Predicting Brain Activity from Tags

COSC 174: Machine Learning and Statistical Data Analysis

Final Project Report

Jessica Thompson

May 30, 2011

1 The Problem

The goal of this project was to be able to predict brain activity evoked by music given text descriptors of the presented musical stimuli. More precisely, the task was 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). Ultimately I wanted to be able to use the trained model to predict brain activity for a novel combination of tags. This task is similar to problems such as automatic image or audio annotation, so I decided to borrow methods from these areas to address my problem [1].

2 Dataset

2.1 Brain Activity

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, all training and testing is performed within each individual subject. Performance

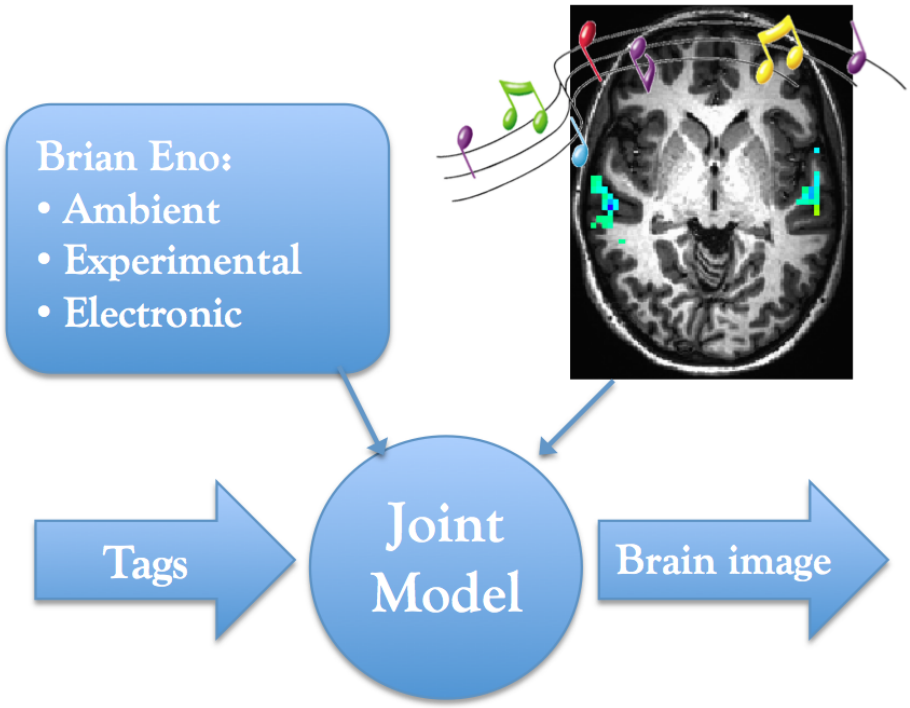

Figure 1: Brain activity evoked by music is combined with tags that describe the music to learn a joint tag-brain image model that can be used to predict brain activity to a list of tags.

measures are then averaged across subjects. Previous work with this dataset has shown that musical style can be automatically predicted from brain activity evoked by music [2]. 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. Raw BOLD signal has been used successfully in machine learning applications in the past [3].

2.2 Tags

Tags were pulled from last.fm using a hierarchical search. Tags were collected in the following order:

1. Tags for the exact track
2. Artist tags
3. Tags of similar artists

Track	Top Tag
Brian Eno - A Clearing	Ambient
Brian Eno - Theme from 'Creation'	Ambient
Eno, Moebius & Roedelius - Old Land	Cool trip
Galerie Stratique, Horizons Lointains	Space Ambient
Anugama - IO-Moon of Jupiter	Ambient
Elvis Presley - Jailhouse Rock	Rock n roll
Bill Haley - Shake Rattle and Roll	Rock n roll
Little Richard - Bama Lama Bama Loo	Rock n roll
Ritchie Valens - Come On Let's Go	Rockabily
Eddie Cochran - Money Honey	Morose Deep Dilate Crimson
Ozzy Osbourne - Fire in the Sky	Heavy Metal
Judas Priest - You've Got Another Thing Coming	Heavy Metal
Metallica - Of Wolf & Man	San Francisco
ACDC - You Shook Me All Night Long	Rock
Scorpions - Rock You Like A Hurricane	Hard Rock
Beethoven - Symphony No. 9 Mvt. 2	Classical
Tchaikovsky - Symphony No. 4 Mvt. 4	Classical
Sibelius - Symphony No. 2 Mvt. 4	Classical
Schubert - Symphony No. 5 Mvt. 1	Classical
Beethoven - Symphony No. 6 Mvt. 1	Classical
Waylon Jennings - Are You Sure Hank Done It This Way	Country
Willie Nelson - Me and Paul	Country
Merle Haggard - Pancho and Lefty	Country
Hank Williams Jr - Whiskey Bent and Hell Bound	Singer-songwriter
Willie Nelson - Welfare Line	Country

Figure 2: Top Tags.

