



Novelty and Diversity Enhancement and Evaluation in Recommender Systems and Information Retrieval

Saúl Vargas

Universidad Autónoma de Madrid

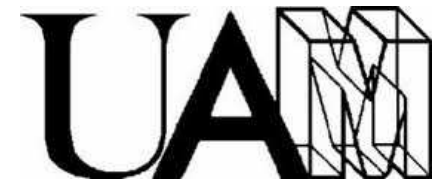
saul.vargas@uam.es

IRG



About me

- PhD student and Teaching Assistant at Universidad Autónoma de Madrid (Spain).
- Supervised by Professor Pablo Castells.
- In 2012 I presented my Master Thesis, which started this doctoral research.



Recommender Systems (1/2)



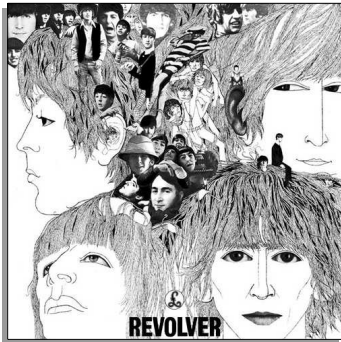
Recommender Systems (2/2)

- Personalized Information Retrieval Systems.
- No query, information need is implicit:
 - *“I would like to listen to (new) music.”*
 - *“I would like to watch a movie.”*
 - *“What products would I be interested in buying?”*
- Previous interactions as indicator of user preferences.

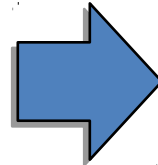
Diversity and Novelty in Recommendations

You bought

Revolver



Abbey Road



Recommendations:



Rubber Soul



With The Beatles



Beatles for Sale



Let it be



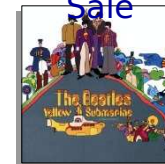
Help!



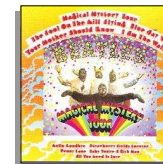
A Hard Day's Night



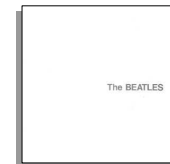
Sgt. Pp's Lonely Hearts Club Band



Yellow Submarine



Magical Mystery Tour



The White Album



Please Please me



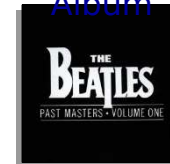
1967-1970 (Blue)



1962-1966 (Red)



Past Masters



Past Masters Vol 2



Dark Side of the Moon



Some Girls



Bob Dylan

...

Diversity in Search



State of the Art

“Each diversity or novelty paper in RS has its own definition, metrics and methods”

- Lack of formalization and standardization in Recommender System.
- There are few studies connecting Search Result Diversification with Diversity in Recommender Systems.

Research Methodology

- Comprehensive study of the State of the art for both IR and RS.
- Definition and formalization of tasks.
- Development of metrics and algorithms.
- Offline experiments:
 - Publicly available data sets:
 - MovieLens1M (movies, 6K users, 4K items, 1M ratings).
 - Netflix (movies, 480K users, 18K items, 100M ratings).
 - MSD (music, 1M users, 380K items, 48M play counts).
- Online evaluation:
 - Crowdsourced evaluation (Crowdfunder, Amazon Mechanical Turk)

Research Goals

- Unification and formalization of novelty and diversity metrics for Recommender Systems.
- Connection between principles in Search Result Diversification and Diversity in Recommender Systems.
- New novelty and diversity enhancement methods.

Unified Framework for Diversity and Novelty Metrics (1/2)

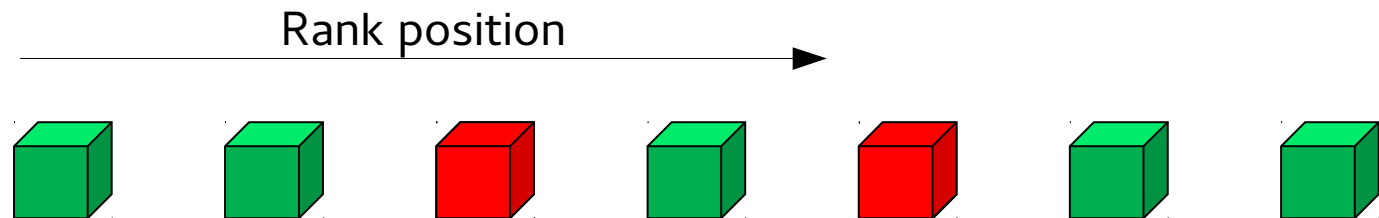
- (RecSys 2011) Expressing many novelty and diversity metrics for RS and incorporate rank and relevance awareness:

$$m(R|\theta) = C \sum_{i \in R} p(\text{choose}|i) \text{nov}(i|\theta)$$

- Item novelty models:
 - θ ="all users" \rightarrow global novelty (anti-popularity)
 - θ ="user profile" \rightarrow personalized novelty
 - θ ="previous recommendation" \rightarrow temporal diversity
 - θ ="other items in recommendation" \rightarrow intra-list diversity

Unified Framework for Diversity and Novelty Metrics (2/2)

- Choice model: rank and relevance-aware!
 - An item is chosen if it is seen and found relevant.
 - Items not chosen, however novel, do not contribute to the recommendation novelty.



