Music classification using lyrics

John Chia

December 15, 2009

Outline

Problem Description

- Music classification: label songs with genres
- Framed as automatic classification of text (lyrics)
- Supervised learning
- Applications in other systems: recommendation, summarization, information retrieval
- Related work: [Baumann and Hummel, 2003] (TF-IDF vector cosine), [Logan et al., 2004] (PLSA) and [Brochu and Freitas, 2003]

Lyrics Database

- Automatically pulled lyrics from AZlyrics
- Downloaded and parsed 2163 out of 8764 songs in uspop2002 (24.7%)
- Distribution of songs by genre:

  208 Others (6 lowest occuring genres combined)
  209 Rap
  270 R&B
  1478 Rock

Multinomial Naive Bayes (MNB)

- Find the posterior probability of a category c (i.e. a genre) given the document d

$$P(c|d) = P(c|w_{1:n}) = \frac{P(c)}{P(w_{1:n})} \prod_{i=1}^{n} P(w_i|c)$$

- Classify according to:

$$\underset{j}{\text{argmax}}\ P(c_j) \prod_{i=1}^{n} P(w_i|c_j)$$

- Also used: Transformed Weight-normalized Complement Naive Bayes (TWCNB) [Rennie et al., 2003]

Probabilistic Latent Semantic Analysis

- Views joint on words and documents as marginalized over a hidden latent variable (a topic z)[Hofmann, 1999]

$$P(wd) = \sum_z P(z)P(d|z)P(w|z)$$

- Expectation-maximization used to find $P(z)$, $P(d|z)$ and $P(w|z)$
- Relation to latent semantic analysis

$$U_{i,j} = P(d_j|z_i) \quad \Sigma_{i,i} = P(z_i) \quad V_{i,j} = P(w_i|z_j)$$
$$P(w_i d_j) = [U\Sigma V]_{i,j}$$

Probabilistic Latent Semantic Analysis

- Idea: truncate Σ to reduce topics
  $\longrightarrow$ reduce semantic associations to most likely topics
- Perform MNB classification using joint

$$P(c|w_{1:n}) \propto P(c) \sum_i P(w_{1:n} d_i|c)$$

Empirical N-gram cross-entropy

- Empirically constructed n-gram models

$$P(w_i|w_{i-n-1:i}) = \frac{c(w_{i-n-1:i}) + \alpha}{(1+\alpha)\sum_i c(w_{i-n-1:i})}$$

- One for each genre q_c, one for each document to classify (p)

$$\begin{aligned} H(p, q_c) &= -\sum_x p(x) \log q_c(x) \\ &= H(p) + D_{KL}(p||q) \end{aligned}$$

- Classify according to:

