## Diversity and Novelty in Web Search, Recommender Systems and Data Streams

**WSDM 2014 TUTORIAL**

Rodrygo L. T. Santos – UFMG, Brazil
Pablo Castells – UAM, Spain
Ismail Sengor Altingovde – METU, Turkey
Fazli Can – Bilkent, Turkey

## Outline

- PART I
  - Background (Santos)
- PART II
  - Diversity and Novelty in Search (Altingovde+Santos)
- PART III
  - Diversity and Novelty in Recommendation (Castells)
- PART IV
  - Diversity and Novelty in Data Streams (Can)

## Diversity and Novelty in Web Search, Recommender Systems and Data Streams

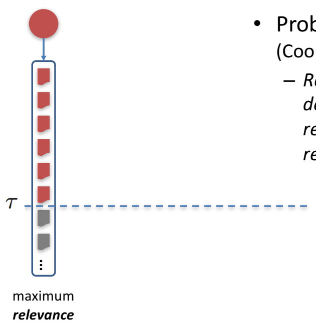
**PART I: BACKGROUND**

Rodrygo L. T. Santos – UFMG, Brazil

## Information Retrieval (IR)

- The goal of an IR system is to provide relevant items given a user's information need
  - Relevant search results given a user's query
  - Relevant recommendations given a user's profile
- Determining relevance is the key challenge (Goffman, 1964)
  - A user's prerogative
  - Can be estimated at best

## Relevance-oriented Retrieval




- Probability Ranking Principle (Cooper, 1971; Robertson, 1977)
  - *Ranking information items by decreasing probability of relevance results in optimal retrieval effectiveness*

## Ranking Optimality

- PRP is optimal under certain assumptions (Gordon & Lenk, 1991, 1992)
  - A1. The probability of relevance is estimated with certainty, with no measure of risk
  - A2. The probability of relevance is estimated independently for every document

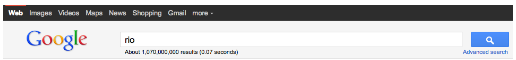
## Limiting Assumption A1

- Assumption
  - A1. The probability of relevance is estimated with certainty, with no measure of risk
- Limitation: ***ambiguity***
  - Information needs and items are ambiguously represented (Turtle & Croft, 1996)

## Ambiguity


***Which 'rio'?***

## Ambiguity




***Brazilian city?***

## Ambiguity


***Animation movie?***

## Ambiguity

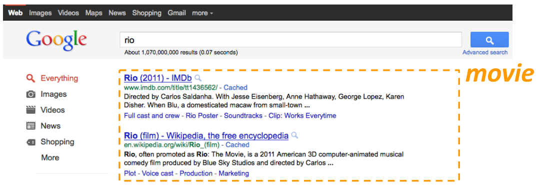

- Wikipedia lists over 50 meanings for 'rio'…
  - … but we can only display '10 blue links'
- 16% of all web search queries are ambiguous (Song, 2009)
  - Every query is underspecified to some extent (Cronen-Townsend & Croft, 2002)

## Limiting Assumption A2

- Assumption
  - A2. The probability of relevance is estimated independently for every document
- Limitation: ***redundancy***
  - *"The relationship between a document and a query is necessary but not sufficient to determine relevance"* (Goffman, 1964)

*Pablo Castells*
*Escuela Politécnica Superior*
*Universidad Autónoma de Madrid*

## Redundancy



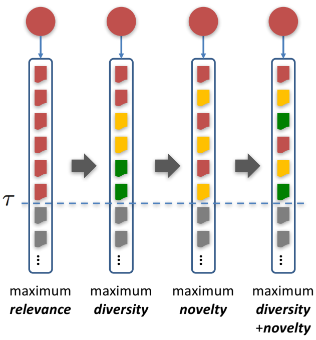

- Any need for more results about the movie?
  - Users are unlikely to inspect the results any further once they find something relevant (Craswell et al., 2008)

## Diversity and Novelty

- Diversity
  - *"the need to resolve **ambiguity**" [in the retrieval request]*
- Novelty
  - *"the need to avoid **redundancy**" [in the retrieval response]*

(Clarke et al., 2008)

## Diversity and Novelty




- Diversity
  - *"cover as many request aspects as possible"*
- Novelty
  - *"cover request aspects as early as possible"*
- *Diversity+Novelty*
  - *"cover as many aspects as early as possible"*

