Intent-Oriented Diversity in Recommender Systems

Saúl Vargas, Pablo Castells and David Vallet
Universidad Autónoma de Madrid
{saul.vargas,pablo.castells,david.vallet}@uam.es
IR Group @ UAM

Motivation

Diversity in Information Retrieval

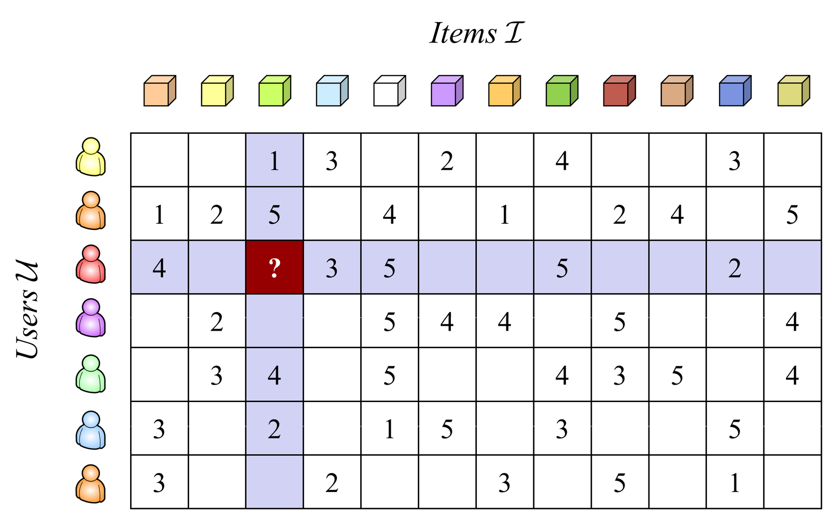
- Addressed as an issue of uncertainty in queries: ambiguity, underspecification (Zhai 2003, Chen 2006, Agrawal 2008, Clarke 2009, Santos 2010)
- Maximize the probability of returning at least some relevant doc
- Revision of document relevance independence assumption
- Formulated/solved in terms of query aspects / subtopics / subqueries / nuggets / intents / categories…
- Diversification algorithms: MMR, IA-Select, xQuAD, risk/return, etc.
- Metrics: α-nDCG, nDCG-IA, ERR-IA, NRBP, subtopic recall, etc.
- Diversity is also studied in many other fields: biology, ecology, genetics, demographics, telecommunications, finance… recommender systems

The recommendation task

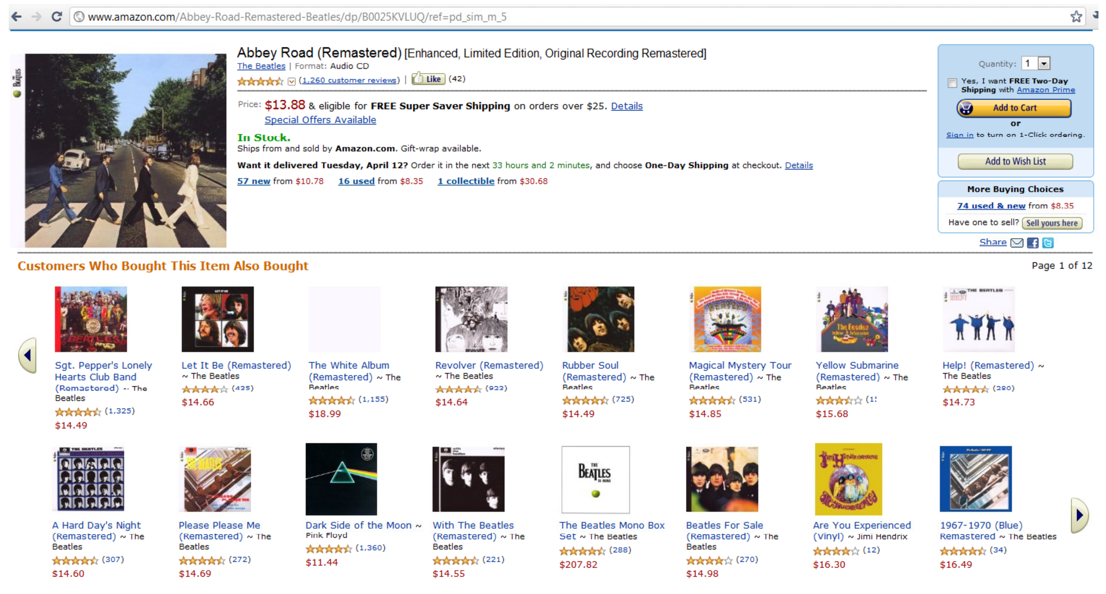
- Given:
  - A set of users $\mathcal{U}$, a set of items $\mathcal{I}$
  - A history of observed interaction (evidence of user preference for items)
    - Users accessing items: $\mathcal{L} \subset \mathcal{U} \times \mathcal{I} \times \mathcal{T}$ $\mathcal{T}$ = set of timestamps (e.g. Last.fm, Amazon)
    - Or users rating items: $r : \mathcal{U} \times \mathcal{I} \rightarrow \mathcal{V}$ $\mathcal{V}$ = set of rating values (e.g. Netflix, Amazon)
  - No query
- Predict:
  - A personalized ranking of items that each user may like
    - For each $u \in \mathcal{U} \rightarrow R = \langle i_1, i_2, \ldots, i_n \rangle$ $i_k \in \mathcal{I}$
  - Equivalently, define a retrieval function to rank items
    - $f : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}$ (often stated as rating prediction)
- Common methods: content-based, CF, kNN, matrix factorization…
  - Focused on accuracy, driven by similarity
- Evaluation: split user preference data into training and test

