

Audio Tag Annotation and Retrieval by Using Tag Count Information

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

¹Academia Sinica, ²National Taiwan University

MMM 2011

Jan. 06, 2011

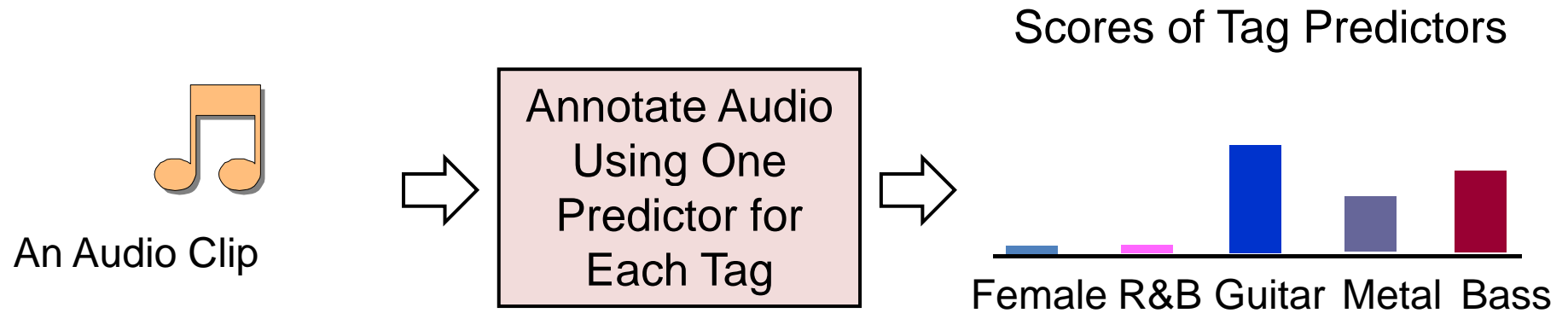
Social Tagging to Music

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, a secondary bar contains a promotional message and language options. The main content area is divided into a left sidebar with navigation links (Artist, Biography, Pictures, Videos, Albums, Tracks, Events, News, Charts, Similar Artists, Tags) and a main content area. The main content area displays the artist 'The Beatles' and the track 'Let It Be'. Below the track name, there is a 'Tags' section with a grid of blue text tags. The size of each tag indicates its popularity, with 'classic rock' and 'the beatles' being the largest. Other visible tags include '60s', 'beatles', 'british', 'classic', 'rock', 'piano pop', and 'oldies'.

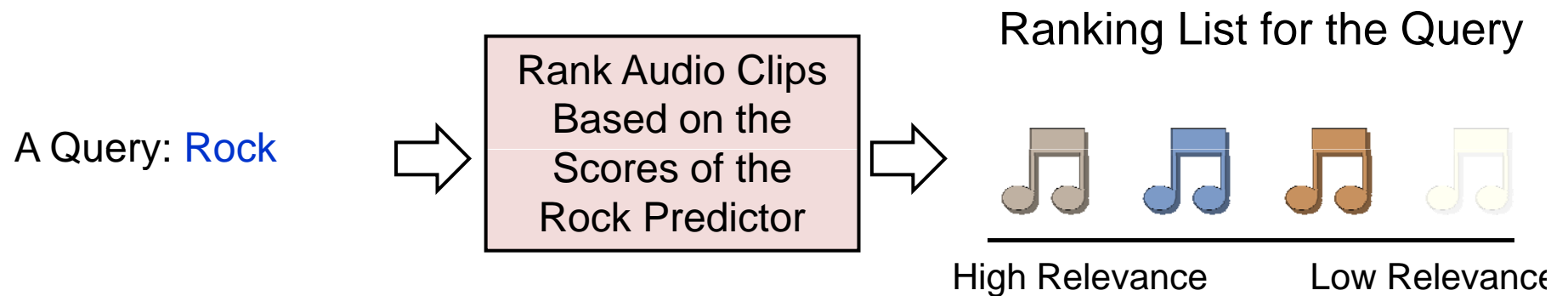
Tag count : the number of different users who have annotated the tag (a larger font indicates a higher tag count)

