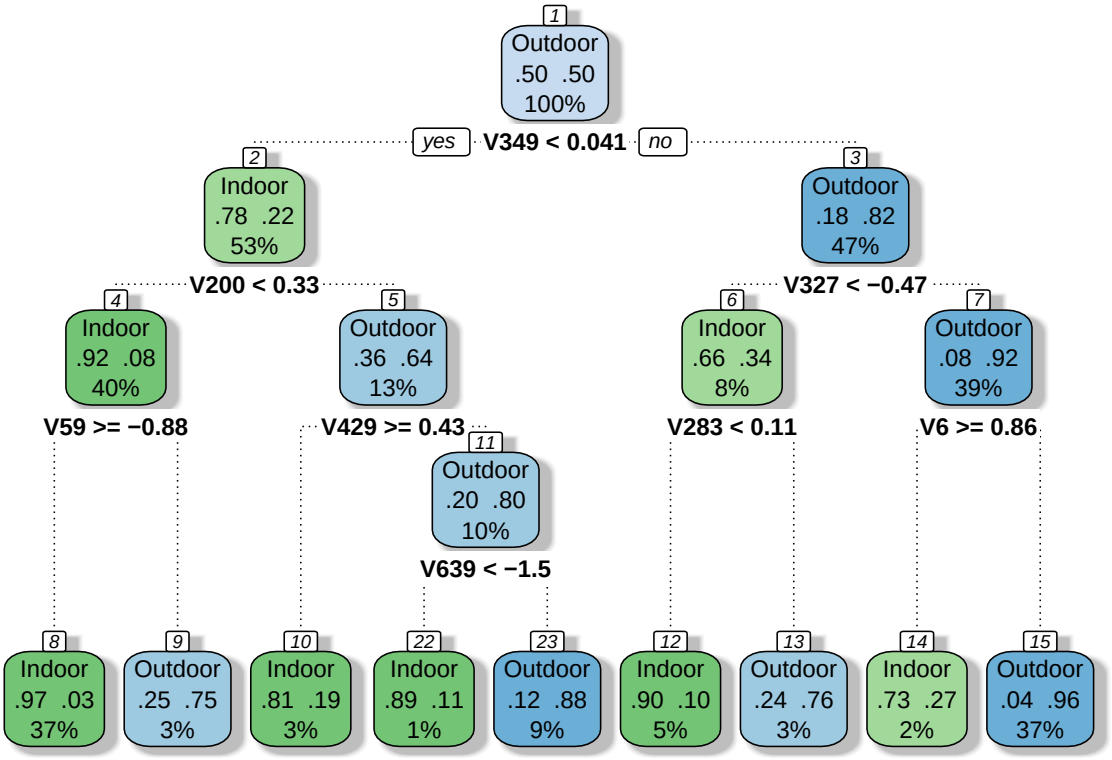
I assessed performance for each version by calculating area under the curve, accuracy, and precision (summarized in table below). Based on these results, I would choose logistic regression with ridge penalty, as it had the highest accuracy and precision, and close to the highest area under the curve. Strikingly, all models have very high performance overall.

	LL	AUC	ACC	TPR	TNR	PRE
Log Reg	0.1740108	0.9994863	0.9911111	0.9954955	0.9868421	0.9866071
Log Reg with Ridge	0.0370500	0.9994666	0.9955556	1.0000000	0.9912281	0.9910714
Log Reg with Lasso	0.0354410	0.9994468	0.9911111	0.9954955	0.9868421	0.9866071