Recommendation accuracy vs. diversity

- Research in the Recommender Systems field has focused on accuracy in matching user tastes
- Yet diversity is as important or more in recommendation than in search
  - Accuracy alone is not enough to achieve useful recommendations in real scenarios
  - Novelty and diversity are key dimensions of utility, along with accuracy
  - Raising awareness in the field –diversity has become an important emerging research topic in Recommender Systems
- The rationale for search diversity also applies in recommendation
  - User needs are conveyed implicitly → they involve even more uncertainty than queries
  - A diverse recommendation increases the chances that at least some item is liked by avoiding a too narrow array of choice
  - The utility of recommended items is not mutually independent
- Further reasons to diversify
  - Diverse recommendations enrich the user experience
  - The chances that the user discovers new interests is increased
  - The business is enhanced by increasing the diversity of sales
- Diversity and accuracy understood as separate, opposing goals



User-item preference data in the recommendation task



Research on diversity in Recommender Systems

- Diversification approaches
  - Optimization of diversity/accuracy tradeoff
  - Objective functions, greedy optimization, promotion of long-tail items in the ranking, etc.
- Metrics
  - Average intra-list diversity (Ziegler 2005, Zhang 2008) $\text{ILD} = \frac{2}{|R|(|R|-1)} \sum_{i_j, i_k \in R,\, j<k} d(i_j, i_k)$
  - Aggregate diversity (Adomavicius 2011) $\text{AD} = \left| \bigcup_{u \in \mathcal{U}} R_u \right|$
  - Novelty (not the focus of our present work –which is on diversity)
  - Diversity metric framework, rank and relevance sensitiveness (Vargas 2011)
- Some gaps
  - Problem statement and formalization not quite the same as in search diversity
  - Not the same level of methodological consensus/convergence as in search diversity
  - Metrics are rank-insensitive (except Vargas 2011)
  - Diversity addressed independently from relevance: complementary accuracy metrics
  - Room for further studies on metric properties in general

- Since recommendation can be stated as an IR task, is it possible to find convergence of diversity theories, methods, and metrics from both fields?

From search diversity to recommendation diversity

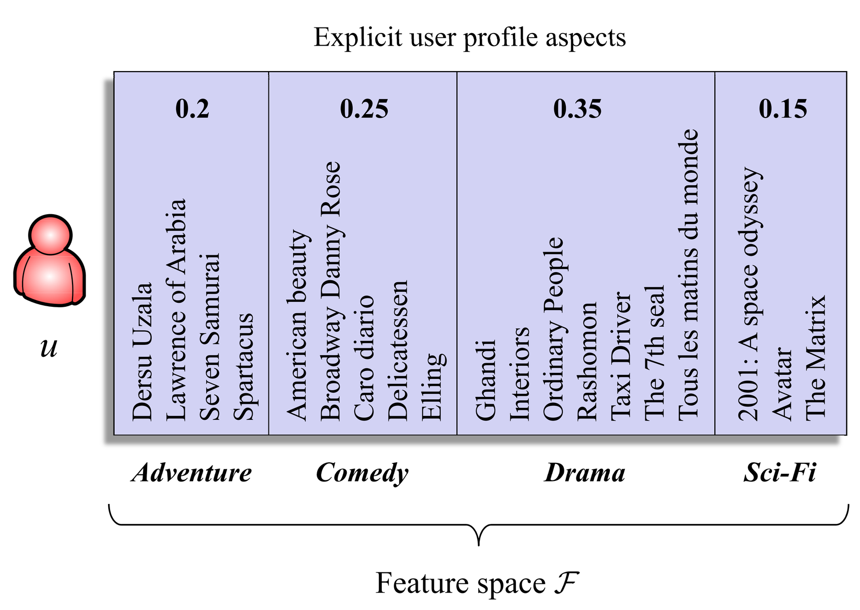
Mapping search diversity to recommendation diversity

