Discovering Relevant Preferences in a Personalised Recommender System using Machine Learning Techniques

Alejandro Bellogín, Iván Cantador, Pablo Castells, Álvaro Ortigosa

Escuela Politécnica Superior
Universidad Autónoma de Madrid
Campus de Cantoblanco, 28049 Madrid, Spain
{alejandro.bellogin, ivan.cantador, pablo.castells, alvaro.ortigosa}@uam.es

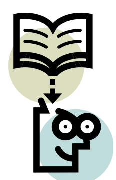
Motivation

- Complexity of recommender systems
- Several variables affect the effectiveness of recommendations
  - Apply personalisation?
  - Use current context?
  - Consider all the users equally?
  - Consider all the items similarly?
  - …
- Effectiveness is achieved by adequately handling these variables, but it is also an issue of knowing how relevant each one is
- Can we learn which preferences are relevant to achieve effective recommendations?

Our approach

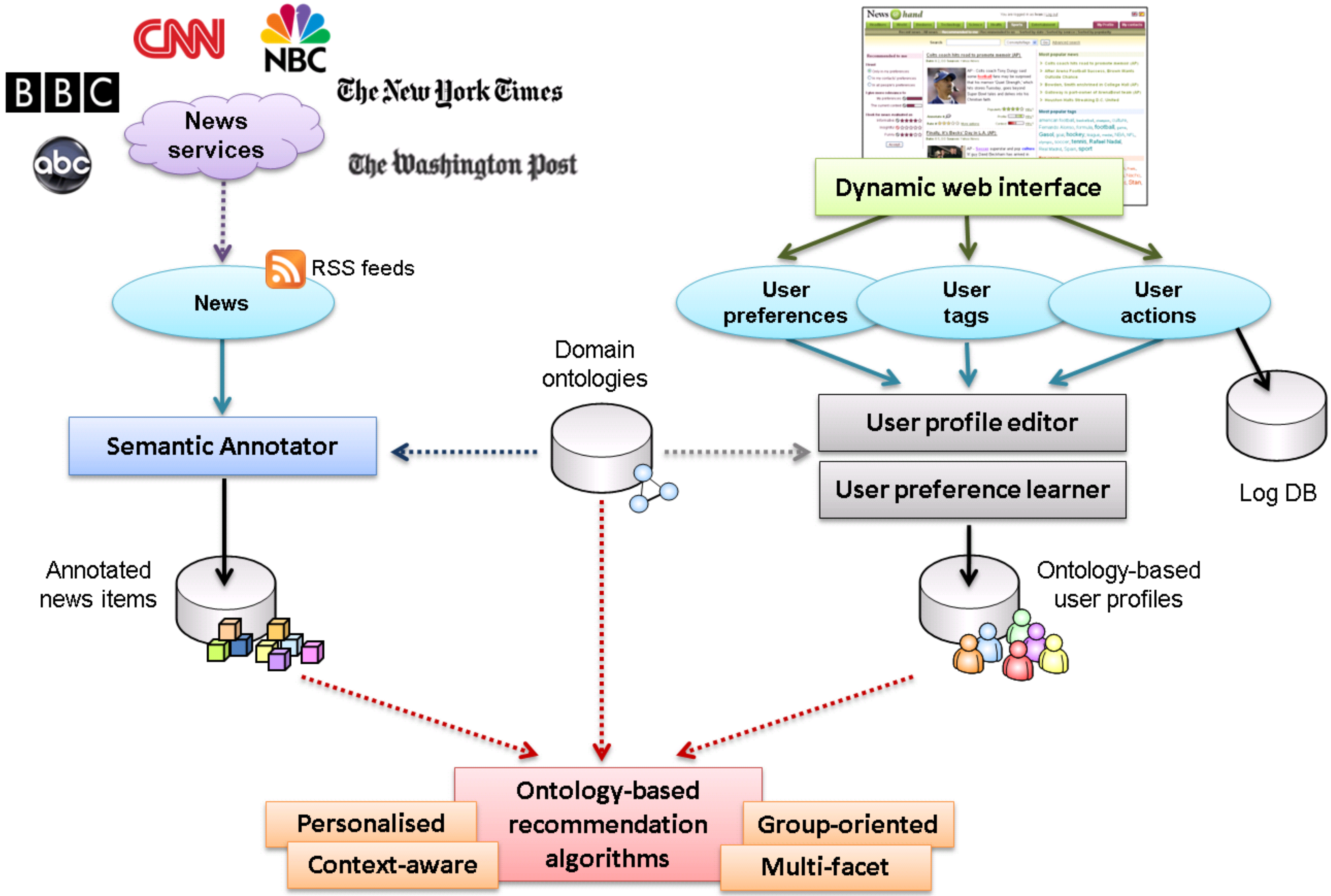
- Log analysis of a personalised news recommender system
- For each user-rated recommendation, a pattern is created
    - The attributes of the pattern correspond to the characteristics we aim to analyse, and their values are obtained from log information databases.
    - The class of the pattern can be assigned two possible values, correct or incorrect, depending on whether the user evaluated the recommendation as relevant or irrelevant

