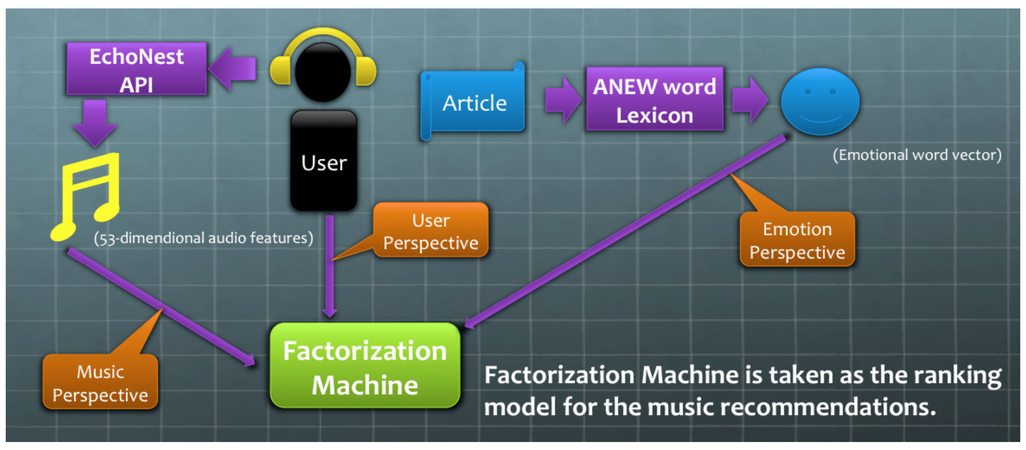
Using Emotional Context from Article for Contextual Music Recommendation

Chih-Ming Chen, Ming-Feng Tsai,
Department of Computer Science &
Program in Digital Content and Technology
National Chengchi University
Taipei 11605, Taiwan
{g10018, mftsai}@cs.nccu.edu.tw

Jen-Yu Liu, Yi-Hsuan Yang
Research Center for Information
Technology Innovation
Academia Sinica
Taipei 11564, Taiwan
{ciaua, yang}@citi.sinica.edu.tw

Overall Construction

Music usually carries people's emotions, and people sometimes express their feelings by writing articles while listening to music. In light of this observation, we proposes a context-aware approach that recommends music to a user based on the user's emotional state predicted from the article the user writes.



Dataset - Livejournal

LiveJournal is a well-known blog website where users can write blogs, or online diaries. The users are able to listen to a song and label a mood tag that reflects his or her emotional state while writing an article, as exemplified in following Figure. The collected data contains 19,596 users, 225,652 articles, and 30,260 songs.

one step, two step, fall behind
Jul. 13th, 2013 at 8:55 PM
The cookies smell great, though, and are super easy to make, so consider that part a thumbs up. And I am just using the vanilla ice cream I always make - it's the recipe that came with the ice cream maker.

Current Mood: accomplished
Current Music: Grandmother Song - Vienna Ten - LiveJournal Post Example

Our Ranking Approach

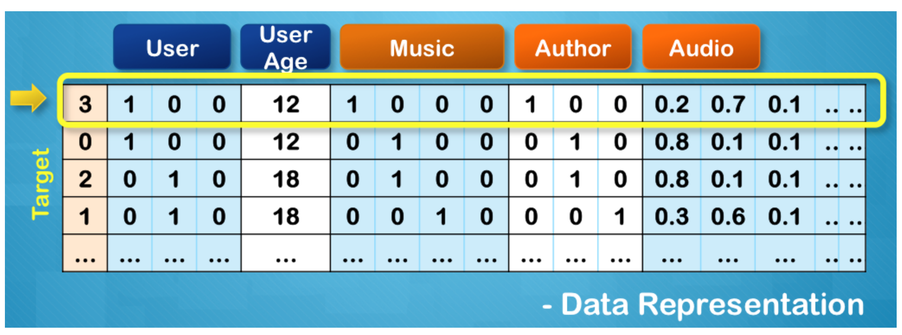
What we want for this problem

- A competitive model for ranking problem.
- Easy to embed various kinds of feature in the data.
- Capable of learning the relationship between user's emotional states and their listening behavior.