## Greedy Best First Search

- In each iteration, select the document $d$ that maximizes a **scoring/ranking function** $f$
- Algorithm outline:
  - R: input ranking
  - D: final ranking
  - τ: ranking cutoff

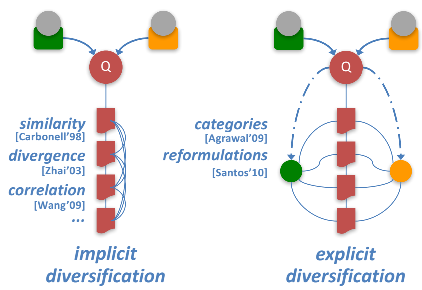
$D \leftarrow \emptyset$
**while** $|D| < \tau$ **do**
  $d \leftarrow \arg\max_{d \in R \setminus D} f(q, d, D)$
  $R \leftarrow R \setminus \{d\}$
  $D \leftarrow D \cup \{d\}$

## Diversity and Novelty in Web Search, Recommender Systems and Data Streams

**PART II:**
**DIVERSITY AND NOVELTY IN WEB SEARCH**

Ismail Sengor Altingovde – METU, Turkey
Rodrygo L. T. Santos – UFMG, Brazil

## How to Diversify?



MMR
Max-Sum Dispersion
MPT
Swap
Faceted Retrieval

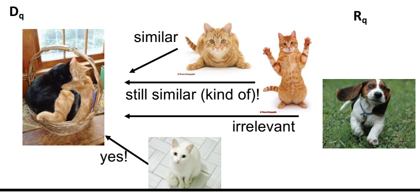
## Implicit Diversification Methods

Greedy Local Search
GMC
GRASP
QPRR
Clustering

## Maximal Marginal Relevance (MMR)

[Carbonel and Goldstein, SIGIR 1998]

- Selects the document $d$ that maximizes:
  - $f_{\text{MMR}}(q, d, D_q) = \lambda \, \text{rel}(q,d) - (1-\lambda) \max_{d_i \in D_q} \text{sim}(d, d_i)$
- $f_{\text{MMR}}$ discounts the document's relevance by its max similarity to already selected docs.




RC
OptSelect
PM-2
IA-SELECT
RAB
RCB
xQuAD

## Explicit Diversification Methods

## Category Coverage
(Agrawal et al., 2009)

- Basic idea: score a document proportionally to its coverage of different query categories

dmoz open directory project

**Arts** Movies, Television, Music...
**Business** Jobs, Real Estate, Investing...
**Computers** Internet, Software, Hardware...
**Games** Video Games, RPGs, Gambling...
**Health** Fitness, Medicine, Alternative...
**Home** Family, Consumers, Cooking...
**Kids and Teens** Arts, School Time, Teen Life...
**News** Media, Newspapers, Weather...
**Recreation** Travel, Food, Outdoors, Humor...
**Reference** Maps, Education, Libraries...
**Regional** US, Canada, UK, Europe...
**Science** Biology, Psychology, Physics...
**Shopping** Clothing, Food, Gifts...
**Society** People, Religion, Issues...
**Sports** Baseball, Soccer, Basketball...
**World** Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...

## Category Coverage
(Agrawal et al., 2009)

- Basic idea: score a document proportionally to its coverage of different query categories

$$f_{\text{IA-Select}}(q, d, \mathcal{D}_q) = \sum_{\underbrace{c \in \mathcal{T}}_{\text{query categories}}} \underbrace{f(c|q, \mathcal{D}_q)}_{\text{category utility}} \; \underbrace{f(d|q, c)}_{\text{category coverage}}$$

## Sub-Query Coverage
(Santos et al., 2010)

- Basic idea: score a document proportionally to its relevance to different sub-queries

*suggested sub-queries*

rio |
rio **car**
rio **duran duran**
rio **de janeiro**
rio **movie**
rio **grande**
rio **salado**
rio **las vegas**
rio **hondo**
rio **bravo**
rio **tinto**

*related sub-queries*

Related searches for **rio**:

| | |
|---|---|
| rio **las vegas** | rio **de janeiro** |
| rio **mp3 player** | rio **brazil** |
| rio **grande** | rio **salado** |
| rio **tinto** | **kia** rio |
| rio **gaithersburg** | rio **de janero** |
| rio **hotel** | rio **casino** |
| rio **carnival** | **24**rio **car** |