Recommendation Algorithms

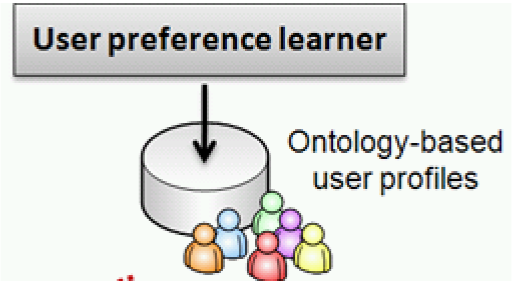


Logs → Patterns $\xrightarrow{\text{ML}}$ Relevant preferences for effective recommendations

Architecture



User profile representation



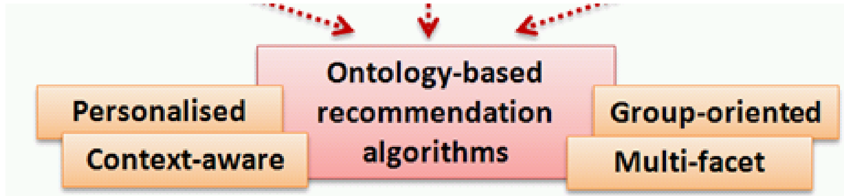
- User preferences are described as vectors

  $\mathbf{u}_m = (u_{m,1}, u_{m,2}, \ldots, u_{m,K})$ where $u_{m,k} \in [-1,1]$ measures the intensity of the interest of user $u_m \in \mathcal{U}$ for concept $c_k \in \mathcal{O}$ (a class or an instance) in a domain ontology $\mathcal{O}$

- Items $d_n \in \mathcal{D}$ are assumed to be annotated by vectors

  $\mathbf{d_n} = (d_{n,1}, d_{n,2}, \ldots, d_{n,K})$ of concept weights, in the same vector-space as user preferences

Recommendation models

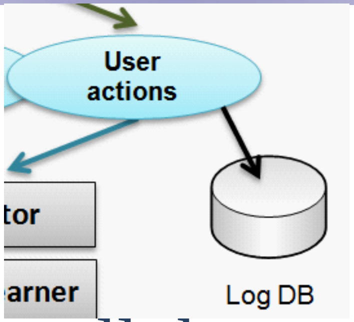


- Preference expansion mechanism
  - Through explicit semantic relations with other concepts in the ontology
- Personalised and context-aware

  $$score\ (d_n, u_m) = \mathrm{w_p} \cdot pref\ (d_n, u_m) + \mathrm{w_c} \cdot pref\ (d_n, u_m, context)$$

  - *pref(·,·)* measures the relevance of a document for a user, using a cosine-based vector similarity
  - *context* is represented as a set of weighted concepts

Log database



- The system monitors all the actions the user performs, and records them in a log database

Table	Attributes
Browsing	**actionID**, actionType, timestamp, **sessionID**, itemID, **itemRankingPosition**, itemRankingProfile, itemRankingContext, itemRankingCollaborative, itemRankingHybridUP, itemRankingHybridNUP, itemRankingHybridUPq, itemRankingHybridNUPq, topicSection, interestSituation, **userProfileWeight**, **contextWeight**, collaborative, scoreSearch
Context updates	actionID, actionType, timestamp, sessionID, context, origin, changeOfFocus
Queries	actionID, actionType, timestamp, sessionID, keywords, topicSection, interestSituation
Recommendations	actionID, actionType, timestamp, sessionID, recommendationType, userProfileWeight, contextWeight, collaborative, topicSection, interestSituation
User accesses	actionID, actionType, timestamp, sessionID
User evaluations	**actionID**, actionType, timestamp, **sessionID**, itemID, rating, **userFeedback**, tags, comments, topicSection, interestSituation, duration
User preferences	actionID, actionType, timestamp, sessionID, concept, weight, interestSituation
User profiles	actionID, actionType, timestamp, sessionID, userProfile
User sessions	sessionID, userID, timestamp