This search procedure helped to ensure that several tags were collected for each track, even when last.fm had few or no tags at all for the exact track. This process returned 1499 unique tags. However, most of these only occurred once. 385 appeared for more than one track. 208 appeared for more than 2 tracks.

The most popular tag, "classic rock", appeared for 18 out of the 25 tracks. I selected the top 100 most popular last.fm tags across my 25 stimuli to be my dictionary. These tags are shown in Figure 3.

Dictionary of Tags				
country	classical	rock and roll	ambient	rock
heavy metal	rockabilly	hard rock	rock n roll	classic rock
oldies	50s	electronic	metal	meditation
new age	classic country	outlaw country	instrumental	romantic
downtempo	composers	singer-songwriter	cool trip	thrash metal
symphony orchestra	san francisco	psytrance	morose deep dilate crimson	of wolf and man
s and m	folk	experimental	electronica	80s
americana	classic	willie nelson	lo-fi	catchy
american	beautiful	rockin	dark	piano
easy	cowboys	meditative	ballad	dramatic
60s	world	southern rock	marchosa	perficta
500	rock roll	indie	chillout	male vocalists
artful	moral	tough guys	canadian	paul
art rock	supernatural	russian	rock'n'roll	chill
german	country rock	a15	dark ambient	hell yeah
willie and waylon	latin	supercla	50s pop and rock	indian
krautrock	under 2000 listeners	yeah	less than 200 listeners	psychedelic trance
astronomic entities	soundscape	true ambient	drjazzmrfunkmusic	names
british	scorpions	nwobhm	elvis presley	elvis
blues	idm	alternative	hank williams	70s

Figure 3: Dictionary of 100 most popular last.fm tags

3 Model

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 this project I propose learning a joint model using both tags and brain images such that one can be predicted from the other, borrowing 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 images.

They rank the possible annotations of a given image such that the highest ones best describe the semantic content of the image [6]. This is represented by the following model:

$$f(x) = \Phi_W(i)^T \Phi_I(x) = W_i^T V x \quad (1)$$

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.

4 WARP Loss Optimization Algorithm

This joint model is trained using the Weighted Approximate Ranked Pairwise (WARP) loss optimization algorithm. This algorithm learns mapping matrices W and V by repeatedly choosing one negative label and one positive label for a random image and then taking a gradient step to minimize the error function of the rankings generated by the model. Pseudocode of this model is shown in Figure 4.

Algorithm 1 Online WARP Loss Optimization

Input: labeled data (x_i, y_i), $y_i \in \{1, \dots, Y\}$.
repeat
  Pick a random labeled example (x_i, y_i)
  Let $f_{y_i}(x_i) = \Phi_W(y_i)^\top \Phi_I(x_i)$
  Set $N = 0$.
  repeat
    Pick a random annotation $\bar{y} \in \{1, \dots, Y\} \setminus y_i$.
    Let $f_{\bar{y}}(x_i) = \Phi_W(\bar{y})^\top \Phi_I(x_i)$
    $N = N + 1$.
  until $f_{\bar{y}}(x_i) > f_{y_i}(x_i) - 1$ or $N \geq Y - 1$
  if $f_{\bar{y}}(x_i) > f_{y_i}(x_i) - 1$ **then**
    Make a gradient step to minimize:
      $L(\lfloor \frac{Y-1}{N} \rfloor)|1 - f_y(x_i) + f_{\bar{y}}(x_i)|_+$
    Project weights to enforce constraints (2)-(3).
  end if
until validation error does not improve.

Figure 4: WARP pseudocode.