Pablo Castells
Escuela Politécnica Superior
Universidad Autónoma de Madrid

## Sub-Query Coverage
(Santos et al., 2010)

- Basic idea: score a document proportionally to its relevance to different sub-queries

$$f_{\text{xQuAD}}(q,d,\mathcal{D}_q) = \underbrace{(1-\lambda)\,p(d|q)}_{\text{document relevance}} + \underbrace{\lambda}_{\text{diversification trade-off}} \underbrace{\sum_{s\in\mathcal{S}_q} p(s|q)\,p(d|q,s) \prod_{d_j\in\mathcal{D}_q}(1-p(d_j|q,s))}_{\text{document diversity}}$$

## Sub-Query Coverage
(Santos et al., 2010)

- Basic idea: score a document proportionally to its relevance to different sub-queries

$$f_{\text{xQuAD}}(q,d,\mathcal{D}_q) = \underbrace{(1-\lambda)\,p(d|q)}_{\text{document relevance}} + \underbrace{\lambda}_{\text{diversification trade-off}} \sum_{\underbrace{s\in\mathcal{S}_q}_{\text{sub-queries}}} \underbrace{p(s|q)}_{\text{sub-query importance}}\,\underbrace{p(d|q,s)}_{\text{document coverage}} \underbrace{\prod_{d_j\in\mathcal{D}_q}(1-p(d_j|q,s))}_{\text{document novelty}}$$

- Top performing approach at TREC 2009-2012

## Diversity Evaluation

**benchmarks**

**D metrics**

**cascade metrics**

**IA metrics**

## Evaluation Benchmarks

- Basic test collection layout
  - Set of documents
  - Set of queries
  - Set of relevance assessments
    - (query, document) → label
    - (query, document, aspect) → label
- Advanced features
  - Graded relevance labels
  - Aspect intents
  - Aspect probabilities
    - (query, aspect) → label

## Evaluation Benchmarks

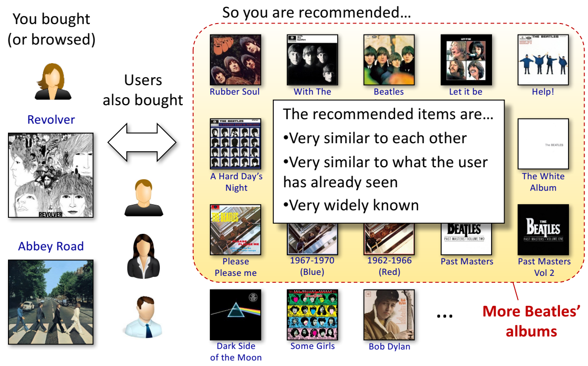
- TREC 6-8 Interactive track (Over, 1997, 1998; Hersh & Over, 1999)
  - 20 queries
- TREC 2009-2012 Web track (Clarke et al., 2009-2012)
  - 200 queries
- NTCIR 9-10 Intent task, NTCIR 11 iMine task (Sakai & Song, 2012)
  - 200 + 50 queries

## Diversity and Novelty in Web Search, Recommender Systems and Data Streams

**PART III:**
**DIVERSITY AND NOVELTY IN RECOMMENDER SYSTEMS**

Pablo Castells – UAM, Spain

## Diversity in recommender systems




## Diversity in recommender systems

Search diversity:

"Avoid redundancy of possible user intents (aspects) as a means to cope with the uncertainty in the query"

## Diversity in recommender systems

Recommendation diversity:

"Avoid redundancy of possible user intents (aspects) as a means to cope with the uncertainty ~~in the query~~"

## Diversity in recommender systems

Recommendation diversity:

"Avoid redundancy of possible user intents (aspects) as a means to cope with the uncertainty ~~in the query~~ in the observed evidence of user interests"

## Why diversify recommendations

(Vargas, Castells & Vallet SIGIR 2011, 2012)

For better **system** effectiveness ("a safer bet")

- Uncertainty in user preferences
  - Ambiguity, underspecification
  - Preferences are multiple, dynamic, contextual…
  - Much broader needs than in search
- Increase chances of at least some relevant item
- Equivalent to search diversity principle
  - Infer the user's intent based on what the user asks vs. what the user does