The News@hand system

- A **hybrid news recommender system** which makes use of Semantic Web technologies to provide several on-line news recommendation services

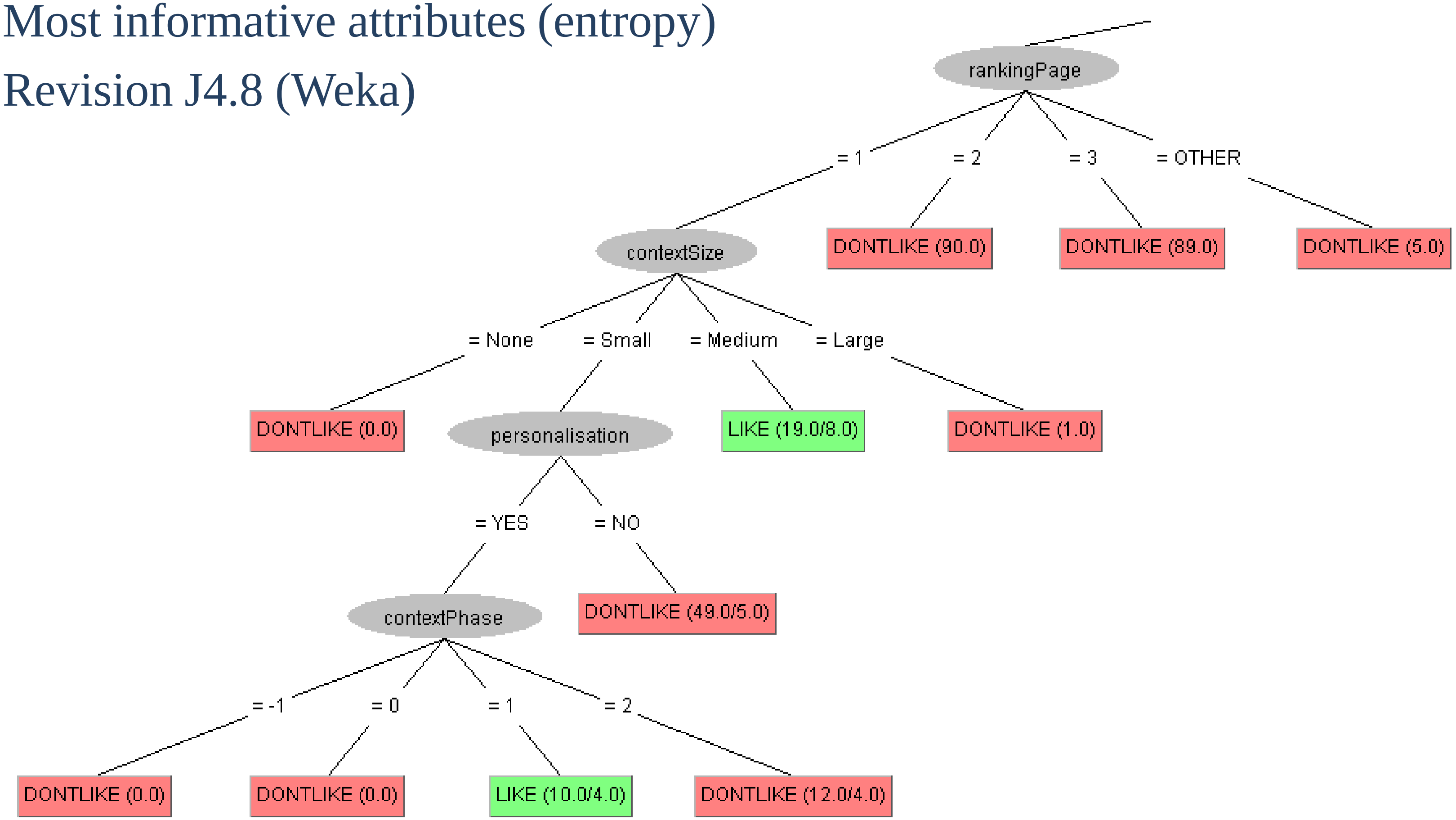
Long-term (profile) user preferences

Short-term (context) user preferences

Semantic expansion

Machine learning algorithms

- Decision Trees:
  - Interpretable
  - Most informative attributes (entropy)
  - Revision J4.8 (Weka)