$$\operatorname*{argmin}_c \; -\sum_x p(x) \log q_c(x)$$

TWCNB

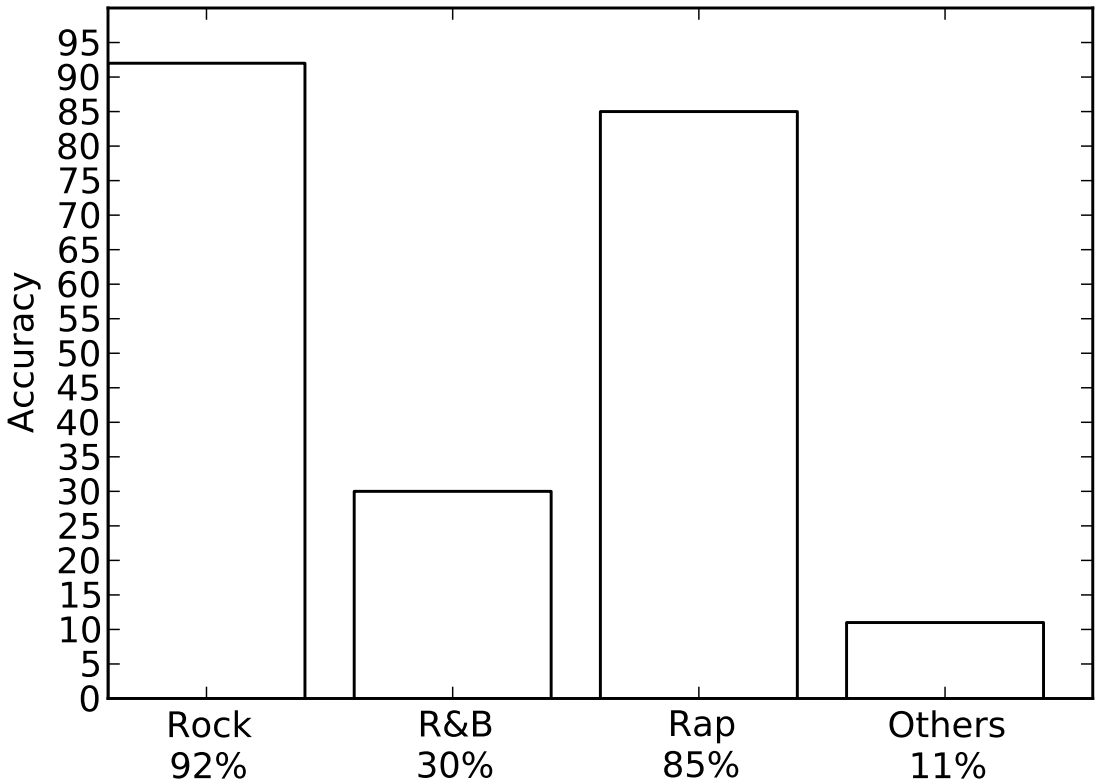

Figure: TWCNB classifier accuracy by genre.

Bigram Cross-entropy (BGCE)

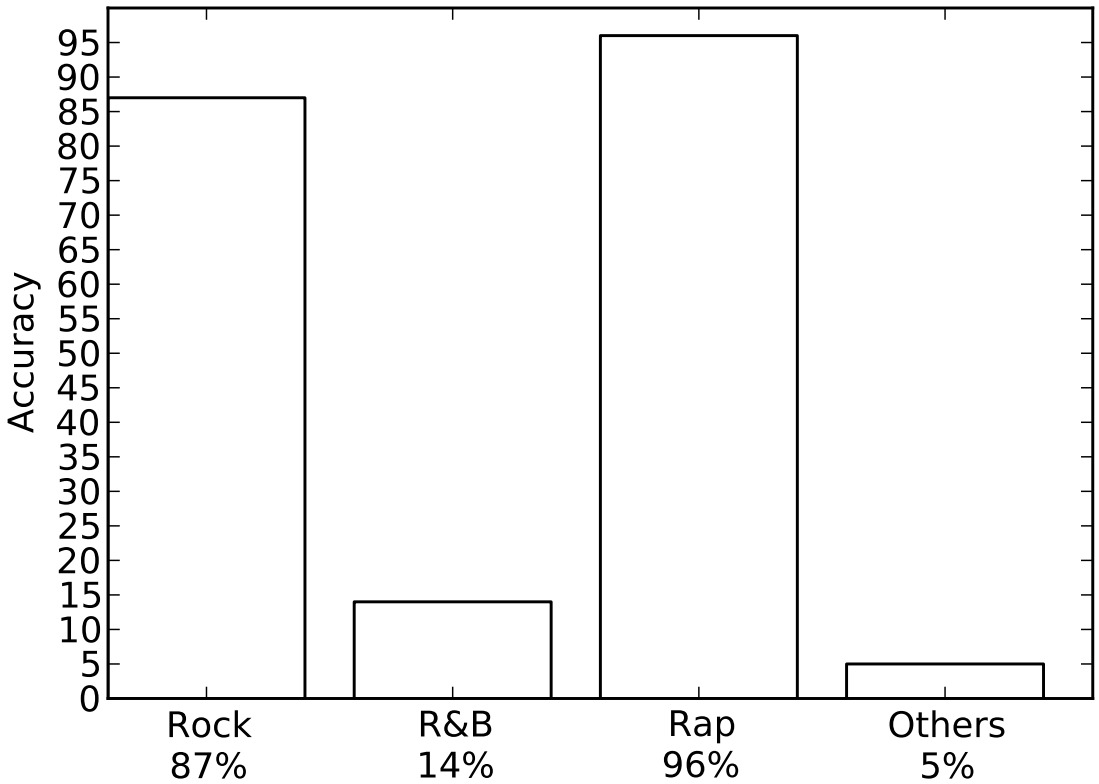

Figure: Bigram cross-entropy classifier accuracy by genre.