Audio Tag Annotation and Retrieval

Annotating audio clips with tags



Retrieving audio clips using a tag query



Motivation (1/2)

- Noisy Social Tags:
 - Social tags are assigned by people with **different levels of musical knowledge**, they Inevitably contain noisy information
 - Social tags are unstructured, free-form text that may contain **misspellings**
 - Tags may even assigned by **malicious** users
- **Tag count** information should be considered in automatic tagging because the count reflects the **confidence degree** of the tag
 - Tags with high counts are more reliable and credible

Motivation (2/2)

- In previous works, the tag count is transformed into 1 (with a tag) or 0 (without a tag), by using a threshold
 - Then train a binary classifier for each tag and make predictions about untagged audio clips
- **The tag count information are lost**
 - a tag assigned twice is treated in the same way as a tag assigned hundreds of times
- **Question:** how to use the tag count information for audio tag annotation and retrieval?
- **Answer:** Cost-sensitive Learning with the Tag Counts as Costs

Factors Affect the Tag Counts

1. Consistent Agreement

- When a large number of users consider that an audio clip should be associated with a particular tag, the label information is **more reliable**
- Only a small portion of an audio clip is related to a certain tag then the tag count will be small
- Tags with higher counts are **major property** of the audio

2. Tag Bias

- Some types are more often used. (such as “rock”)
- Some others are unpopular or hard to recognize. (“Baroque” is less popular than “classic”, “drum machine” might easily be recognized as “drum”)

3. Song/Album/Artist Popularity:

- Popular songs, albums, and artists usually receive more tags, since people tend to tag music that they like or they are familiar with

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 bar with a promotional message and language options. The main content area is divided into a left sidebar with navigation links (Artist, Biography, Pictures, Videos, Albums, Tracks, Events, News, Charts, Similar Artists, Tags) and a main content area. The main content area displays the artist name 'The Beatles', the track name 'Let It Be', and a large collection of tags. The tags are arranged in a word cloud format, with 'classic rock' being the most prominent. Other notable tags include '60s', 'beatles', 'british', 'pop', 'rock', and 'the beatles'. 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 Tags with Higher Counts

The screenshot shows the last.fm website interface. At the top, the navigation bar includes 'last.fm', 'Music', 'Radio', 'Events', 'Charts', and 'Community'. Below this, a banner reads 'New! Festival recommendations based on your taste »' with 'English | Help' and 'Mus' options. The main content area is for the track 'Let It Be' by 'The Beatles'. A sidebar on the left lists navigation options: Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, Charts, and Similar Artists. The track page features a small album cover and the text 'The Beatles » Tracks » Let It Be'. Below this, the word 'Tags' is followed by a word cloud. The most prominent tags in the word cloud are '60s', 'beatles', 'classic rock', 'british', 'oldies', 'pop', and 'rock'. A red banner at the bottom of the screenshot contains the text 'More Reliable, Major Properties, and Important Tags'.

Cost-sensitive Learning

- Given some training example (\mathbf{x}, y, c) , where c is the **misclassification cost** of this example
- Learn a classifier h which minimizes the expected cost on unseen instances:

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

where $I(\cdot)$ is an indicator function that yields 1 if its argument is true

- A more general setup of traditional classification problem

Cost-sensitive Learning Applications

- Business Marketing
 - Given some features of potential customers, each customer has a purchasing amount as misclassification cost
 - Decide which customers to mail a new catalog
 - Cost-sensitive learning with purchasing amount as cost
 - Goal: total profit obtained from some unseen testing customers
- Audio Tag Annotation with Tag Counts as Costs
 - Given some acoustic features of an audio with its tags and tag counts
 - Goal: minimize misclassified tag counts
 - If 100 users annotate an audio with “rock”, but the classifier causes a false negative, then it has to pay a cost equal to 100
 - Paid more attention on the reliable, major, and Important tags

Cost-insensitive Support Vector Machine

- Optimization problem:

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

- Train a binary classifier for each tag and gather all instances annotated with this tag as positive example

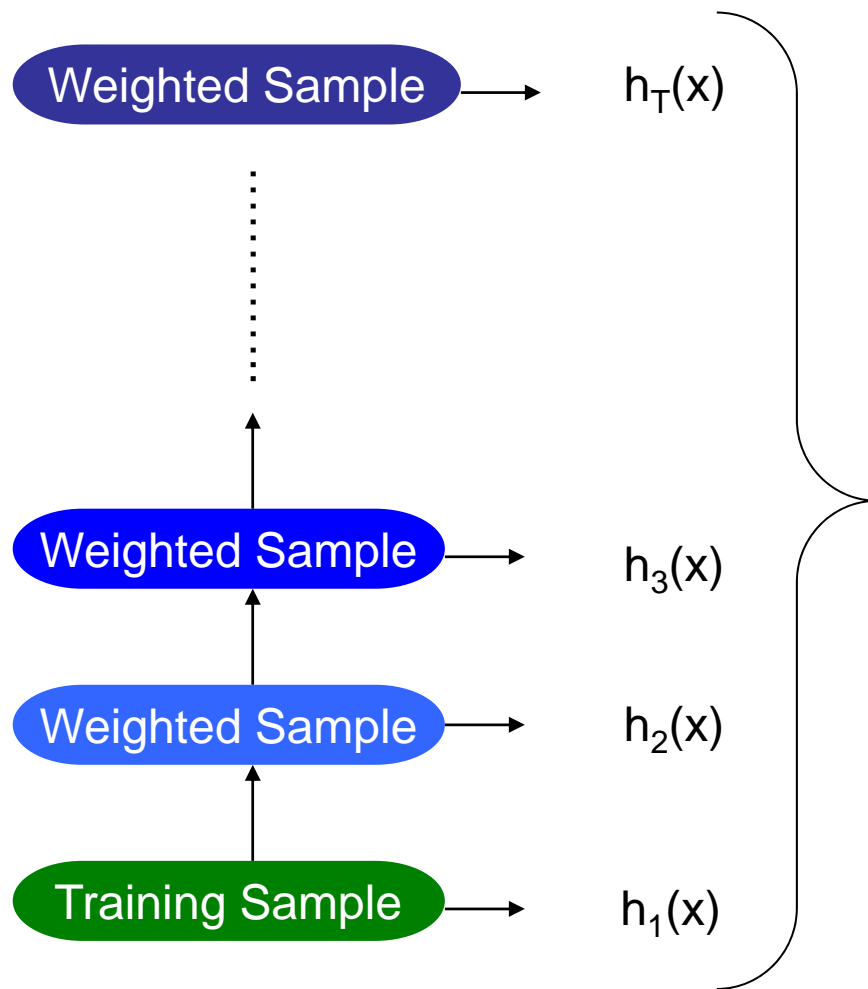
Cost-sensitive Support Vector Machine

- Optimization problem:

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

- The cost c_i is assigned as the tag count for positive examples
- The cost c_i is assigned uniformly for negative examples
 - Since people do not use negative tags like “non-rock” and “no drum”

