Semantic Coherence-based User Profile Modeling in the Recommender Systems Context

Roberto Saia Ludovico Boratto Salvatore Carta

Department of Mathematics and Computer Science
University of Cagliari, Italy

KDIR 2014 - 6th *International Conference on Knowledge Discovery and Information Retrieval*
October 21-24, 2014 Rome, Italy

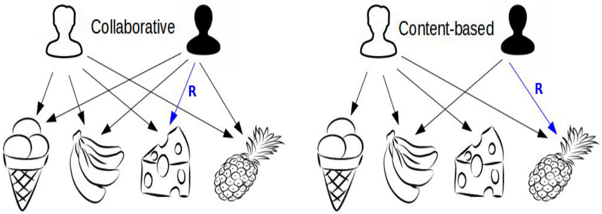
Outline

RECOMMENDER SYSTEMS

- A Recommender System (**RS**) produces personalized content for users in the form of suggestion of items
- It can operate in different ways to obtain this result (e.g., *Collaborative Filtering*, *Content-based Filtering* or *Hybrid Approaches*)

RECOMMENDER SYSTEMS

- A Recommender System (**RS**) produces personalized content for users in the form of suggestion of items
- It can operate in different ways to obtain this result (e.g., *Collaborative Filtering*, *Content-based Filtering* or *Hybrid Approaches*)

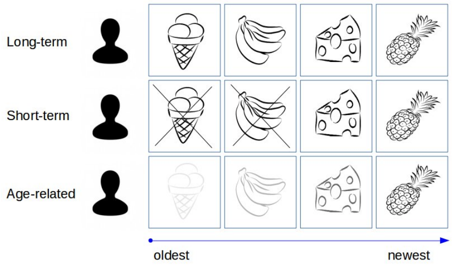


USER PROFILING

- There are different ways to define the user profiles
- Usually they take into account the age of information in different ways (e.g. *long-term*, *short-term* or *age-related* approaches)

USER PROFILING

- There are different ways to define the user profiles
- Usually they take into account the age of information in different ways (e.g. *long-term*, *short-term* or *age-related* approaches)



PROBLEM DEFINITION

- Regardless of the ways to use the information stored in the user profiles, there are several factors that could make the information non-adherent to real tastes of the users
- e.g., this could happen when a personal account is used by other people

PROBLEM DEFINITION

- Regardless of the ways to use the information stored in the user profiles, there are several factors that could make the information non-adherent to real tastes of the users
- e.g., this could happen when a personal account is used by other people

OBTAIN RELIABLE USER PROFILES

- We would like to define a strategy able to make the user profiles as adherent as possible to the real user tastes
- Then we need a way to identify and remove this kind of noise from the user profiles

OBTAIN RELIABLE USER PROFILES

- We would like to define a strategy able to make the user profiles as adherent as possible to the real user tastes
- Then we need a way to identify and remove this kind of noise from the user profiles

A GENERAL PURPOSE APPROACH

- There are several approaches designed to face this problem, but each of them usually is designed to operate within a precise context (e.g., news or goods)
- What we want to achieve is instead a *general purpose* approach able to operate in almost all *text-based* contexts

A GENERAL PURPOSE APPROACH

- There are several approaches designed to face this problem, but each of them usually is designed to operate within a precise context (e.g., news or goods)
- What we want to achieve is instead a *general purpose* approach able to operate in almost all *text-based* contexts

IMPLEMENTATION

- We use a *coherence-based* model to identify the *noise* in a user profile
- The idea is to measure the semantic similarity of each item with respect to remaining items (*Global Coherence*)
- The semantic similarity measures are performed exploiting WordNet (conceptual relations between synsets[1])

[1] sets of cognitive synonyms

IMPLEMENTATION

- We use a *coherence-based* model to identify the *noise* in a user profile
- The idea is to measure the semantic similarity of each item with respect to remaining items (*Global Coherence*)
- The semantic similarity measures are performed exploiting WordNet (conceptual relations between synsets[1])

[1] sets of cognitive synonyms

IMPLEMENTATION

- We use a *coherence-based* model to identify the *noise* in a user profile
- The idea is to measure the semantic similarity of each item with respect to remaining items (*Global Coherence*)
- The semantic similarity measures are performed exploiting WordNet (conceptual relations between synsets[1])

[1] sets of cognitive synonyms

DETECTION AND NOISE REMOVAL

- We remove an item **i** from the user profile when it meets two requirements:
  1. it is located in the oldest part **o** of the user interaction history;
  2. its semantic similarity $\mathbf{S_i}$ with the set of other items is below the mean value of the semantic similarity of all items $\overline{\mathbf{GS}}$ stored in the profile.
- Regarding both requirements, we can introduce an arbitrary tolerance value (α_1 and α_2)

$$Remove(i) = \begin{cases} yes, & \text{if } S_i < (\overline{GS} \mp \alpha_1) \text{ AND } Time(i) <= (o \mp \alpha_2) \\ no, & \text{otherwise} \end{cases}$$

