Automatic Generation of Social Tags for Music Recommendation
(Extra Material)

Douglas Eck[*]
Sun Labs, Sun Microsystems
Burlington, Mass, USA
douglas.eck@umontreal.ca

Paul Lamere
Sun Labs, Sun Microsystems
Burlington, Mass, USA
paul.lamere@sun.com

Thierry Bertin-Mahieux
Sun Labs, Sun Microsystems
Burlington, Mass, USA
bertinmt@iro.umontreal.ca

Stephen Green
Sun Labs, Sun Microsystems
Burlington, Mass, USA
stephen.green@sun.com

This annex is meant provides some tables and graphics which did not fit in the main 8-page NIPS paper. This document won't make much sense unless it is read along with the paper itself.

Some statistics about tags mined from Last.FM

Tag Type	Frequency	Examples
Genre	68%	heavy metal, punk
Locale	12%	French, Seattle, NYC
Mood	5%	chill, party
Opinion	4%	love, favorite
Instrumentation	4%	piano, female vocal
Style	3%	political, humor
Misc	3%	Coldplay, composers
Personal	1%	seen live, I own it

Table 1: The distribution of the types of tags for the 500 most frequently applied tags at Last.fm.

---

[*]Eck and Bertin-Mahieux currently at Dept. of Computer Science, Univ. of Montreal, Montreal , Canada

1

60 selected tags from Last.fm (total 6590293 tag applications)					
80S	65128 (0.99%)	Alternative	366550 (5.56%)	Alternative Rock	144022 (2.19%)
Ambient	67834 (1.03%)	Black Metal	90384 (1.37%)	Blues	44364 (0.67%)
British	48206 (0.73%)	Britpop	46010 (0.70%)	Chillout	53912 (0.82%)
Classic Rock	171529 (2.60%)	Dance	73541 (1.12%)	Death Metal	95745 (1.45%)
Doom Metal	30929 (0.47%)	Electronic	247122 (3.75%)	Electronica	116782 (1.77%)
Emo	109985 (1.67%)	Experimental	70555 (1.07%)	Favorites	35806 (0.54%)
Female Vocalists	136860 (2.08%)	Folk	106730 (1.62%)	Funk	38395 (0.58%)
German	34245 (0.52%)	Gothic	31538 (0.48%)	Gothic Metal	34350 (0.52%)
Grunge	39610 (0.60%)	Hardcore	103291 (1.57%)	Hard Rock	94303 (1.43%)
Heavy Metal	79631 (1.21%)	Hip-Hop	160607 (2.44%)	Hiphop	92389 (1.40%)
Indie	393588 (5.97%)	Indie Pop	60155 (0.91%)	Indie Rock	166477 (2.53%)
Industrial	84430 (1.28%)	Instrumental	44823 (0.68%)	Japanese	64960 (0.99%)
Jazz	105152 (1.60%)	Melodic Death Metal	48753 (0.74%)	Metal	249553 (3.79%)
Metalcore	58249 (0.88%)	Newwave	39981 (0.61%)	Pop	211833 (3.21%)
Post-Punk	34352 (0.52%)	Post-Rock	50936 (0.77%)	Power Metal	51625 (0.78%)
Progressive Metal	51044 (0.77%)	Progressive Rock	79470 (1.21%)	Punk	182593 (2.77%)
Punk Rock	82083 (1.25%)	Rap	109179 (1.66%)	Reggae	40664 (0.62%)
Rock	659670 (10.01%)	Screamo	40665 (0.62%)	Seen Live	490504 (7.44%)
Singer-Songwriter	123163 (1.87%)	Ska	39000 (0.59%)	Soul	46695 (0.71%)
Soundtrack	45209 (0.69%)	Thrash Metal	55955 (0.85%)	Trip-Hop	49179 (0.75%)

Table 2: Statistics on the tags we data-mined from Last.fm. "# Artists" indicates the number of artists for which this tag was the most frequent.