Cost-sensitive AdaBoost



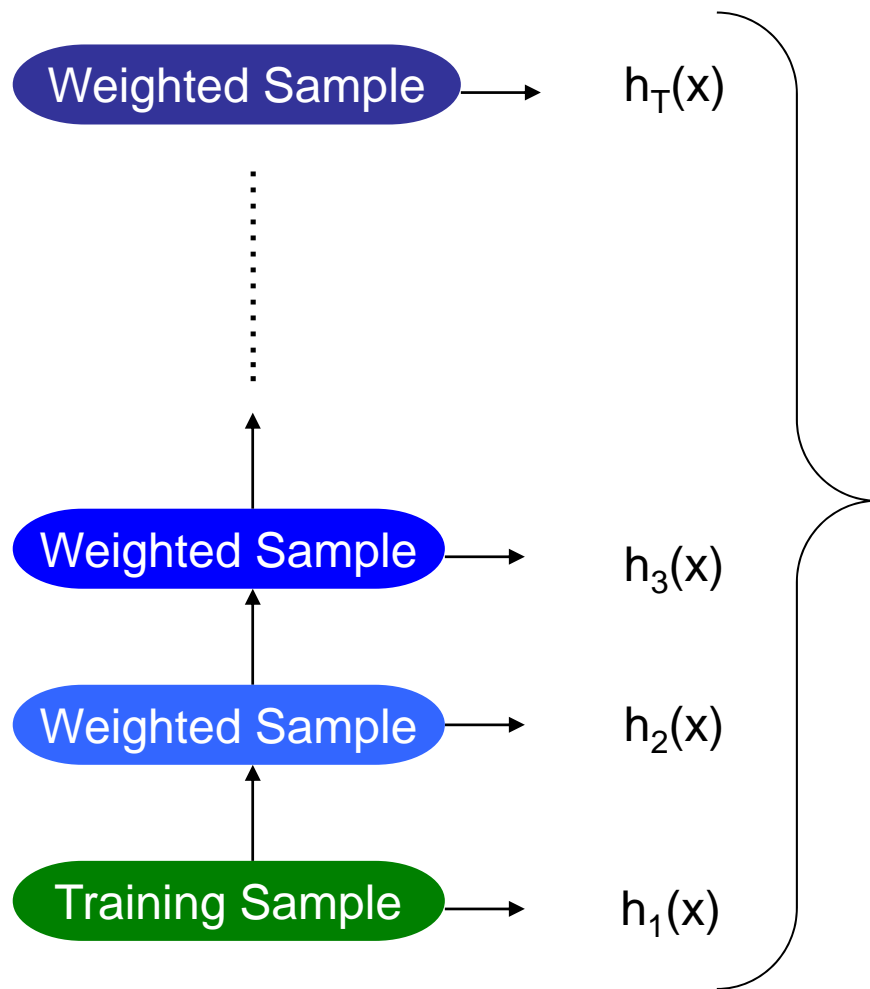
Minimize Weighted Error

$$\sum_i D(i) \exp(-y_i h_t(x_i))$$

Cost-sensitive Weight Update Rule:

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

Cost-sensitive AdaBoost



Final Classifier:

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x)$$

Cost-sensitive Evaluation Metrics: Which One is Better?

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 content area. The sidebar contains a list of menu items: Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, Charts, Similar Artists, and Tags. The main content area features a header for "The Beatles » Tracks » Let It Be" with a small album cover icon. Below the header is a "Tags" section containing a word cloud. The word cloud consists of various music-related terms in different sizes and colors (primarily blue and yellow), including "60s", "beatles", "british", "classic rock", "oldies", "pop", "rock", and "the beatles". At the bottom left of the word cloud area is a "Tag" button with a tag icon.

Cost-sensitive Evaluation Metrics: Which One is Better?

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: "New! Festival recommendations based on your taste »" and language options: "English | Help" and "Mus".

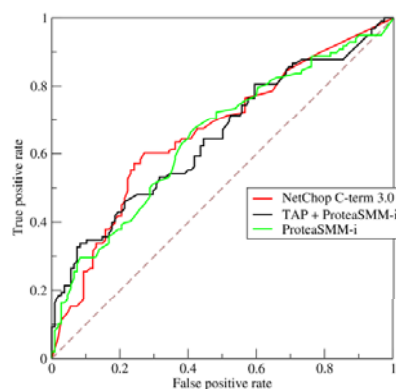
The main content area is divided into a left sidebar and a main right section. The sidebar contains a list of menu items: Artist, Biography, Pictures, Videos, Albums, Tracks (highlighted in red), Events, News, Charts, Similar Artists, and Tags.

