

Cost-Sensitive Stacking for Audio Tag Annotation and Retrieval

Hung-Yi Lo^{1,2}, Ju-Chiang Wang¹, Hsin-Min Wang¹
and Shou-De Lin²

¹Academia Sinica, ²National Taiwan University

ICASSP 2011

May 25, 2011

Social Tagging to Music

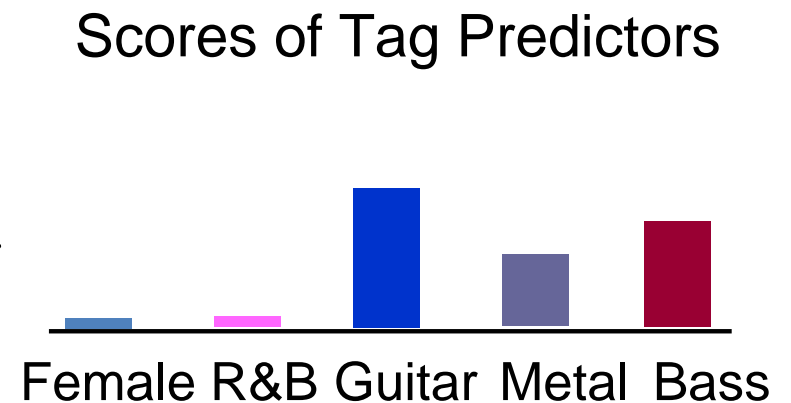
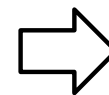
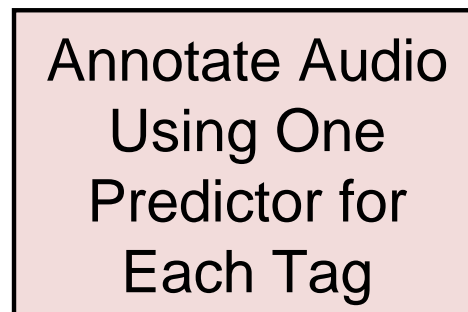
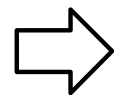
The screenshot shows the last.fm website interface. At the top, there is a red navigation bar with the last.fm logo and links for Music, Radio, Events, Charts, and Community. Below this is a secondary red bar with a promotional message: "New! Festival recommendations based on your taste »" and language options: "English | Help" and "Mus".

The main content area is divided into a left sidebar and a main right section. The sidebar contains a list of menu items: Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, and Charts. The main section displays the artist "The Beatles" and the track "Let It Be". Below the track name, there is a "Tags" section with a word cloud of user-generated tags. The most prominent tags are "60s", "beatles", "classic rock", "british", "oldies", "piano pop", and "rock". Other visible tags include "70s", "acoustic", "alternative", "alternative rock", "amazing", "awesome", "ballad", "ballads", "beautiful", "brilliant", "british invasion", "britpop", "calm", "chill", "chillout", "classics", "cool", "downtempo", "easy listening", "english", "favorite", "favorites", "favourite", "favourite songs", "favourites", "good", "great", "guitar", "indie", "john lennon", "love", "male vocalist", "male vocalists", "melancholic", "melancholy", "mellow", "moody", "night", "paul mccartney", "perfect", "pop rock", "psychedelic", "rock ballad", "rolling stones top 500 songs of all time", "sad", and "singer-songwriter".