$disc(k) = 0.9^{k-1}$	1	0.90	0.81	0.73	0.66	0.59	0.53
$p(rel i, u) = r_{u,i}$	1	1	0	1	0	1	1
$p(choose i, u, R)$	1	0.9	0	0.73	0	0.59	0.53

Connection between IR and RS (1/3)

- Search Result Diversification:
 - Avoiding redundant documents.
 - Coping with query ambiguity.
 - Coping with query underspecification.
- Diversity in Recommendation Lists:
 - Avoiding redundant items.
 - Users usually have different tastes.
 - Users expect varied recommendations.

Connection between IR and RS (2/3)

- A straightforward translation of concepts...

- Document → Item
- Query → User
- Subtopics → Tastes



0.2	0.25	0.35	0.15
Dersu Uzala Lawrence of Arabia Seven Samurai Spartacus	American beauty Broadway Danny Rose Caro diario Delicatessen Eiling	Ghandi Interiors Ordinary People Rashomon Taxi Driver The 7th seal Tous les matins du monde	2001 : A space odyssey Avatar The Matrix
<i>Adventure</i>	<i>Comedy</i>	<i>Drama</i>	<i>Sci-Fi</i>

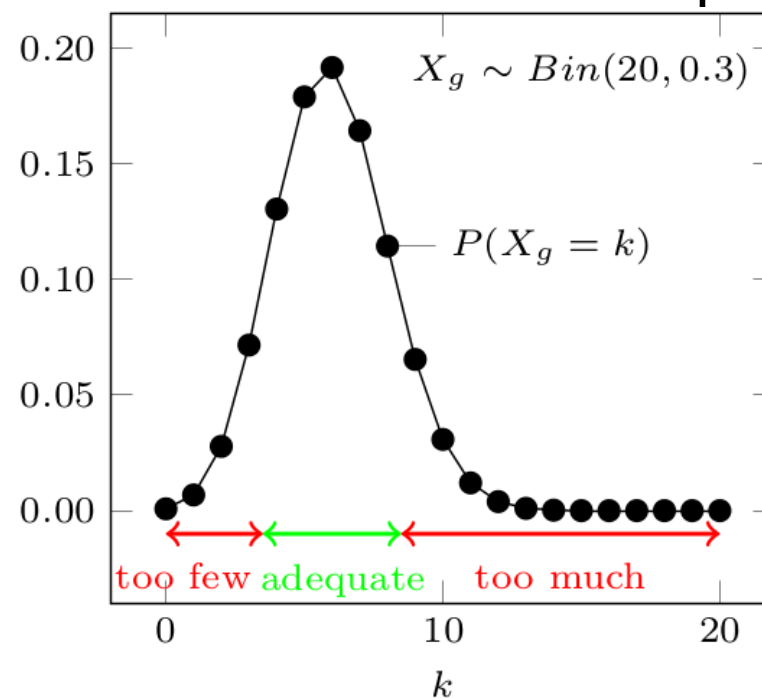
- ...allows to adapt IR diversity metrics and diversification techniques to RS (**SIGIR 2011**):

- Metrics: ERR-IA, α -nDCG, S-recall, ...
- Algorithms: MMR, IA-Select, xQuAD, ...

Connection between IR and RS (3/3)

- IR diversity metrics present some inconveniences:
 - They consider a infinite size ranking of documents, they do not “target” small, fixed-size results lists.
 - Presenting a redundant document w.r.t. some subtopics is fine as long as it covers other non-redundant subtopics.

- We propose a **Binomial Framework** for considering coverage, redundancy and size-awareness in diversity in RS (RecSys 2014??).



Novelty and Diversity Enhancement (1/3)