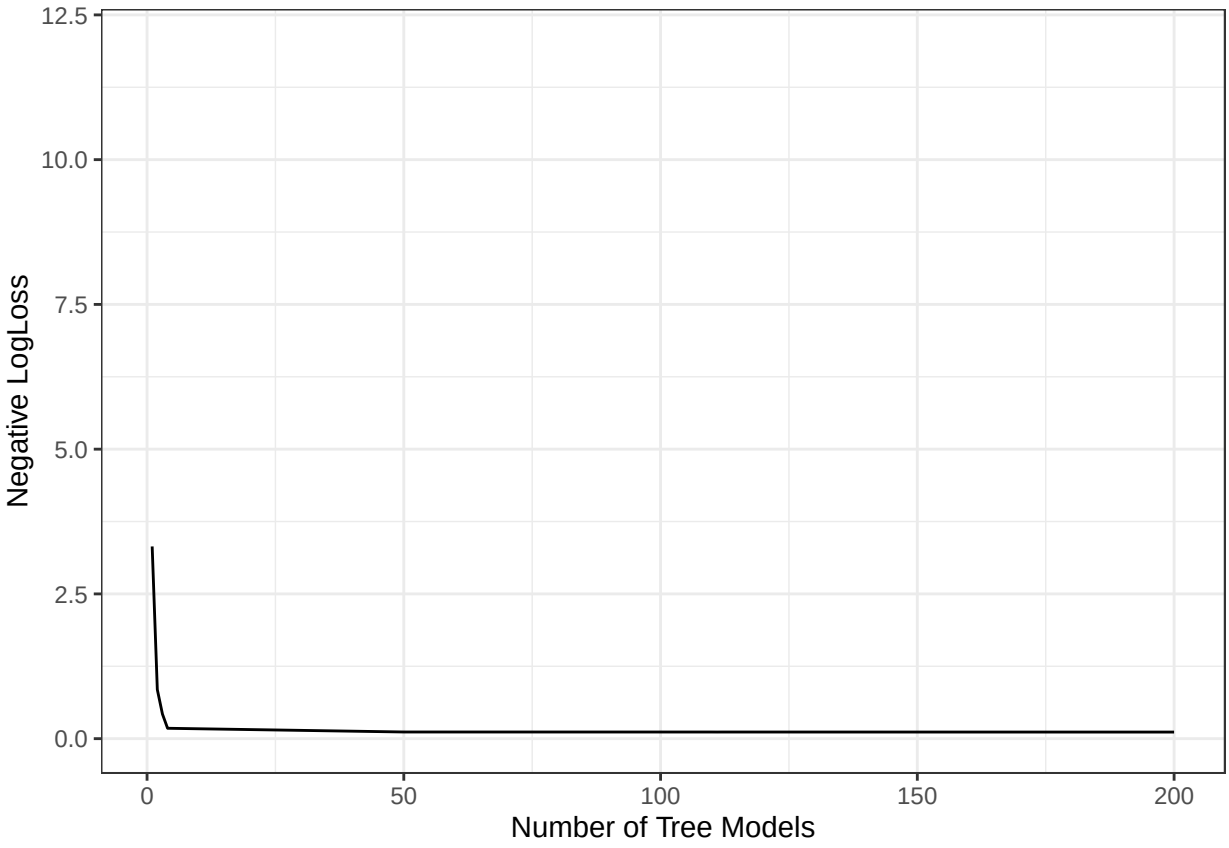
For the decision tree model, I tested multiple complexity parameter values as seen in the figure below.



See figure below for diagram of decision tree. I also assessed performance for the decision tree model by calculating area under the curve, accuracy, and precision (summarized in table at the end).



For the bagged tree model, I tested multiple number of tree values (1-200). Due to high computational time, the figure below shows only a few sample values. I also assessed performance for the bagged tree model by calculating area under the curve, accuracy, and precision (summarized in table at the end).



The table below shows results for all three models considered based on measures of area under the curve, accuracy, and precision. From these measures, logistic regression with ridge regression performs the best, followed by the bagged tree model, then the decision tree model.

	LL	AUC	ACC	TPR	TNR	PRE
Log Reg with Ridge	0.0370500	0.9994666	0.9955556	1.0000000	0.9912281	0.9910714
Decision Tree	0.3135789	0.9416390	0.9000000	0.9144144	0.8859649	0.8864629
Bagged Tree	0.1141214	0.9984491	0.9844444	0.9954955	0.9736842	0.9735683

Discussion/Conclusion:

It was interesting to find that, despite using more complex models like decision trees and bagged trees, logistic regression still had the best performance. Since the variables used were NLP output, we do not actually know what information specific variables represent.

As such, the variables used in the decision tree are interesting in that they show which of the 768 NLP output values may represent indoor/outdoor information. The decision tree performed noticeably poorer compared to the logistic regression and bagged tree models. This may be because decision tree models are better suited to categorical values, and when used here had to be limited whether the variable is greater than or less than some value. These findings are useful in showing that overall, there was quite high accuracy in decoding whether images were indoor or outdoor. This suggests the written image descriptions are descriptive enough that they may prove useful in a future memory experiment.

References

Chanales, A. J., Tremblay-McGaw, A., Drascher, M., & Kuhl, B. A. (2021). Adaptive repulsion of long- term memory representations is triggered by event similarity. *Psychological Science*, *32*, 705–720.

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994–4008.

Mensink, G. M., & Raaijmakers, J. G. (1988). The production of social capital in US counties. *Psychological Review*, *95*, 434–455.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*(6), 661–682.

Song, K., Tan, X., Qin, T., Lu, J., & Liu, T. (2020). MPNet: Masked and permuted pre-training for language understanding. *ArXiv*.