Integrating Quality Criteria in a Fuzzy Linguistic Recommender System for Digital Libraries

A. Tejeda-Lorente, J. Bernabé-Moreno, C. Porcel, E. Herrera-Viedma

ITQM 2014
3-6 JUNE 2014, Moscow
Contents

• Introduction.
• Proposed system.
• Conclusions.
Introduction

- Web: Main source of information generation and transmission.
- We focus on an academic environment: **University Digital Libraries (UDL)**.

**Information Access Problems**

- Need for automatic search systems and access to the information in the Web:
  - **Recommender Systems (RecSys)**: They aid users in the information access process through prediction and item recommendation that can be interesting for them → users’ profile.
Introduction

• **Main problem** in the Web: exponential and uncontrolled:

• **Consequence:** the users of UDL still having serious difficulties to access to relevant information.
Proposed solution

- We split the process of generating recommendations in two phases:
  1. Identify relevant resources.
  2. Identify valid resources from a quality point of view.

- Hybrid recommendation $\rightarrow$ Switched hybrid RecSys: To alternate between a content-based scheme and a collaborative one depending on the number of existing ratings.

- To add the Re-ranking module which combines the estimated relevance degree with the quality of the item.

- To adopt a multi-granular fuzzy linguistic modeling.
Contents

• Introduction.
• Proposed system.
• Conclusions.
Description of the proposed system
Proposed system : representation of Information

• We use different sets of labels selected from a linguistic hierarchy.
• Concepts assessed:
  1. **Relevance degree** of a discipline with respect to a resource scope, which is assessed in $S_1$.
  2. **Similarity degree** among resources or among users, which is assessed in $S_2$.
  3. **Predicted relevance degree** of a resource for a user, which is assessed in $S_3$.
  4. **Satisfaction degree** expressed by a user to evaluate a recommended resource, which is assessed in $S_4$.
  5. **Preference degree** of a resource regarding another one, which is assessed in $S_5$.
• We use 5 labels to $S_1 \sim S_5$, and 9 to $S_2, S_3 \sim S_4$. 
Proposed system: representation of Information

Resources representation

• To represent the resource scope, we use a vector model.

• We use a classification of by 25 disciplines.

• A resource $i$, is represented as:

$$VR_i=(VR_{i1}, VR_{i2}, ..., VR_{i25})$$

• where $VR_{ij}$ ($S_1$ labels) shows the importance degree of discipline $j$ regarding to resource scope $i$. 

ITQM 2014: Integrating Quality Criteria in a Fuzzy Linguistic RecSys for Digital Libraries
User profiles

1. To acquire users’ preferences over the 5 most representative resources.
   - It is enough for users to provide a row of the relation and the system will complete the relation ($S_5$ Labels).

2. To calculate user resource preference degrees over each considered resource $\rightarrow$ arithmetic mean.
   - Now we can obtain the user preference vector as the aggregation of vectors representing selected resources characteristics, weighted through preference degrees.
Proposed system : Recommendation scheme

Hybrid scheme

• It allows us to face the cold start problem.
• Similarity measures: standard cosine measure, but defined in a linguistic context ($S_2$ labels).

• Content-based approach: when a new resource is inserted.
• Collaborative approach: when a new user is inserted.

• Then, the relevance of a resource for a user is estimated ($S_3$ labels).
Proposed system: Quality estimation

- Idea: If a resource is usually preferred over others that show a certain quality.

- At the stage of completing the incomplete preference relations we count the number of times a resource $i$ is chosen to be shown among the outstanding resources, $(s_i)$ is the total of times the resource $i$ has been selected and the total number of times $i$ has been preferred over other $(p_i)$:

\[ q(i) = \frac{p_i}{s_i} \]

- Advantages: It avoids to collect additional information about users and to increase the complexity.
Proposed system : Reranking

- We aggregate the estimated relevance with the quality score obtained.
- We use a **multiplicative aggregation** and we normalize it in the range of the label set $S_3$.

- **Advantages**: ease of application and good results obtained.
Proposed system: Feedback

• The activity of generating recommendations is completed with this phase.

• Users provide the system with their satisfaction ratings about the items received ($S_4$ labels).
Contents

• Introduction.
• Proposed system.
• Conclusions.
Conclusions

• We have addressed the recommendations process from two perspectives:
  1. Find relevant resources.
  2. Resources of good quality.
• We have presented a hybrid fuzzy linguistic recommender system applied to a UDL.
• We performed online studies → satisfactory results.

• Future works:
  – Techniques for automatic resource representation.
  – Incorporate new techniques in the recommendation process.
Any question?