- Search diversity is based on query uncertainty, intent… –but no query in the recommendation task!
- Notion of **user profile aspect** as an analogous to query aspect
  - A natural idea: the interests of a user may have many different sides and subareas: professional, politics, movies, travel, etc.
  - Different user preference aspects can be relevant or totally irrelevant at different times
    → uncertainty about what user interest area should play in a given context

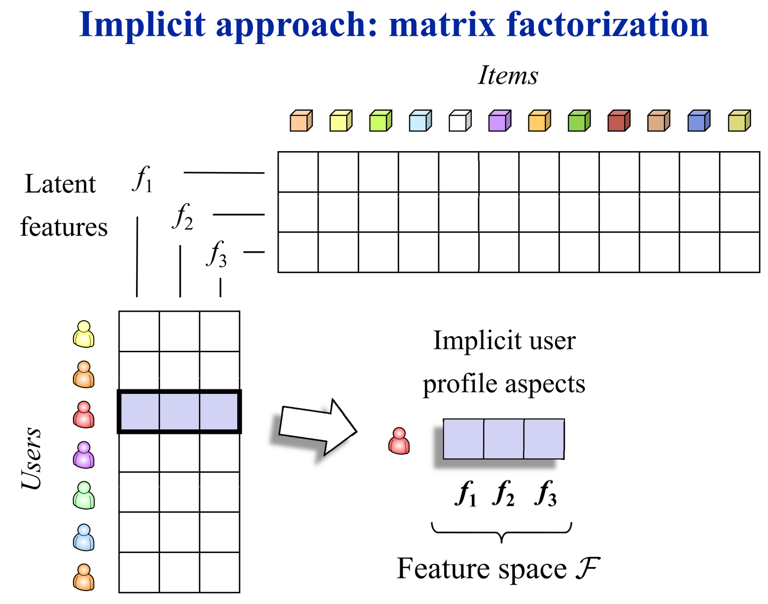
Search task	Recommendation task
Query (*representation of information need*)	*User profile* (*evidence of broad, implicit user need*)
Document	*Item* (*movie, book, music track*)
Document content (*words*)	*Item features* (*movie genre / director / cast, track artist, etc.*)
Relevance judgment	*User preference data* (*rating, purchase / access records*)
Subtopic, query aspect, intent, category	*User profile aspect*

- Problem: how to extract user profile aspects?

Explicit approach: available item features



Implicit approach: matrix factorization



Adaptation of search diversity techniques to recommender systems

Diversification algorithms

- Reranking baseline recommendation by greedy maximization of objective function
- IA-Select (Agrawal 2008), objective function:

$$\sum_{f \in \mathcal{F}} p(f|u)\, \hat{r}_{\text{norm}}(u,i)\, p(f|i) \prod_{j \in S} \left(1 - p(f|j)\, \hat{r}_{\text{norm}}(u,j)\right)$$

$$p(f|u) \sim \frac{\left|\{i \in \mathbf{u} \,|\, f \in \mathbf{i}\}\right|}{\sum_{g \in \mathcal{F}} \left|\{i \in \mathbf{u} \,|\, g \in \mathbf{i}\}\right|} \qquad p(f|i) \sim \frac{\chi_i(f)}{|\mathbf{i}|}$$

Example: $p(Comedy|u) \sim \frac{5}{20}$

- u Target user for recommendation
- R Baseline recommendation
- S Reranked recommendation
- $\mathcal{F}$ The feature space (explicit or implicit)
- $\hat{r}_{\text{norm}}$ The rating prediction function (i.e. retrieval function) of the recommender, normalized to [0,1]
- $\mathbf{u}$ The set of items accessed by user u (his user profile)
- $\mathbf{i}$ The set of features of item i
- $\chi_i(f) = \begin{cases} 1 & \text{if } f \in \mathbf{i} \\ 0 & \text{otherwise} \end{cases}$

- Maximum Marginal Relevance (MMR, Carbonell 1998), objective function:

$$(1-\lambda)\, \hat{r}_{\text{norm}}(u,i) + \lambda \operatorname*{avg}_{j \in S} \left(1 - sim(i,j)\right)$$

- sim An item similarity function (cosine on feature vectors in our experiments)