4-gram cross-entropy with backoff (4GCE)

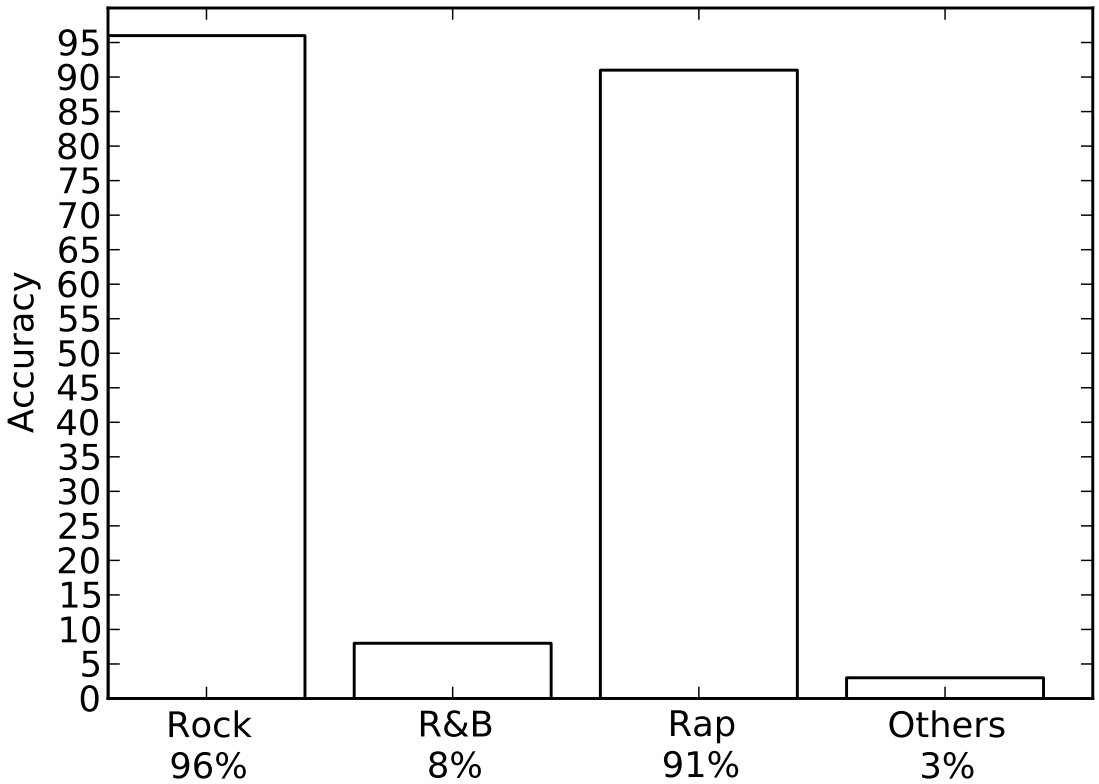

Figure: 4-gram with back-off cross-entropy classifier accuracy by genre.