Evaluation

- 16 users (12 undergraduate/graduate students, 4 lecturers)
- Task: find and evaluate those news items that were relevant to a given goal

Profile	Section	Query		Task goal
1 Telecom	World	$Q_{1,1}$	pakistan	News about media: TV, radio, Internet
	Entertainment	$Q_{1,2}$	music	News about software piracy, illegal downloads, file sharing
2 Banking	Business	$Q_{2,1}$	dollar	News about oil prices
	Headlines	$Q_{2,2}$	fraud	News about money losses
3 Social care	Science	$Q_{3,1}$	food	News about cloning
	Headlines	$Q_{3,2}$	internet	News about children, young people, child safety, child abuse

- Different configurations based on activation/deactivation of: personalisation and context-aware recommendations, semantic expansion
- Each user classifies an item as relevant in general, relevant to the current goal or relevant to the profile

Results

- Relevant preferences for personalised recommendations:
  - The bigger the user profile size, the more relevant the retrieved news
  - The system retrieves relevant news in the first page (top 5)
  - If the expansion is activated it is a very important preference
- Relevant preferences for context-aware recommendations:
  - Best performance with a medium context size
  - Context usually needs personalisation to obtain good results
- Meta-evaluation conclusions:
  - The evaluation was unbalanced in terms of difficulty of obtaining relevant news items for each task
  - Profile specificity (*Business*, *Entertainment* sections; manual profiles)

Conclusions and future work

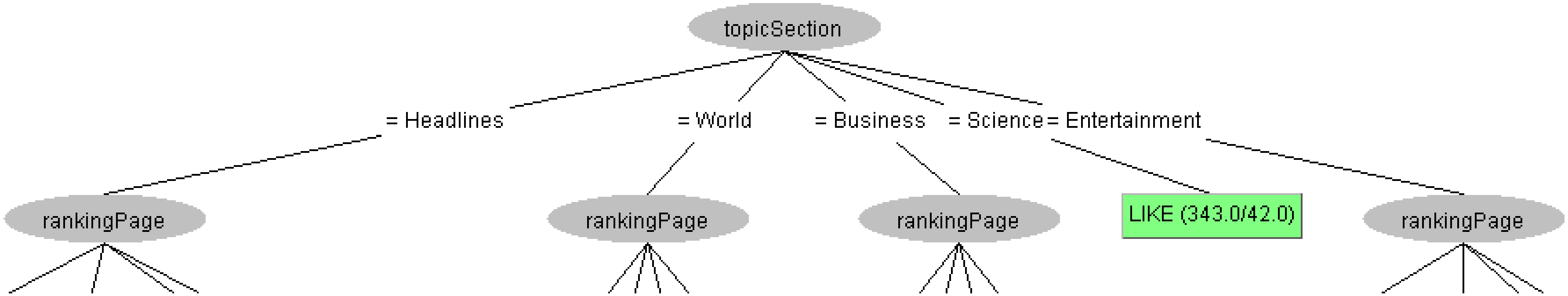
- Machine Learning techniques are useful:
  - to learn those user and system features of a recommender system that are (un)favourable for correct recommendations
  - to learn deficiencies and weaknesses of the experiments conducted to measure the system performance

- Next steps:
  - Make the system adaptive to the current status of the analysed preferences and evaluate the (potential) improvement
  - Similar evaluations with the collaborative group-oriented and multi-facet recommendation strategies of News@Hand