The WARP algorithm [4] minimizes the following error function which can be written as a function of the mapping matrices W and V.

$$err = L(\lfloor \frac{Y-1}{N} \rfloor)|1 - f_y(x_i) + f_{\bar{y}}(x_i)| = L(\lfloor \frac{Y-1}{N} \rfloor)|1 - W_y^T V x_i + W_{\bar{y}}^T V x_i| \quad (2)$$

From this we can calculate the partial derivatives with respect to each variable

to compute the gradient:

$$\frac{\partial err}{\partial W_y^T} = L(\lfloor \frac{Y-1}{N} \rfloor)(-Vx_i)$$

$$\frac{\partial err}{\partial W_{\bar{y}}^T} = L(\lfloor \frac{Y-1}{N} \rfloor)(Vx_i)$$

$$\frac{\partial err}{\partial V} = L(\lfloor \frac{Y-1}{N} \rfloor)(-W_y^T x_i^T + W_{\bar{y}}^T x_i^T)$$

These partial derivatives define the update rules for the stochastic gradient descent on the error function to be implemented in code:
$V \leftarrow V - L(\lfloor \frac{Y-1}{N} \rfloor)(-W_y^T x_i^T + W_{\bar{y}}^T x_i^T)$
$W_y \leftarrow W_y - L(\lfloor \frac{Y-1}{N} \rfloor)(-Vx_i)$
$W_{\bar{y}} \leftarrow W_{\bar{y}} - L(\lfloor \frac{Y-1}{N} \rfloor)(Vx_i)$

Both W and V are regularized such that

$$V_i \leq C \text{ for i=1...d}$$
$$W_i \leq C \text{ for i=1...Y}$$

These max norm regularization constraints are enforced by calculating scalar factors $a = \frac{C}{\|V_i\|}$ and $b = \frac{C}{\|W_i\|}$ st $aV_i \leq C$ and $bW_i \leq C$ for all i.

5 Results

5.1 Tag Prediction

Precision@k was evaluated for each subject and k=1, 5, 10 using leave-one-run-out cross validation. Precision values were compared to three baselines: precision achieved using random mapping matrices W and V, and precision achieved using a K-Nearest Neighbors classifier with k=1,5. Tag prediction precision is significantly above random for all subjects and for k=1, 5, 10, however, the WARP algorithm is outperformed by the simple K-NN classifier (k=5) in all cases. These results are summarized in figure 5.

5.2 Brain Activity Prediction

Brain activity prediction was evaluated using a leave-two-out retrieval paradigm as described in [5]. In this evaluation paradigm, two stimuli presentations are left out of the training phase: a target and a decoy. The trained model is then used to predict brain activity for the left out stimuli. A hit is recorded if the predicted target brain is more similar to the true target brain than the predicted decoy brain. This evaluation paradigm is depicted in Figure 6.

	Subject	sj	yw	ad	at	am	ec	ab	jd	hy	mg	mh	sg	sw	kj	zi	Average
p@1	random	.132	.137	.147	.120	.128	.122	.128	.113	.153	.137	.133	.145	.107	.120	.132	**.130**
	KNN (k=1)	.192	.212	.258	.205	.278	.273	.267	.240	.260	.212	.278	.213	.270	.237	.188	**.239**
	KNN (k=5)	.430	.435	.470	.388	.385	.392	.435	.488	.425	.303	.455	.337	.390	.360	.390	**.406**
	WARP	.264	.262	.264	.270	.277	.257	.264	.274	.274	.287	.231	.226	.274	.245	.238	**.260**
p@5	random	.129	.130	.121	.127	.135	.136	.138	.124	.136	.129	.139	.130	.123	.126	.144	**.131**
	KNN (k=1)	.214	.244	.251	.212	.240	.260	.255	.240	.249	.201	.271	.181	.221	.213	.189	**.229**
	KNN (k=5)	.324	.397	.386	.326	.346	.343	.366	.404	.372	.287	.367	.290	.331	.319	.324	**.345**
	WARP	.285	.252	.267	.258	.305	.256	.264	.287	.299	.310	.265	.264	.280	.282	.272	**.276**
p@10	random	.127	.137	.135	.134	.136	.128	.134	.133	.132	.128	.134	.136	.128	.126	.139	**.132**
	KNN (k=1)	.202	.250	.237	.199	.221	.232	.230	.228	.247	.182	.255	.181	.207	.183	.195	**.217**
	KNN (k=5)	.320	.364	.359	.334	.338	.336	.342	.358	.355	.315	.348	.325	.336	.331	.317	**.338**
	WARP	.251	.242	.243	.233	.260	.241	.244	.255	.253	.265	.235	.258	.234	.239	.248	**.247**