## Why diversify recommendations

(McAlister 1982, and many more…)

For the sake of it: direct **user** satisfaction

- Natural variety-seeking drive in human behavior
  - Within a recommendation and over time
  - Desire for the unfamiliar, alternation among the familiar
  - Ideal level of stimulation
- Multiple simultaneous needs and tastes
- Broaden the user's horizon
- The task is often explicitly about discovery

Diversity but also **novelty**

## Why diversify recommendations
(Fleder Mgt. Sci. 2009, McNee CHI 2006)

For enhanced **business** performance

- Sales diversity: mitigate risk, expand the business
- Long tail: draw revenues from market niches
  - "Sell less of more"
  - Higher profit margin on cheaper long-tail products

## Search diversity applied to recommender systems

## Applying **aspect diversity** to recommendation
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

Items, $i$

| Users | | | | $i$ | | | |
|---|---|---|---|---|---|---|---|
| | 4 | | 4 | 2 | | 2 | 2 |
| | 1 | 4 | 4 | | 4 | | |
| $u$ | 4 | 3 | | 2 | 5 | | 2 |
| | 4 | 3 | 3 | | | 2 | 2 |
| | | 1 | 1 | 5 | 1 | 5 | 5 |

## Applying **aspect diversity** to recommendation
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

Items, $i$

| Users | | | | $i$ | | | |
|---|---|---|---|---|---|---|---|
| | 4 | | 4 | 2 | | 2 | 2 |
| | 1 | 4 | 4 | | 4 | | |
| $u$ | 4 | 3 | | 2 | 5 | | 2 |
| | 4 | 3 | 3 | | | 2 | 2 |
| | | 1 | 1 | 5 | 1 | 5 | 5 |

← User profile

## Applying **aspect diversity** to recommendation
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

Items, $i$

| | | | | $i$ | | | |
|---|---|---|---|---|---|---|---|
| $u$ | 4 | 3 | | 2 | 5 | | 2 |

← User profile

## Applying **aspect diversity** to recommendation
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

Items, $i$

| | | | | $i$ | | | |
|---|---|---|---|---|---|---|---|
| $u$ | 4 | 3 | | 2 | 5 | | 2 |

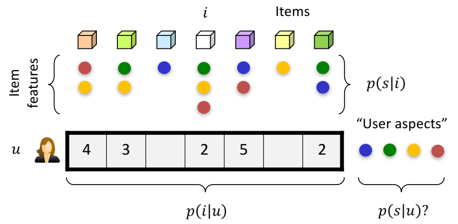
"User aspects"?

Pablo Castells
Escuela Politécnica Superior
Universidad Autónoma de Madrid

## Applying **aspect diversity** to recommendation

(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

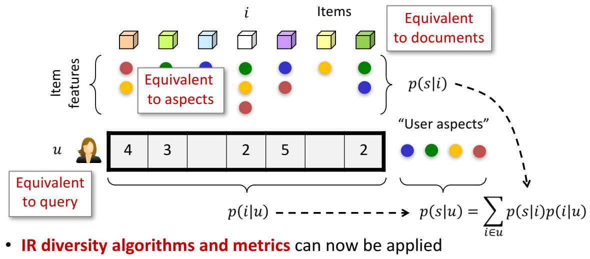
- Similar to IA-Select (Agrawal 2009), aspects from **item features**
- Using a "meaningful" item feature space, derive user aspect distributions




## Applying **aspect diversity** to recommendation

(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; Vargas & Castells OAIR 2013; …)

- Similar to IA-Select (Agrawal 2009), aspects from **item features**
- Using a "meaningful" item feature space, derive user aspect distributions




- **IR diversity algorithms and metrics** can now be applied
- Other approaches to user interest subdivision have been considered
  - E.g. feature-based slicing, clustering (Hurley WI 2009)

## Applying **implicit IR diversification**

(Smyth 2001, Ziegler 2005, Hurley TOIT 2011, Shi SIGIR 2012, Vargas et al SIGIR 2011, …)

- Implicit IR diversification re-ranking algorithms can be applied even more straightforwardly to recommendation
- In fact some state of the art recommendation diversification algorithms are essentially MMR
  - Just a baseline scoring function and a similarity measure are needed

$$f(i|D,u) = (1-\lambda)\,\hat{r}(u,i) + \frac{\lambda}{|D|}\sum_{j \in D} d(i,j)$$

- MPT has also been adapted to recommendation

## Recommendation novelty and diversity evaluation and metrics

## Novelty and diversity metrics

Recommendation diversity metrics…

•Are not defined in terms of ranking

•Do not involve relevance

Different metrics for different notions

•Three most common: intra-list diversity, unexpectedness, inverse popularity

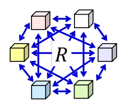
•Several other, more particular metrics

## Novelty and diversity metrics

(Smyth & McClave ICCBR 2001, Ziegler et al WWW 2005, etc.)