- Two scenarios are considered:
  - Feature data is explicit and known to the diversification method → the above algorithms use the explicit features as binary vectors
  - No feature knowledge, only user-item preference data is available → latent feature vectors are extracted by rating matrix factorization (binarized for probability estimations in IA-Select)

Diversity metrics

- α-nDCG (Clarke 2008)
  - User aspects from explicit features play the role of query "nuggets"
- Intent-aware metrics (Agrawal 2008)
  - User aspects from explicit features play the role of "categories", for instance:

Always use the explicit features in the metrics

$$\text{nDCG-IA} = \sum_{f \in \mathcal{F}} p(f|u)\, \text{nDCG}(u|f)$$

where nDCG(u|f) counts as relevant the items that u likes and have the feature f

Preliminary experiments

$\alpha = 0.5$

		α-nDCG@50		ERR-IA@50		nDCG-IA@50		ILD@50	
		kNN	MF	kNN	MF	kNN	MF	kNN	MF
λ = 0.5, IA-Sel	Explicit	0.1589	<u>0.1838</u>	0.0409	0.0516	0.0604	<u>0.0755</u>	<u>0.8659</u>	0.8734
λ = 0.5, IA-Sel	Latent	<u>0.1596</u>	0.1597	<u>0.0465</u>	0.0458	<u>0.0618</u>	0.0637	0.7951	0.7817
λ = 0.5, MMR	Explicit	0.1334	0.1652	0.0367	0.0431	0.0461	0.0555	0.8601	<u>0.8761</u>
λ = 0.5, MMR	Latent	0.1320	0.1742	0.0373	<u>0.0528</u>	0.0492	0.0705	0.7906	0.7740
Baseline RS		0.1213	0.1451	0.0352	0.0425	0.0440	0.0561	0.7787	0.7655

All differences to baseline are statistically significant (Wilcoxon p < 0.005), except gray cells

- Dataset: Movielens 100K ratings, ~1K users, ~1.6K items
- Recommender baselines: user-based kNN, matrix factorization (MF)
- Diversification algorithms: rerank top 500 recommended items
- Explicit features: movie genres
- Relevant items ≡ rating > 3
- Metrics based on explicit features (ILD taking Jaccard distance)
- 5-fold 80% – 20% training-test splits provided in the dataset
- Consistent behavior of diversification, improving baselines
- Diversification on latent profile aspects competitive w.r.t. explicit!

Conclusions and advantages

- Adaptation of search diversity techniques to recommender systems
- New rationale for diversity in recommender systems (theory and models)
- New diversity metrics for recommender systems: α-nDCG, IA metrics
- Introduction of rank sensitiveness in diversity metrics
- Introduction of relevance and diversity in single metrics
- Moving towards shared consensus and common evaluation methodologies
- Further diversification algorithms (e.g. xQuAD, Santos 2010), further unification of recommendation novelty and diversity metrics (Vargas 2011)
- Future direction: user profile aspect extraction is a rich research problem

References

(Adomavicius 2011) G. Adomavicius and Y. Kwon. Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. IEEE Transactions on Knowledge and Data Engineering, In press.

(Agrawal 2008) R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. WSDM 2009, Barcelona, Spain, 2009, pp. 5-14.

(Carbonell 1998) J. G. Carbonell and J. Goldstein. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. SIGIR 1998, Melbourne, Australia, 1998, pp. 335-336.

(Chen 2006) H. Chen and D. R. Karger. Less is More. SIGIR 2006, Seattle, WA, USA, 2006, pp. 429-436.

(Clarke 2008) C. L. A. Clarke et al. Novelty and diversity in information retrieval evaluation. SIGIR 2008, Singapore, 2008, pp. 659-666.

(Santos 2010) R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting Query Reformulation for Web Search Result Diversification. WWW 2010, Raleigh, NC, USA, April 2010, pp. 881-890.

(Vargas 2011) S. Vargas and P. Castells. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. RecSys 2011, Chicago, IL, USA, October 2011.

(Wang 2009) J. Wang and J. Zhu. Portfolio theory of information retrieval. SIGIR 2009, Boston, MA, USA, pp. 115-122.

(Zhai 2003) C. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. SIGIR 2003,Toronto, Canada, 2003, pp. 10-17.

(Zhang 2008) M. Zhang and N. Hurley. Avoiding Monotony: Improving the Diversity of Recommendation Lists. RecSys 2008, Lausanne, Switzerland, October 2008, pp. 123-130.

(Ziegler 2005) C-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. WWW 2005, Chiba, Japan, May 2005, pp. 22-32.