DETECTION AND NOISE REMOVAL

- We remove an item **i** from the user profile when it meets two requirements:
  1. it is located in the oldest part **o** of the user interaction history;
  2. its semantic similarity $\mathbf{S_i}$ with the set of other items is below the mean value of the semantic similarity of all items $\overline{\mathbf{GS}}$ stored in the profile.
- Regarding both requirements, we can introduce an arbitrary tolerance value (α_1 and α_2)

$$Remove(i) = \begin{cases} yes, & \text{if } S_i < (\overline{GS} \mp \alpha_1) \text{ AND } Time(i) <= (o \mp \alpha_2) \\ no, & \text{otherwise} \end{cases}$$

DETECTION AND NOISE REMOVAL

- We remove an item **i** from the user profile when it meets two requirements:
  1. it is located in the oldest part **o** of the user interaction history;
  2. its semantic similarity $\mathbf{S_i}$ with the set of other items is below the mean value of the semantic similarity of all items $\overline{\mathbf{GS}}$ stored in the profile.
- Regarding both requirements, we can introduce an arbitrary tolerance value (α_1 and α_2)

$$Remove(i) = \begin{cases} yes, & \text{if } S_i < (\overline{GS} \mp \alpha_1) \text{ } AND \text{ } Time(i) <= (o \mp \alpha_2) \\ no, & \text{otherwise} \end{cases}$$

DETECTION AND NOISE REMOVAL

- We remove an item **i** from the user profile when it meets two requirements:
  1. it is located in the oldest part **o** of the user interaction history;
  2. its semantic similarity $\mathbf{S_i}$ with the set of other items is below the mean value of the semantic similarity of all items $\overline{\mathbf{GS}}$ stored in the profile.
- Regarding both requirements, we can introduce an arbitrary tolerance value (α_1 and α_2)

$$Remove(i) = \begin{cases} yes, & \text{if } S_i < (\overline{GS} \mp \alpha_1) \text{ } AND \text{ } Time(i) <= (o \mp \alpha_2) \\ no, & \text{otherwise} \end{cases}$$

EXPERIMENTS

- In the performed experiments[2] we have compared the F_1-measure (a metric that combines *Precision* and *Recall* metrics) of a Recommender System at the state of the art, before and after the noise removal
- Aim: show the effectiveness of our approach to remove incoherent items from the user profiles, to improve the accuracy of the recommendations

[2] Using *Yahoo! Webscope Movie Dataset R4*

EXPERIMENTS

- In the performed experiments[2] we have compared the F_1-measure (a metric that combines *Precision* and *Recall* metrics) of a Recommender System at the state of the art, before and after the noise removal
- Aim: show the effectiveness of our approach to remove incoherent items from the user profiles, to improve the accuracy of the recommendations

[2] Using *Yahoo! Webscope Movie Dataset R4*

MAIN RESULTS

- After the noise removal, the percentage increment of the F_1-measure has been around the 13%
- A coherent profile also shows advantages in terms of computational load reduction

- Increase in percent of the F_1-measure:

MAIN RESULTS

- After the noise removal, the percentage increment of the F_1-measure has been around the 13%
- A coherent profile also shows advantages in terms of computational load reduction

- Increase in percent of the F_1-measure:

MAIN RESULTS

- After the noise removal, the percentage increment of the F_1-measure has been around the 13%
- A coherent profile also shows advantages in terms of computational load reduction

- Increase in percent of the F_1-measure:

CONCLUSIONS

- We wanted to define a *general purpose* approach able to locate and remove from the user profiles the items that could make them *non-adherent* to the real tastes of users
- We have done this using a new *coherence-based* metric, obtaining an improvement or in term of accuracy or in term of computational load

CONCLUSIONS

- We wanted to define a *general purpose* approach able to locate and remove from the user profiles the items that could make them *non-adherent* to the real tastes of users
- We have done this using a new *coherence-based* metric, obtaining an improvement or in term of accuracy or in term of computational load

FUTURE WORKS

- We would like to define a criteria to evaluate the *coherence* within a heterogeneous environment, investigating the semantic relation between the different kinds of items
- and evaluate with other experiments the improvement in terms of computational load, a side effect observed during the experiments that requires to be well formalized

FUTURE WORKS

- We would like to define a criteria to evaluate the *coherence* within a heterogeneous environment, investigating the semantic relation between the different kinds of items
- and evaluate with other experiments the improvement in terms of computational load, a side effect observed during the experiments that requires to be well formalized

Questions and Answers

THANK YOU