The main section displays the artist's profile for "The Beatles » Tracks » Let It Be". It features a small album cover image and the heading "Tags". Below the heading, a large collection of blue, clickable tags is displayed in a grid-like fashion. The tags include: 70s, acoustic, alternative, alternative rock, amazing, awesome, ballad, ballads, beautiful, brilliant, british invasion, britpop, calm, chill, chillout, classic, 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, piano, pop rock, psychedelic, rock ballad, rolling stones top 500 songs of all time, sad, singer-songwriter, sweet, uk, uplifting, and 1970.

At the bottom of the tag list, there is a dark button with a tag icon and the text "Tag".

New Evaluation Metrics: Cost-sensitive Area Under ROC

- **ROC Curve**: is a graphical plot of true positives rate vs. false positives rate as the discrimination threshold is varied



- **Cost-sensitive Area Under ROC Curve (AUC)**: replace the true positive rate by cost-weighted true positive rate
- **Clip AUC**: given a audio clip, give correct tag higher scores
 - for audio annotation
- **Tag AUC**: given a tag, give correct tag higher scores
 - for audio retrieval

New Evaluation Metrics: Cost-sensitive F-measure

- Cost-sensitive precision (CSP):

$$\frac{\text{Weighted TP}}{\text{Weighted TP} + \text{Weighted FP}}$$

- Cost-sensitive recall (CSR):

$$\frac{\text{Weighted TP}}{\text{Weighted TP} + \text{Weighted FN}}$$

- Cost-sensitive F-measure

$$2 \times \frac{\text{CSP} \times \text{CSR}}{\text{CSP} + \text{CSR}}$$

- We evaluate on both cost-sensitive metrics and cost-insensitive metrics

Experimental Setup

- Compare to our winning method (cost-insensitive) in MIREX 2009 audio tagging competition
 - MIREX refers to Music Information Retrieval Evaluation eXchange
- Experiments basically follow the MIREX 2009 setup
 - Download audio data from MajorMiner, a [web-based music labeling game](http://majorminer.org/): <http://majorminer.org/>
 - Use the same 45 tags and download all the audio clips that are associated with these tags
 - The resulting audio database contains 2,472 clips
 - Select parameters using inner cross-validation on training data
- Repeat cross-validation twenty times to reduce variance

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 Cost-sensitive Metrics

Mean±St. D.	Cost-sensitive Tag AUC	Cost-sensitive Clip AUC	Cost-sensitive F-measure
AdaBoost	0.8055±0.0027	0.8892±0.0011	0.4099±0.0052
CS AdaBoost	0.8169±0.0023	0.8967±0.0005	0.4469±0.0081
SVM	0.8112±0.0022	0.8957±0.0007	0.4354±0.0077
CS SVM	0.8215±0.0023	0.9005±0.0004	0.4593±0.0056
Ensemble	0.8334±0.0019	0.8979±0.0007	0.4606±0.0067
CS Ensemble	0.8356±0.0018	0.9032±0.0006	0.4808±0.0072

Better Than



Results of Cost-insensitive (Regular) Metrics

Mean±St. D.	Tag AUC	Clip AUC	F-measure
AdaBoost	0.7941±0.0027	0.8773±0.0011	0.3018±0.0035
CS AdaBoost	0.8050±0.0023	0.8854±0.0005	0.3216±0.0049
SVM	0.7992±0.0021	0.8837±0.0007	0.3226±0.0053

Better Than

Tags with smaller counts may contain **noisy labeling information**

CSL method can ignore the noisy information by giving a smaller penalty (cost), and thereby train a more accurate classifier

CS Ensemble	0.8247±0.0017	0.8921±0.0005	0.3442±0.0046
-------------	---------------	---------------	---------------

Conclusion

- This paper has presented our novel idea for exploiting **tag count information** in audio tagging tasks
 - discussed several factors that affect the tag counts
 - consistent agreement is the most important issue
- Formulate the audio tag prediction task as a cost-sensitive classification problem to minimize the misclassified tag counts
- Present cost-sensitive versions of several regular evaluation metrics
- The proposed cost-sensitive methods outperform their cost-insensitive counterparts in terms of not only the cost-sensitive evaluation metrics but also the regular evaluation metrics
 - Since the tag count tell us whether we should trust this tag or not



Thank You