Exploiting Hierarchical Tags for Context-awareness

Alan Said[1], Jérôme Kunegis[2] and Ernesto W. De Luca[1]

DAI-Labor
Distributed Artificial Intelligence Laboratory

[1]DAI Lab @ Technische Universität Berlin, [2]University of Koblenz-Landau

Third Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR 2010)

Abstract

Context-awareness for information retrieval and recommendation is a challenging problem as information about the user's current situation is rarely available. If such information would be available, retrieval and recommendation systems could present their users with search results or recommendations relevant to their current contexts. We present a model for context-aware recommendation based on a dataset where contextual information is implicitly available in the form of tags assigned to movies.

Tags for context-awareness

In the movie recommendation website Moviepilot, each movie can be tagged with any number of tags from five different keyword categories - *Mood*, *Intended Audience*, *Plot*, *Place* and *Time*. The categories are organized in hierarchical trees (Italy being the parent node of Sicily in the place category, etc.).

Figure: Tags for the movie The Godfather shown in the yellow box.

Assuming the tags of movies rated by users reflect their current context, the tags can be used to enhance collaborative filtering by including the user's context.

Example:
A user regularly rates movies tagged with the *Guys* keyword from the Intended Audience category combined with *Exciting* or *Thrilling* from the Mood category. If the same user has reoccurring sessions of movies labeled *Family*, an assumption about the user's context can be made, i.e. the company of the user is different in the two sessions. By studying the user's history, and identifying similar events, the social context can be inferred (taking into consideration other contextual features such as time of day, day of week, etc.) and be used in the recommendation process.

Dataset

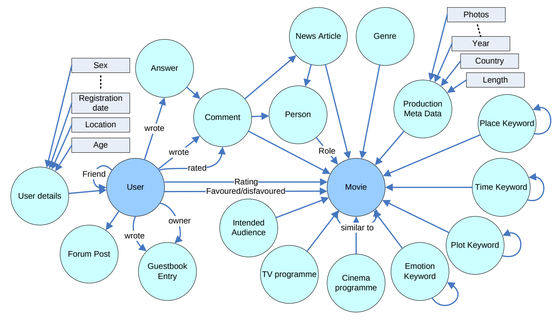
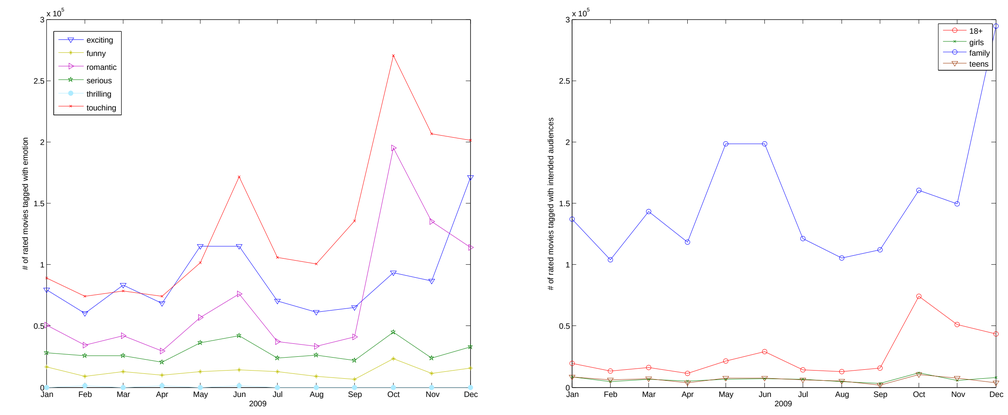


Figure: The entity relationship diagram of the dataset used.

Category	# of ratings	% of total
Mood	4,525,169	72.9
Intended Audience	2,343,939	37.8
Plot	5,887,824	94.8
Place	5,165,311	83.2
Time	4,595,547	74.0
Total_{tagged}	5,910,444	95.2
Total_{all}	6,208,990	100.0

Table: The number of ratings and percentages in each of the keyword categories



(a) Emotion (b) Intended audience

Figure: The number of ratings per month for popular tags in two tag categories. Note how the *family* tag receives considerably more ratings during December than during other months. This can be seen as the context of "Christmas" during which family movies should be considered more relevant.

Two proposed approaches

Hybrid kNN

By identifying a set of contexts, and the ratings performed in each of them, each user can be split into several virtual users, one for each context the user is know to have been in. The similarities to each user's neighbors is then calculated for each virtual user, meaning for every context of every user, creating context-aware neighborhoods.

	u_i	u_j	u_k	u_l	u_m
m_a	1	3		5	
m_b		4			4
m_c			5	2	
m_d	5	3		3	
m_e	3	4	1		1

Figure: User-Movie matrix without context-awareness.

	u_i^x	u_i^y	u_j^x	u_j^y	u_j^z	u_k	u_l	u_m
m_a	1				3		5	
m_b				4				4
m_c						5	2	
m_d	5		3				3	
m_e		3		4		1		1

Figure: User-Movie matrix with context-awareness added for users u_i and u_j.

Latent context modeling

Combining usage data with tags, creates a folksonomy-like structure in which every user-movie-tag relationship additionally has a rating. By creating a tensor representation of this data, where tags are combined with the ratings, we can use a tensor factorization approach for latent context modeling, i.e. HOSVD[1] or HyPLSA[2].

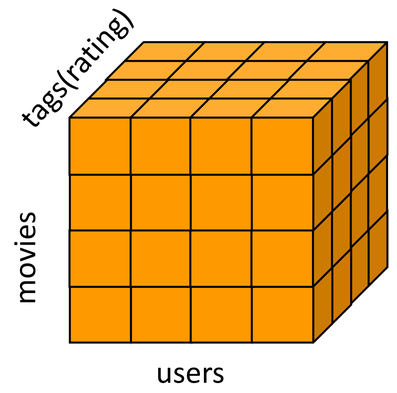


Figure: A tensor representation of the data.

[1] Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering - *A. Karatzoglu et al. 2010*
[2] A hybrid approach to item recommendation in folksonomies *R. Wetzker et al. 2009*

Tag assignments

Mood	TAS	Audience	TAS	Plot	TAS	Place	TAS	Time	TAS
Exciting	1252	Family	1405	Murder	849	USA	468	21st century	458
Funny	1095	Guys	759	Love	643	New York	366	present	412
Touching	1013	Girls	307	Friendship	495	Prison	306	1990's	304
Thrilling	876	18+	263	In Love	476	Paris	254	1980's	199
Serious	631	Teens	232	Father-son relationship	432	London	254	WWII	184

Table: The most popular tags in every category and the number of tag assignments each.