Figure 5: Precision@k for Tag Prediction

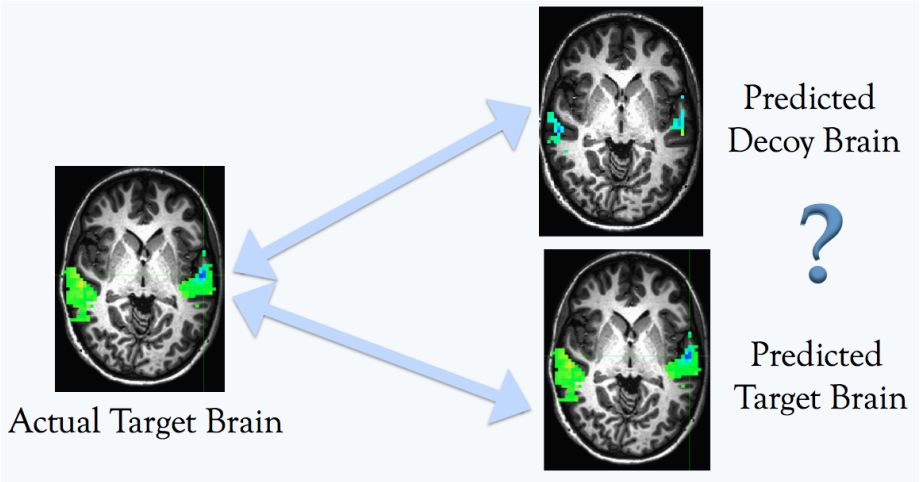

Figure 6: Evaluation paradigm.

This evaluation becomes very computationally expensive very quickly. For the purposes of this project I was forced to make several compromises to this evaluation paradigm in order to make it feasible given the time constraints. I reduced the number of datapoints by grouping samples within stimulus presentations. Since three brain volumes were collected for every 6-second except, 600 samples was reduced to 200 (600/3) samples. Instead of repeating the test for

every pair of stimuli, each sample is used as a target only once. Performance on this measure was no better than chance (.50). Results are summarized in Figure 7.

Subject	sj	yw	ad	at	am	ec	ab	jd	hy	mg	mh	sg	sw	kj	zi
Accuracy	.465	.435	.475	.435	.495	.445	.460	.510	.460	.470	.450	.430	.450	.520	.495
Average	**.4663** (below chance)														

Figure 7: Results of leave-two-out retrieval experiment to evaluate brain activity prediction. Performance is no better than chance (50%)

6 Discussion

Last.fm tags were successfully predicted from brain activity evoked by music using the method proposed in [6], however, I was unable to show that this algorithm could successfully predict brain activity from tags. Additionally, a k-NN classifier (k=5) out performed the WARP algorithm on the tag prediction task. There are several possible explanations for these results.

The method proposed by Weston et al. was designed especially for web-scale data and situations in which algorithms like k-NN are not feasible. My small dataset does not require the random sampling used by the WARP algorithm and algorithms like k-NN are entirely feasible. The main motivation for using this algorithm was not that it was the most appropriate for tag prediction, but rather that it provided a means of predicting brain activity from tags. So it is not so surprising that k-NN outperforms the WARP algorithm at tag prediction.

Although I was unable to show that this algorithm could be used to predict brain activity from tags, it's possible that a more complete evaluation would show otherwise. The compromises that were made to the leave-two-out retrieval experiment could have obscured a positive result. Additionally there are two hyperparameters to this algorithm (C, the constraint from the max norm regularization, and D, the dimensionality of the joint feature space), whose fine tuning might achieve better results.

Intuitively, it seems reasonable that it would be easier to predict tags from brain activity than to predict brain activity from tags since: a) there is more information in brain images than in tags, and b) it is possible that the true mapping from tags to brain-images are non-linear. Perhaps more descriptive tag data (e.g. ranked input tags) could be used to capture more of the subtlety of the brain's response. More informative tag information would also allow for the use of a kernel (e.g. gaussian kernel) which also might help to capture some of the non-linearities in the mapping from tags to brain response [7].

References

[1] Jason Weston, Samy Bengio, and Philippe Hamel. Large-scale music annotation and retrieval: Learning to rank in joint semantic spaces. *Journal of New Music Reserach*, 2011.

[2] 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.

[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] N. Usunier, D. Buffoni, and P. Gallinary. Ranking with ordered weighted pairwise classification. In *Proceedings of the International Conference on Machine Learning*, 2009.

[5] 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.

[6] 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.

[7] O. Bousquet J. Weston, B. Scholkopf. Joint kernel maps. In *Predicting Structured Data*. Springer Verlag, 2005.