Questions

Thank you

Split showing unbalanced anomaly

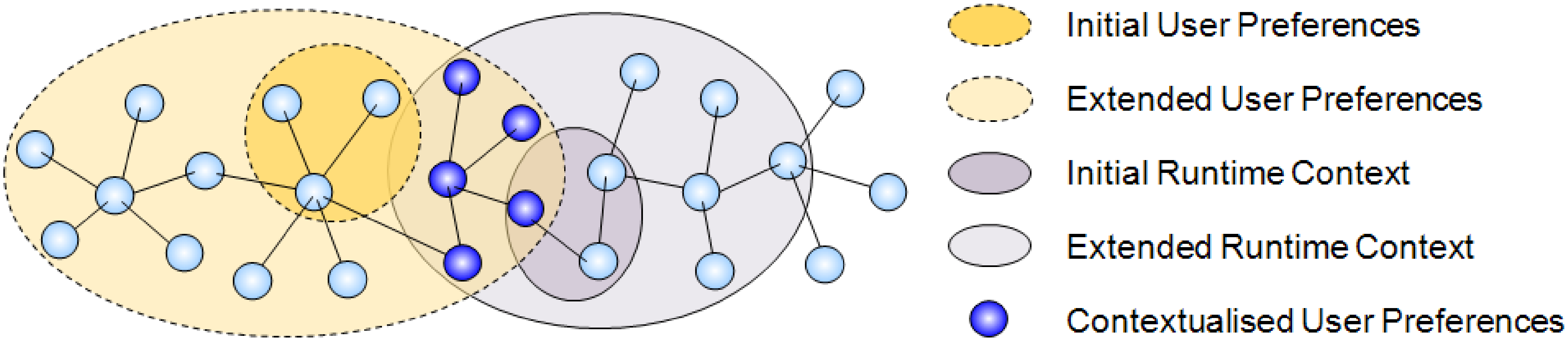


Evaluation: configurations

User	Personalised recommendations		Context-aware recommendations	
	Without expansion	With expansion	With expansion	
	w_p=1 w_c=0	w_p=1 w_c=0	w_p=0 w_c=1	w_p=0.5 w_c=1
1	$*Q_{1,1}$	$Q_{2,1}$	$Q_{3,1}$	$^{A}Q_{1,2}$
2	$Q_{2,2}$	$*Q_{3,2}$	$^{A}Q_{2,1}$	$Q_{1,2}$
3	$Q_{3,1}$	$^{A}Q_{3,2}$	$*Q_{1,1}$	$Q_{2,1}$
4	$^{A}Q_{1,1}$	$Q_{1,2}$	$Q_{2,2}$	$*Q_{3,2}$
5	$Q_{1,2}$	$*Q_{2,2}$	$Q_{3,2}$	$^{A}Q_{2,1}$
6	$Q_{2,1}$	$Q_{3,1}$	$*^{A}Q_{3,2}$	$Q_{1,1}$
7	$Q_{3,2}$	$^{A}Q_{1,1}$	$Q_{1,2}$	$*Q_{2,2}$
8	$*^{A}Q_{2,2}$	$Q_{1,1}$	$Q_{2,1}$	$Q_{3,1}$
9	$Q_{1,1}$	$Q_{2,1}$	$*Q_{3,1}$	$^{A}Q_{3,2}$
10	$Q_{2,2}$	$Q_{3,2}$	$^{A}Q_{1,1}$	$*Q_{1,2}$
11	$*Q_{3,1}$	$^{A}Q_{2,2}$	$Q_{1,1}$	$Q_{2,1}$
12	$^{A}Q_{3,1}$	$*Q_{1,2}$	$Q_{2,2}$	$Q_{3,2}$
13	$Q_{1,2}$	$Q_{2,2}$	$Q_{3,2}$	$*^{A}Q_{1,1}$
14	$*Q_{2,1}$	$Q_{3,1}$	$^{A}Q_{2,2}$	$Q_{1,1}$
15	$Q_{3,2}$	$*^{A}Q_{3,1}$	$Q_{1,2}$	$Q_{2,2}$
16	$^{A}Q_{1,2}$	$Q_{1,1}$	$*Q_{2,1}$	$Q_{3,1}$

Semantic contextualisation of user preferences

- Nodes represent ontology concepts and edges are associated to semantic relations between those concepts.



Knowledge Representation

- **User profiles** and **item descriptions** are represented as vectors $\mathbf{u_m}=(u_{m,1}, u_{m,2}, \ldots, u_{m,K})$ and $\mathbf{d_n}=(d_{n,1}, d_{n,2}, \ldots, d_{n,K})$, where $u_{m,k}$, $d_{n,k}$ in [-1,1] are the weights that measure the relevance of concept c_k for user u_m and item d_n
- **Recommendation models** are based on the definition of matching algorithms which make use of similarity measures based on $cos(\mathbf{u_m}, \mathbf{d_n})$