**Intra-list diversity:** average pairwise distance

$$\mathrm{ILD} = \frac{2}{|R|(|R|-1)} \sum_{\substack{i,j \in R \\ i \neq j}} d(i,j)$$




*Internal diversity*

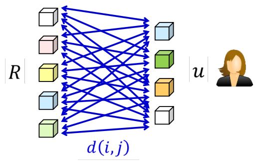
Some distance measure, e.g. $d(i,j) = 1 - sim(i,j)$ with $sim$ = cosine, Jaccard, etc. on item features

## Novelty and diversity metrics

(Adamopoulos ACM TIST 2014, Hurley & Zhang TOIT 2011, Zhang et al WSDM 2012, etc.)

**Unexpectedness:** average distance to items in user profile

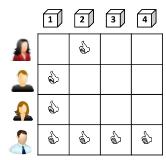
$$\text{Unexp} = \frac{1}{|R||u|}\sum_{\substack{i \in R \\ j \in u}} d(i,j)$$




## Toy example

**System's input**

User-item interaction (e.g. items = music)



Item features (e.g. music styles)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| jazz | 1 | 0 | 1 | 1 |
| reggae | 0 | 1 | 0 | 0 |
| bossa nova | 1 | 0 | 1 | 1 |
| ⋮ | | | | |

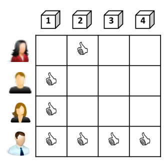
**System's recommendations**

| | |
|---|---|
| 1, 3 | 2, 3 |

## Toy example

**System's input**

User-item interaction (e.g. items = music)



Item features (e.g. music styles)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| jazz | 1 | 0 | 1 | 1 |
| reggae | 0 | 1 | 0 | 0 |
| bossa nova | 1 | 0 | 1 | 1 |
| ⋮ | | | | |

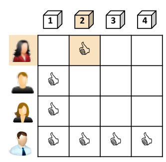
**System's recommendations**

| | | |
|---|---|---|
| | 1, 3 | 2, 3 |
| **ILD** | **0** | |

## Toy example

**System's input**

User-item interaction (e.g. items = music)



Item features (e.g. music styles)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| jazz | 1 | 0 | 1 | 1 |
| reggae | 0 | 1 | 0 | 0 |
| bossa nova | 1 | 0 | 1 | 1 |
| ⋮ | | | | |

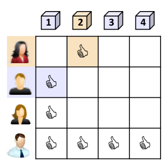
**System's recommendations**

| | | |
|---|---|---|
| | 1, 3 | 2, 3 |
| ILD | 0 | |
| **Unexp** | **1** | |

## Toy example

**System's input**

User-item interaction (e.g. items = music)



Item features (e.g. music styles)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| jazz | 1 | 0 | 1 | 1 |
| reggae | 0 | 1 | 0 | 0 |
| bossa nova | 1 | 0 | 1 | 1 |
| ⋮ | | | | |

**System's recommendations**

| | | |
|---|---|---|
| | 1, 3 | 2, 3 |
| ILD | 0 | **1** |
| Unexp | 1 | **0.5** |

## Novelty and diversity metrics

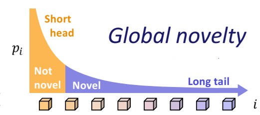
(Zhou et al PNAS 2010, Vargas & Castells RecSys 2011, Zhang et al WSDM 2012, etc.)