2

Visualization

Here is a projection of the booster weights for the tags "rock" and "electronic". No adjustment of tag magnitudes was made. Only a few artists are shown.
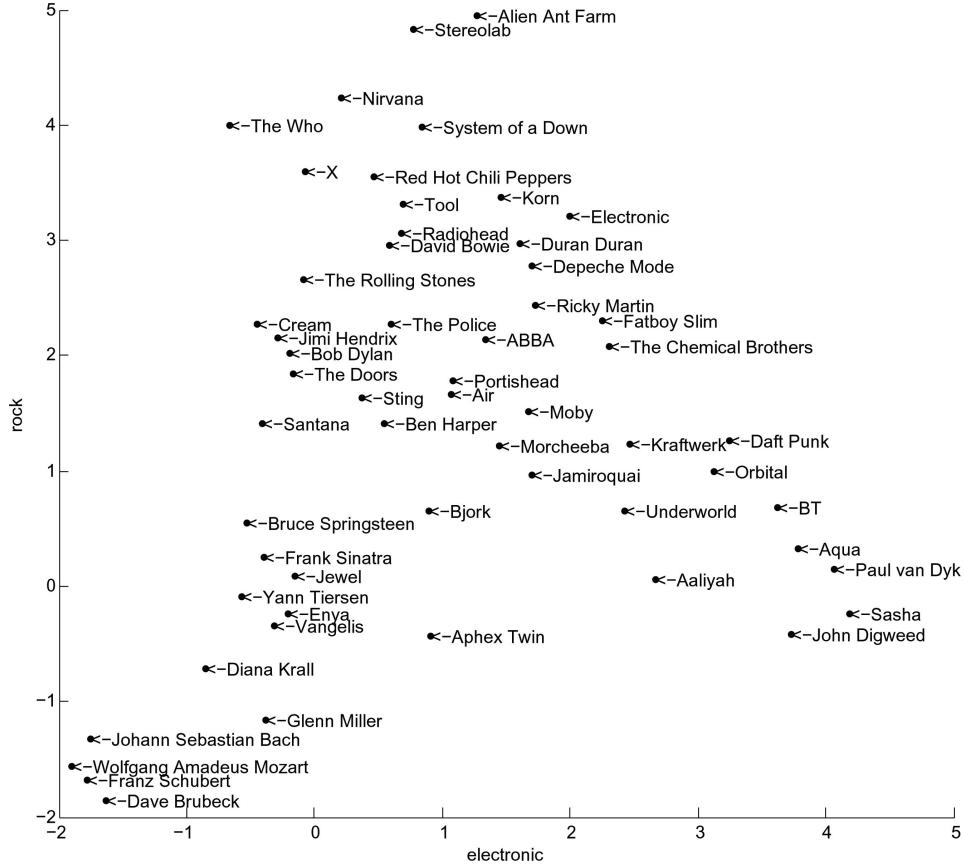
Figure 1: "rock" versus "electronic" for several artists.

3

Breakdown of learning error rates by tag

Summary of error rates for selected boosters					
Tags	Segments	Songs	Tags	Segments	Songs
80s	45.2	42.1	indie	49.6	45.4
alternative	45.8	41.5	indie pop	45.3	42.5
alternative rock	45.7	43.1	indie rock	48.3	45.1
ambient	42.3	38.5	industrial	39.6	36.8
black metal	21.3	17.8	instrumental	31.0	27.2
blues	46.2	42.5	japanese	39.2	37.3
british	42.1	39.5	jazz	43.0	40.0
britpop	47.3	45.4	melodic death metal	22.8	19.2
chillout	46.0	42.9	metal	43.4	39.2
classic rock	47.6	43.5	metal core	35.2	30.8
dance	41.2	38.1	new wave	47.1	44.1
death metal	29.5	26.1	pop	49.4	46.2
doom metal	43.31	39.57	post punk	43.2	40.1
electronic	48.7	45.1	post rock	34.4	32.3
electronica	44.5	35.15	power metal	25.0	20.7
emo	42.2	38.8	progressive metal	30.2	26.1
experimental	48.7	46.4	progressive rock	46.0	43.1
favorites	46.29	42.48	punk	47.5	45.2
female vocalists	44.1	40.1	punk rock	44.3	41.4
folk	46.0	42.6	rap	37.1	35.2
funk	46.6	43.3	reggae	30.9	28.2
german	24.2	20.9	rock	43.0	39.4
gothic	34.6	31.3	screamo	37.7	33.9
gothic metal	36.3	32.2	seen live	45.7	42.8
grunge	42.6	39.2	singer songwriter	48.2	45.0
hard core	38.6	35.8	ska	35.0	31.4
hard rock	46.4	44.0	soul	49.4	46.6
heavy metal	30.1	27.0	soundtrack	42.5	39.8
hip-hop	42.5	40.5	thrash metal	36.6	32.4
hiphop	35.6	31.9	trip-hop	43.7	41.8

