Predicting Brain Activity from Tags

COSC 174: Machine Learning and Statistical Data Analysis

Final Project Proposal

Jessica Thompson

April 12, 2011

1 The Problem

The goal of this project is to be able to predict brain activity evoked by music given text descriptors of the presented musical stimuli. More precisely, the task is to predict blood oxygen level-dependent (BOLD) signal values in the superior temporal sulcus (STS), an area of the brain known to be implicated in auditory categorization, given human-made tags of the music that was presented to the subjects while their brains were scanned using functional magnetic resonance imaging (fMRI). I would like to be able to use the trained model to predict brain activity for a novel combination of tags. This model is depicted in Figure 1. I will evaluate the model in a leave-two out retrieval experiment similar to that used in[1]. In this evaluation paradigm, two brain volumes are left out of the training phase: a target and a distractor. The trained model is then used to predict brain activity for the left out volumes. A hit is recorded if the predicted target brain is more similar to the true target brain than the predicted distractor brain. If we are able to accurately select the target brain significantly more often than 50% of the time, we can conclude that our model is performing better than chance.

2 Methodology

There exists a large corpus of work on automatic text annotation (i.e tagging) of several different types of data (e.g. images, music), but there's limited work that attempts to go in the opposite direction: synthesizing data from tags. For the project I propose it's necessary to learn a joint model using both tags and

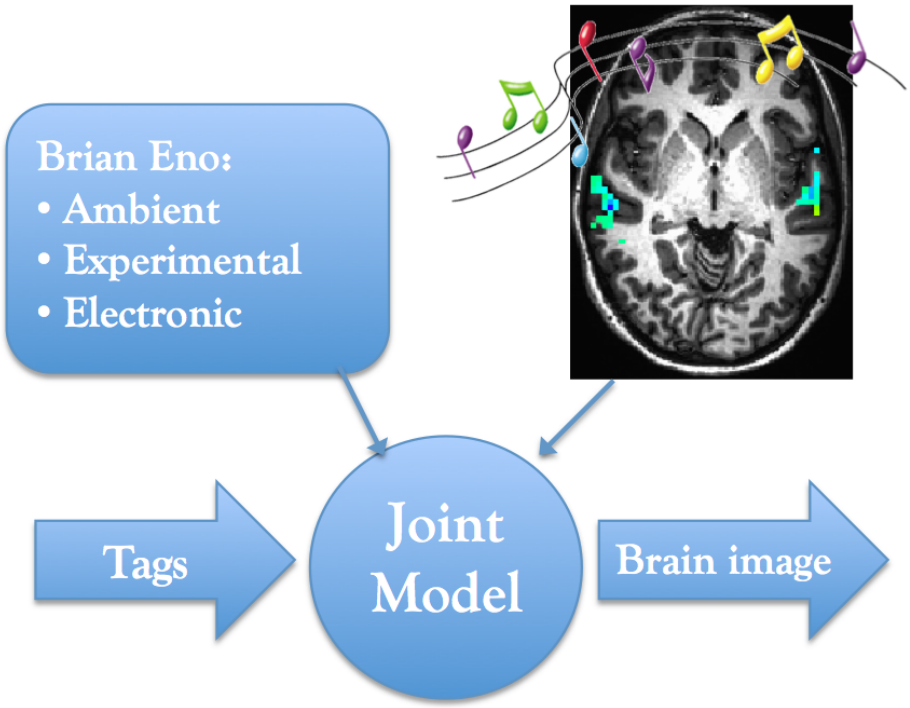

Figure 1: Brain activity evoked by music is combined with tags that describe the music to learn a joint tag-brain image model that can be used to predict brain activity to a list of tags.

brain images such that one can be predicted from the other. I intend to borrow methodology from recent work by Weston et al. in which they learn a joint word-image model from annotated images with the end goal of automatically annotating novel images. They rank the possible annotations of a given image such that the highest ones best describe the semantic content of the image. This is represented by the following model:

$$f(x) = \Phi_W(i)^T \Phi_I(x) = W_i^T V x \tag{1}$$

where $\Phi_I(x)$ is the mapping from image feature space to the joint space, $\Phi_W(i)$ is the mapping for words, and the possible annotations i are ranked according to the magnitude of $f_i(x)$ in descending order.

The authors of [2] use this model to automatically annotate a large database of 10 million images and 100 thousand annotations. My application differs from theirs in that I want to predict images instead of tags, my images are 3-dimensional instead of 2-dimensional, and my dataset is much smaller (only 600 images per subject). However, I don't expect any of these differences to be problematic. The joint-model should be able to predict either tags or images, the

3-dimensional brain data can be represented as a vector, just like the 2D image, and there is nothing to suggest that the model won't work with a smaller dataset. Another difference is that the authors use bags-of-visual terms (a.k.a bag-of-words) features to represent their images while I intend to use raw BOLD values from STS. One could argue that bag-of-words-like features are inappropriate for brain activity which is not necessarily composed of discrete objects and these types of features are perhaps not necessary in this case since I'm not dealing with millions of images. Raw BOLD signal has been used successfully in machine learning applications in the past [3].

3 Dataset

The dataset consists of BOLD signals that were recorded via fMRI while 15 subjects listened to 25 clips of music from 5 different musical genres: Ambient, Symphonic, Country, 50s Rock and Roll, and Heavy Metal. The 25 clips were all 6 seconds long and were each presented 8 times. During the presentation of the music, a complete brain volume was collected every 2 seconds (3 per stimulus presentation). This results in 600 brain volumes per subjects to use for training the model. Given the challenges of combining data from different subjects, I intend to perform all training and testing within each individual subject. Performance measures can then be averaged across subjects. Previous work with this dataset has shown that musical style can be automatically predicted from brain activity evoked by music [4]. Reformulating the problem in terms of lists of tags rather than genre labels allows us to capture more complex information and negates some of the over-simplification inherent to genre labels.

Tags will come from Pandora, last.fm, and the EchoNest. Last.fm provides tags with probability values, based on data from millions of users. Pandora has music annotated by human experts. EchoNest mines the web to find word-track co-occurances. From all of these sets of tags, I will select a subset of the most popular tags to be used as my dictionary for the joint word-brain image model.

4 Expected Timeline

By the milestone deadline I will have collected and organized and chosen a subset of tags to be the dictionary of terms used to train the joint model. This will require an analysis of the most popular tags across all genres and within each genre. I also expect to have implemented the weighted approximate-rank pairwise loss optimization function from Weston et al. (2011) in python. Training and evaluation will happen after the milestone submission.

References

[1] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, and M.A. Just. Predicting human brain activity associated with the

meanings of nouns. *Science*, 320(5880):1191–1195, 2008. pdf.

[2] Jason Weston, Samy Bengio, and Nicolas Usunier. Wsabie: Scaling up to large vocabulary image annotation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2011. pdf.

[3] N. Staeren, H. Renvall, F. De Martino, R. Goebel, and E. Formisano. Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19(6):498–502, March 2009. pdf.

[4] Michael Casey, Jessica Thompson, Olivia Kang, and Thalia Wheatley. Population codes representing musical timbre for high-level fMRI categorization of music genres. In *Springer Lecture Notes on Artificial Intelligence (LNAI) - Survey of the State of The Art Series*. Springer, 2012, In press. pdf.