**Inverse popularity:** mean self-information → recommend in the long tail

How "not popular" are the recommended items on average

$$\text{MSI} = -\frac{1}{|R|}\sum_{i \in R} \log_2 p_i$$

Popularity of $i$




$$p_i = p(known|i) \sim |\{u \in \mathcal{U} | r(u,i) \in \mathcal{R}\}| / |\mathcal{U}|$$

$$p_i = p(i|known) \sim |\{u \in \mathcal{U} | r(u,i) \in \mathcal{R}\}| / |\mathcal{R}|$$

Equivalence

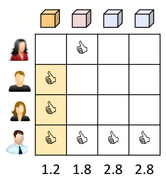
Set of all observed ratings

$$p(i|known) \sim p(known|i) / \textstyle\sum_{j \in \mathcal{I}} p(known|j) \propto p(known|i)$$

## Toy example

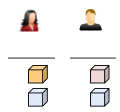
**System's input**

User-item interaction




e.g. $p(a|known) \sim 3/|\mathcal{R}| = 3/7$

**System's recommendations**

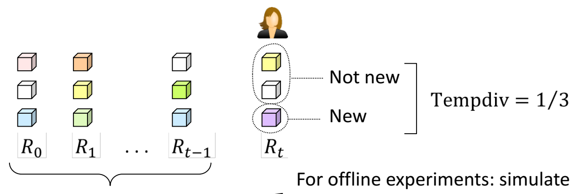


| | | |
|---|---|---|
| ILD | 0 | 1 |
| Unexp | 1 | 0.5 |
| **MSI** | **2.01** | **2.30** |

## Other metrics: temporal diversity

(Lathia et al SIGIR 2010)

- Ability of the recommender system to evolve over time
- Measured as the ratio of different recommended items over time $\text{Tempdiv} = |R_t - \cup_{t'<t} R_{t'}|/|R_t|$



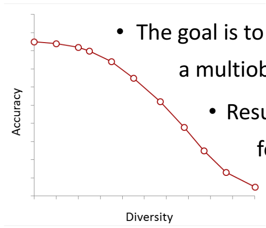

For offline experiments: simulate by temporal splits of data

## Recommendation novelty and diversity enhancement algorithms

## Novelty and diversity enhancement algorithms

Diversity and accuracy viewed as opposing objectives

- Enhancing diversity is expected to involve some sacrifice in accuracy
  - The goal is to achieve an optimal trade-off: a multiobjective optimization problem
    - Results assessed by two metrics, for relevance vs. diversity/novelty




## Diversity enhancement

(Smyth & McClave ICCBR 2001, Ziegler et al WWW 2005; Hurley & Zhang TOIT 2011)

- Enhance diversity by greedy maximization of:

$$f(i|D,u) = (1-\lambda)\,\hat{r}(u,i) + \frac{\lambda}{|D|}\sum_{j\in D} d(i,j)$$

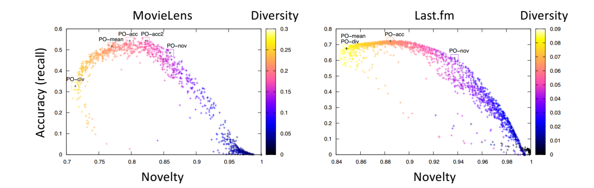
Tested as cosine on item features (Amazon's book taxonomy, trip characteristics)

- Equivalent to MMR
- Evaluated by ILD
- Variant of approach (Hurley & Zhang 2011)

## Novelty and diversity enhancement

(Ribeiro et al RecSys 2012)

- Novelty and diversity by linear recommender hybridization
  - Pareto optimization problem: multi-objective maximization of **accuracy & novelty & diversity** → evolutionary algorithm
- Tests with ensemble of 8 common recommender algorithms
  - Find the Pareto frontier on tradeoffs between the 3 metrics
  - Effective individuals better than constituents and baseline ensembles




Images from Ribeiro, M. T., Lacerda, A., Veloso, A., Ziviani, N. Pareto-efficient hybridization for multi-objective recommender systems. RecSys 2012.

## Unified perspective

## Tentative definitions

Diversity
- Generally applies to a set of items: how **different** the items are **from each other**
- Variants: simple, aggregate, inter-user, inter-system, temporal…

Novelty
- Refers to an item, **different from prior experience**
- Can apply to a set as the average novelty of its elements
- Variants : different, not known, unexpected / surprising / unfamiliar, serendipitous…

## Conclusions

- Novelty & diversity as important in recommendation as in search
- IR diversity principles, metrics and algorithms can be applied
- Plus further particular motivation and techniques
- Wide variety of novelty and diversity metrics and methods
- Most can be unified in a common scheme