Cumulative Classification Results

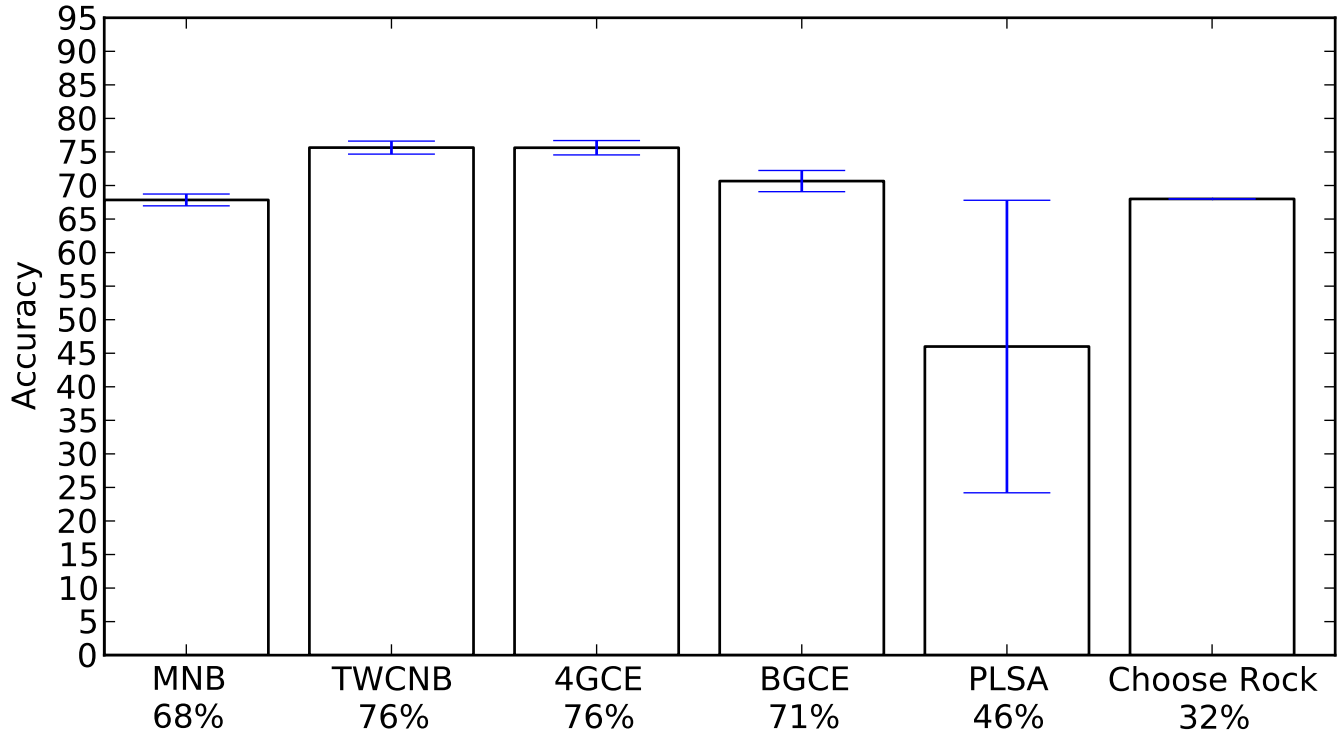

Figure: Accuracy of several classifiers.

PLSA problems

- Early convergence to relatively uniform distribution over topics (z)
- Slow to train:
  - hard to cross-validate parameters
  - used a significantly reduced number of features (100 vs 2000+)
- Doubts about application as presented to classification

Language Noise

- Noise from transcription errors, lack of transcriber consistency, and free-style nature of song writing
  - Slang: gimmie, gimme
  - Incorrect spelling: partna, pardner
  - False contractions: carseat
  - Lyrical emphasis: liaaaaar
- Phoneme transform (e.g. SOUNDEX) could be used to deal with this.

Most Relevant Words by Genre

Rock	R&B	Rap	Other
na	baby	n***as	world
oh	love	n***a	baby
love	na	s**t	love
gon	whoa	n***az	girl
know	girl	yo	oh
baby	ooh	yall	let
got	wan	ya	na
ll	doo	f**k	way
want	yeah	got	night
yeah	oh	em	heart

Table: Most relevant words per genre according to the tf-idf measure.

References


Baumann, S. and Hummel, O. (2003).
Using cultural metadata for artist recommendations.
In *Proceedings of 1st International Conference on Web Delivering of Music*.


Brochu, E. and Freitas, N. D. (2003).
"name that song¡‘: A probabilistic approach to querying on music and text.


Hofmann, T. (1999).
Probabilistic latent semantic indexing.
In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, New York, NY, USA. ACM.


Logan, B., Kositsky, A., and Moreno, P. (2004).