Cost-Sensitive Stacking for exploiting tag count
and tag correlation jointly

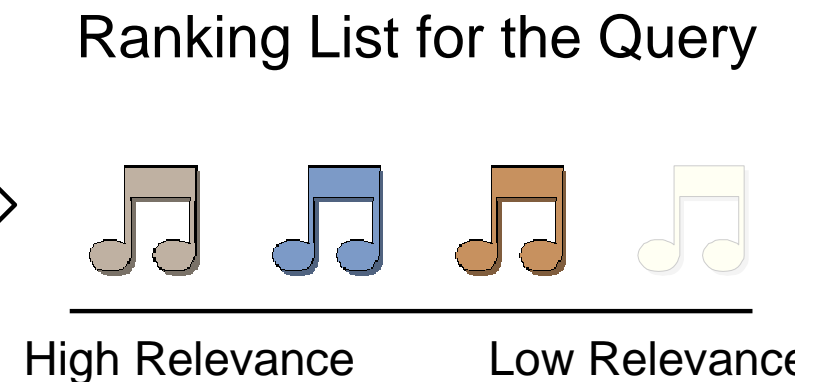
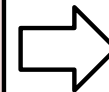
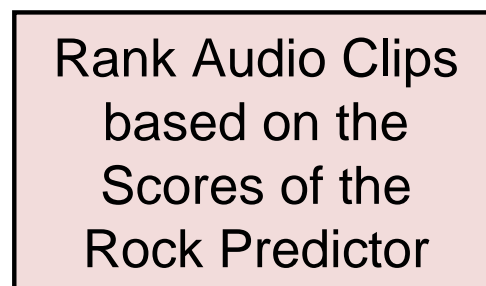
Audio Tag Annotation and Retrieval

Annotating audio clips with tags



Retrieving audio clips using a tag query

A Query: **Rock**



Tag Count Reflects Confidence

- Noisy Social Tags:
 - Social tags are assigned by people with **different levels of musical knowledge**, they inevitably contain noisy information
- **Tag count** information should be considered in automatic tagging because the count reflects the **confidence degree** of the tag
 - High count tags are more **reliable** and **salient**

	High Count Tag	Low Count Tag
Number of Clip-Tag Pairs	686	4,418
False Negative Rate	26.38	64.42

Much Smaller Than



An Example: "Let it Be" and its Tags

The screenshot shows the last.fm website interface. At the top, there is a red navigation bar with the last.fm logo and links for Music, Radio, Events, Charts, and Community. Below this is a secondary red bar with a promotional message and language options. The main content area features a sidebar on the left with navigation links for Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, Charts, Similar Artists, and Tags. The main content displays the artist 'The Beatles' and the track 'Let It Be'. Below the track name, a large collection of tags is shown in various sizes and colors, with 'classic rock' and 'the beatles' being the most prominent. A 'Tag' button is visible at the bottom left of the tag area.

last.fm Music Radio Events Charts Community

New! Festival recommendations based on your taste » English | Help Mus

Artist

Biography

Pictures

Videos

Albums

Tracks


Events

News

Charts

Similar Artists

Tags

 The Beatles » Tracks » Let It Be

Tags

60s 70s acoustic alternative alternative rock amazing awesome ballad ballads
beatles beautiful brilliant **british** british invasion britpop calm chill chillout
classic **classic rock** classics cool downtempo easy listening
english favorite favorites favourite favourite songs favourites good great guitar indie
john lennon love male vocalist male vocalists melancholic melancholy mellow moody
night **oldies** paul mccartney perfect **piano pop** pop rock psychedelic
rock rock ballad rolling stones top 500 songs of all time sad singer-songwriter
sweet **the beatles** uk uplifting 1970

Tag

An Example: “Let it Be” and its High Count Tags

The screenshot shows the last.fm website interface. At the top, there is a red navigation bar with the last.fm logo and links for Music, Radio, Events, Charts, and Community. Below this, a secondary bar contains a promotional message and language options. The main content area features a sidebar on the left with navigation links for Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, Charts, and Similar Artists. The main content displays the track 'Let It Be' by The Beatles, including a small album cover and a breadcrumb trail. Below the track information is a word cloud of tags. The most prominent tags are '60s', 'beatles', 'british', 'classic rock', 'oldies', 'pop', and 'rock'. The 'Tracks' link in the sidebar and the 'Let It Be' text in the breadcrumb trail are highlighted in red.