Factorization machine (FM) provides a good framework to tackle with the task:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \hat{w}_{ij} x_i x_j \,, \qquad \text{where } \hat{w}_{ij} = \sum_{f=1}^{\kappa} v_{if} v_{jf}. \tag{1}$$

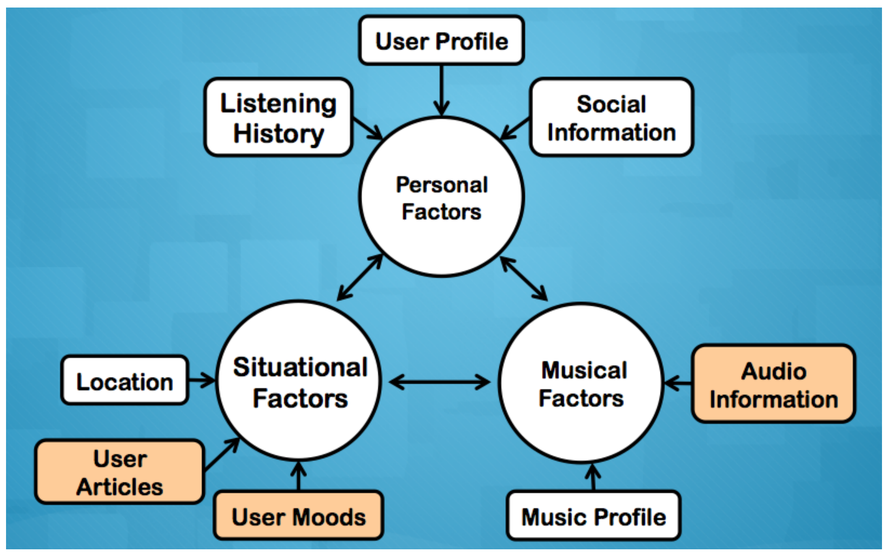
w_0 learns the global bias, w_i learns each weight of features x_i, and $\hat{w}_{ij}$ models the interaction of each pair of features. Instead of using single parameter for each interaction, FM factorizes it as the dot product of two vectors, where κ is the model complexity. This way allows high-quality parameters estimated by higher-order interactions under sparsity.



Factorization Machines
For more details and related works about Factorization Machine, please refer to (http://www.libfm.org/)

Feature Extraction

All the extracted features can be divided into following three aspects: Personal Factors, Situational Factors and Musical Factors.



Music Perspective Feature

Considering the abundant information within music, we use 53 audio features to represent various perceptual dimensions of music, including danceability, loudness, mode, and tempo. They are extracted by using the EchoNest API (http://developer.echonest.com/), a commonly used audio feature extraction tool developed in the field of music information retrieval.

Emotion Perspective Feature

The text of articles is converted to an emotional word vector by using the lexicon of Affective Norms for English Words (ANEW), which provides a set of normative emotional ratings for English words. The emotional words are rated by *valence*, *activation* and *dominance*. Each emotional word is weighted by Term-Frequency Inverse-Document-Frequency (TFIDF) measure,

$$tf(t,d) = \frac{f(w,d)}{\max\{f(w,d) : w \in d\}}, \qquad idf(t,d) = \log \frac{|D|}{|\{d \in D : t \in d\}|}, \tag{2}$$

where D is the total number of articles. Each term in vector is scored by $tf(t,d) \times idf(t,d)$. After the TFIDF weighting, we can get the Valence, Arousal, and Dominance (VAD) values of an article by a weighted summation of the VAD values of the emotional words occur in the article.

ANEW Lexicon

Description	Valence	Arousal	Dominance
dream	6.73	4.53	5.53
good	7.47	5.43	6.41
hate	2.12	6.95	5.05
...	...	...	...

Experimental Results

- Both the performance of the music perspective and the user perspective outperforms the CF-based one (i.e., U + S) by a great margin.
- This result implies that the VAD feature provides more emotional information of the user context, which might not be easily captured by mood tags or words only.
- Combining the content-based audio information with the contextual emotion information further improves the quality of recommendations.

- U: User
- S: Song
- M: Mood
- TFIDF: TF-IDF of Article
- Au: Audio Feature
- VAD: VAD of article

Music Perspective

Emotion (User) Perspective

Features	MAP@10	Recall
U + S	0.3817	0.5216
U + S + Au	0.4708	0.6185
U + S + M	0.4159	0.5628
U + S + TFIDF	0.4212	0.5643
U + S + VAD	0.4483	0.5905
U + S + M + VAD	0.4113	0.5547
U + S + Au+ M + VAD	0.5026	0.6540

Hybrid Method