Table 3: Test error (%) on predicting bins for songs and segments. Note that classical has only 2 bins. All boosters used 2000 single-stump weak learners.

4

Ground truth	Social	Auto
The Beatles		
Radiohead	The Zombies	Rolling Stones
Rolling Stones	The Lovin' Spoonful	Queen
Led zeppelin	Elvis Presley	Donovan
Pink Floyd	The Who	Billy Joel
David Bowie	Fleetwood Mac	Traffic
The Who	Rolling Stones	Los Lobos
Coldplay	George Harrison	The Doors
White Stripes	Queen	The Band
Beck	John Lennon & Yoko Ono	Stevie Ray Vaughn
Red Hot Chili Peppers	The Doors	The Lovin' Spoonful
Wilco		
Sufjan Stevens	Sparklehorse	Kelley Stoltz
Elliott Smith	Silver Jews	John Cale
The Flaming Lips	Calexico	John Linnell
The Shins	Grandaddy	Lloyd Cole
Modest Mouse	Clem Snide	Son Volt
Pavement	Brian Jonestown Massacre	Blonde Redhead
Beck	Modest Mouse	The Delgados
Radiohead	Sebadoh	Death Cab for Cutie
New Pornographers	Malajube	Gomez
The Beatles	Mercury Rev	Mark Knopfler
Bob Dylan		
The Band	Flying Burrito Brothers	Townes Van Zandt
Woody Guthrie	Don McLean	Crosby, Still & Nash
Grateful Dead	Crosby, Still, Nash & Young	Traffic
Rolling Stones	Buffalo Springfield	The Band
Johnny Cash	The Youngbloods	Emerson, Lake & Palmer
Tom Waits	Johnny Cash Massacre	The Lovin' Spoonful
The Beatles	The Band	Crosby, Still, Nash & Young
Crosby, Stills, Nash & Young	Warren Zevon	Johnny Cash
George Harrison	Crosby, Still & Nash	Tom Paxton
Townes Van Zandt	Jerry Garcia & David Grisman	Howlin' Wolf
Ella Fitzgerald		
Sarah Vaughan	Sarah Vaughan	Anita O'Day
Billie Holiday	Dinah Washington	Sarah Vaughan
Duke Ellington	Diana Krall	Betty Carter
Dinah Washington	Bilie Holiday	Doris Day
Count Basie	Betty Carter	Gerry Mulligan
Louis Armstrong	Stacey Kent	Charles Mingus
Nina Simone	Peggy Lee	Stan Getz
Chet Baker	Shirley Horn	Clifford Brown
Frank Sinatra	Madeleine Peyroux	Ben Webster
Diana Krall	Patricia Barber	Etta James
Aqua		
Britney Spears	Elvis Crespo	Ministry of Sound
Madonna	Lou Bega	La Bouche
Eiffel 65	Madonna	Jennifer Lopez
Avril Lavigne	Britney Spears	Supreme Beings of Leisure
Evanescence	Jennifer Lopez	Olive
Shakira	ABBA	Sven Vath
"Weird Al" Yankovic	Cher	Judge Jules
ABBA	Janet Jackson	M.I.A.
Black Eyed Peas	Gloria Estefan	Cher
Enya	Roger Whittaker	dZihan & Kamien

Table 4: Close artists for our 3 similarity matrices: ground truth (TF/IDF on Last.fm user data), from social tags, and from autotags

5