last.fm Music Radio Events Charts Community

New! Festival recommendations based on your taste » English | Help Mus

Artist

Biography

Pictures

Videos

Albums

Tracks

Events

News

Charts

Similar Artists

The Beatles » Tracks » Let It Be


Tags

60s
beatles
british
classic rock
oldies
pop
rock

More Reliable, Salient, and Important Tags

Considering Tag Count in Music Tagging

- In previous work, the tag count is transformed into 1 (with a tag) or 0 (without a tag), by using a **threshold**
 - A binary classifier is trained for each tag to make predictions
- Some problems:
 1. **The tag count information is lost**
 - A tag assigned twice is treated in the same way as a tag assigned hundreds of times
 2. **Hard to determine the threshold**
 3. **Ambiguity on the class membership of the instances nearby the threshold**
 - For example, if we set tag count threshold to 10:

- Tag count is 10 => Positive Instance
 - Tag count is 9 => Negative Instance
- 

Considering Tag Count in Music Tagging

- **Question:** how to use the tag count information for audio tag annotation and retrieval?
- **Answer:** Cost-Sensitive Learning with the Tag Counts as Costs

Cost-Sensitive Learning

- Given a training set (\mathbf{x}_i, y_i, c_i) , $i=1, \dots, n$, where \mathbf{x}_i is the feature vector, y_i is the class label, and c_i is the **misclassification cost** of the i -th training sample
- The goal is to learn a classifier h which **minimizes the expected cost** on unseen instances:

$$E[cI(h(\mathbf{x}) \neq y)]$$

where $I(\square)$ is an indicator function that returns 1 if its argument is true, and 0 otherwise

- A more general setup of traditional classification problem

Cost-Sensitive Learning with Tag Counts as Costs

- The goal is to **minimize misclassified tag counts** for audio tag annotation and retrieval
 - If 100 users annotate an audio clip with “rock”, but the classifier yields a **false negative**, then the **cost is 100**
 - Paying more attentions on the **reliable**, **salient**, and **important** tags

- We exploit two cost-sensitive binary classifiers:

- Cost-Sensitive Support Vector Machine

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_i c_i \xi_i \\ \text{s. t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad \square \end{aligned}$$

- Cost-Sensitive AdaBoost

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t c_i y_i h_t(x_i))}{Z_t}$$

Tag Correlation Information

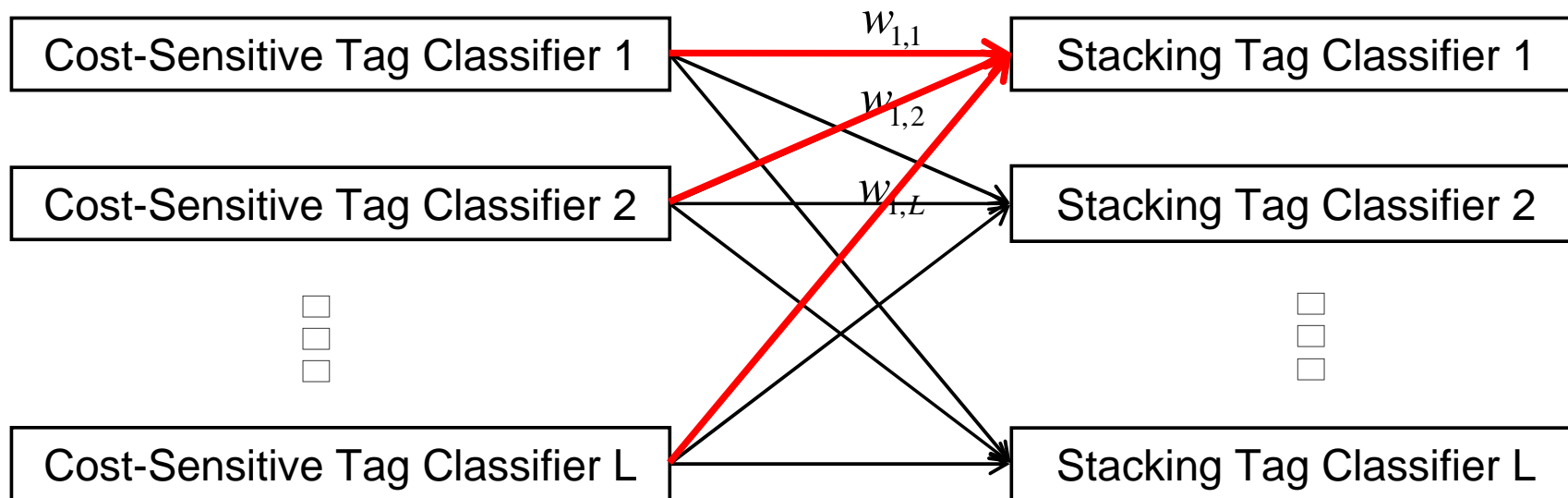
- In some previous works, the tag annotation task is separated into several binary classification problems
 - The tags are assumed **independent**
- **The label correlation information is lost**
 - For example, “Hip hop” and “Rap” often co-occur