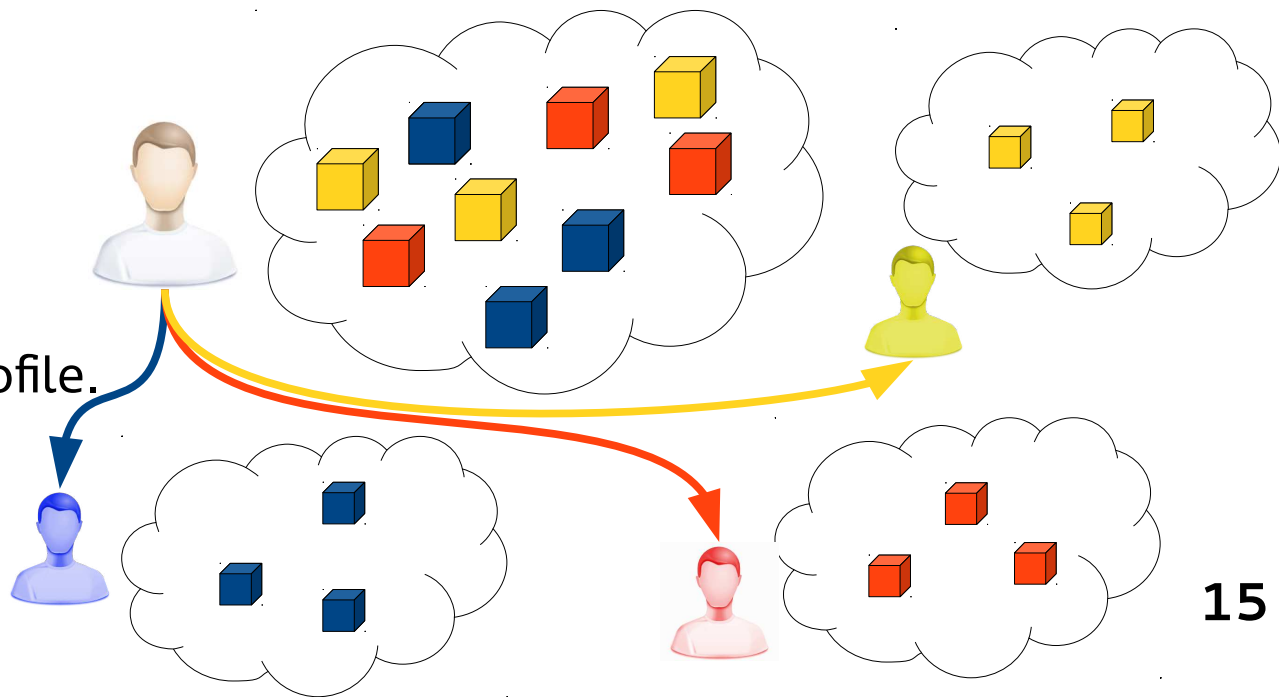
- Explicit relevance models for intent-oriented search result diversification (**SIGIR 2012**).
 - Alternative formulation of well-known aspect-based diversification algorithms: IA-Select and xQuAD.
 - From a generative model to a relevance model.
$$p(d|c, q) \rightarrow p(rel|d, c, q)$$
$$p(i|c, u) \rightarrow p(rel|i, c, u)$$
 - Competitive or better performance than the original algorithms.

Novelty and Diversity Enhancement (2/3)

- RS diversification with user-sub-profiles (**OAIR 2013**).
 - xQuAD: query reformulations (sub-queries) as proxies for subtopics.
 - We propose sub-profiles as an analogy to sub-queries.

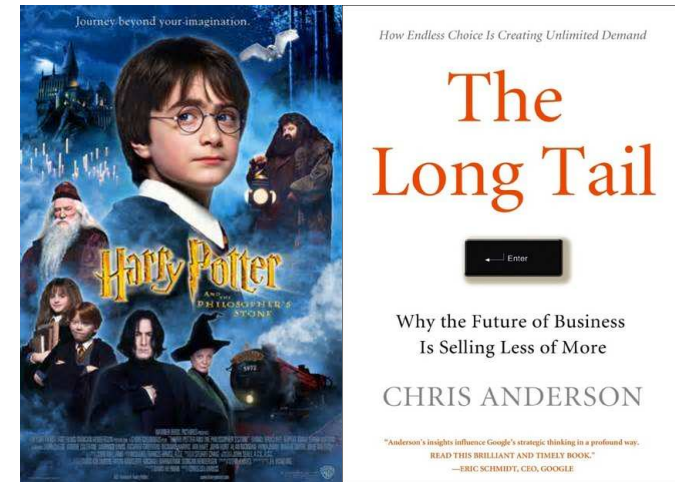
- Method:

- Profiles are partitioned.
- Recommendations are created for each sub-profile.
- Combined with xQuAD.



Novelty and Diversity Enhancement (3/3)

- Recommending users to items in Collaborative Filtering (**RecSys 2014??**):
 - Improving Item novelty
 - Improving Sales diversity
 - Concept: recommending users to items.
 - Two approaches:
 - User-item rating matrix transposition: inverted neighborhoods for nearest neighbors approaches.
 - Probabilistic reformulation: isolate the popularity bias by means of the Bayes rule.



$$p(i|u) = \frac{p(u|i)}{p(u)}p(i)$$

Open Issues

- Connection between IR diversity and RS diversity:
 - Further analysis required?
 - Other ways to exploit the similarities between them?
 - What other fundamental differences are there?
 - Ideas from RS to IR?
- Conducting online evaluations:
 - How to do perform them?
 - What to evaluate? Metrics, algorithms?

Thank you for you attention!