	Hip hop	~Hip hop
Rap	160	17
~Rap	36	2259

- We propose **cost-sensitive stacking** to exploit tag count and correlation information jointly

Cost-Sensitive Stacking

- Stacking classifiers use the outputs of tag classifiers as inputs



- The stacking classifier for tag i is linear:

$$\sum_{j=1}^L w_{i,j} CS_j$$

$w_{i,j} > 0$ means tag j is **positively** correlated to tag i

$w_{i,j} < 0$ means tag j is **negatively** correlated to tag i

Experimental Setup

- The baseline is our winning method (cost-insensitive binary classifier) in MIREX 2009 audio tagging task
 - MIREX refers to Music Information Retrieval Evaluation eXchange
- Our experiments basically follow the MIREX 2009 setup
 - 45 tags were used and the audio clips associated with the tags were downloaded from MajorMiner, a [web-based music labeling game](http://majorminer.org/):
<http://majorminer.org/>
 - The resulting audio database contains 2,472 clips
 - Parameters selected based on inner cross-validation on training data
- Three-fold cross-validation one hundred times

metal	instrumental	horns	piano	guitar
ambient	saxophone	house	loud	bass
fast	keyboard	vocal	noise	british
solo	electronica	beat	80s	dance
jazz	drum machine	strings	pop	r&b
female	distortion	voice	rap	male
slow	electronic	quiet	techno	drum
funk	acoustic	rock	organ	soft
country	hip hop	synth	trumpet	punk

Results of Audio Annotation and Retrieval

Mean±St. D.		Clip AUC	Tag AUC	F-measure
AdaBoost	MIREX	87.73±0.09	79.41±0.25	30.27±0.46
	CS Only	88.54±0.07	80.56±0.20	32.20±0.41
	Stacking Only	88.50±0.11	79.91±0.31	31.18±0.45
	CS Stacking	88.82±0.09	80.69±0.28	32.42±0.45
SVM	MIREX	88.29±0.10	80.01±0.27	31.77±0.37
	CS Only	88.96±0.06	81.12±0.20	32.93±0.38
	Stacking Only	89.00±0.08	81.41±0.19	32.70±0.36
	CS Stacking	89.64±0.07	82.06±0.23	34.22±0.41
Ensemble	MIREX	88.47±0.07	81.89±0.19	33.35±0.40
	CS Only	89.21±0.06	82.54±0.18	34.32±0.41
	Stacking Only	89.12±0.07	82.37±0.18	33.59±0.37
	CS Stacking	89.57±0.06	82.85±0.17	34.69±0.46

Better Than

Conclusion

- Tag count and tag correlation are two important information for social tag prediction on multimedia data
- We have first formulated the audio tag prediction task as a cost-sensitive classification problem to minimize the misclassified tag counts
- We have then formulated the task as a cost-sensitive multi-label classification problem and proposed cost-sensitive stacking to exploit tag count and correlation information jointly
- The experimental results show that the new approach outperforms our MIREX 2009 winning method
- Hung-Yi Lo, Ju-Chiang Wang, Hsin-Min Wang, and Shou-De Lin, Cost-Sensitive Multi-Label Learning for Audio Tag Annotation and Retrieval, IEEE Trans. on Multimedia, June 2011.

Thank You