UNIVERSITÉ DE TUNIS EL MANAR (LIPAH)
UNIVERSITÉ BLAISE PASCAL CLERMONT FERRAND II (LIMOS)

Towards more targeted recommendations in folksonomies

Mohamed Nader Jelassi
Supervised by:
Engelbert Mephu Nguifo
Sadok Ben Yahia

27 mars 2014

Plan

Plan

Plan

Plan

Plan

Plan

1. Introduction and Motivations

2. Related Works

3. PersoRec : the personalized recommender system

4. Results and Discussion

5. Conclusion and Perspectives

Folksonomies

Three-dimensional context

- The user participates via keywords called **Tags**
- With Tags, the user shares and annotates **Resources** (books, movies, bookmarks, pictures, etc.)
- The sets of **Users**, **Tags** and **Resources** constitute a structure called **Folksonomy**

3 Main Actors

USERS   **TAGS**   **RESOURCES**

Users : Kings of the World !

Focus on Users

- Main Actors : contribute to creation of the information
- Everyone can do it !
- So easy to share, annotate
- Tags/Resources exclusively added and shared by users
- Personal choice : depends of the user's **profile** (its age, occupation, location, etc.)

Profile : 4th actor of the Folksonomy

Four-dimensional context

P-Folksonomy = (Users, Tags, Resources, **Profile**)

Different Users = Different Profiles

What Issues ?

Issues

- Information overload
- Lack of data organisation : How to choose ?
- What resources to share ?
- What tags to use for annotating ?

Solution : Personalized Recommandations

Solution

- Folksonomies have to reflect the profile of each user during the recommendation process
- Recommender systems need a deeper understanding of users
  $\Rightarrow$ Personalizing the Recommendations
- Helping each user to :
  - Share the « good » resources : the interesting ones
  - Use the « good » tags : the adequate ones

Users' Profile : How it can help ?

Profile

In the process of the personalization of recommendations, it is helpful to have further informations about users :

age, gender, occupation, country, etc.

(Tags, Resources)

Plan

1. Introduction and Motivations

2. Related Works

3. PersoRec : the personalized recommender system

4. Results and Discussion

5. Conclusion and Perspectives

Three Main Categories

Three Main Categories

1. Works using similarity measures
2. Works relying on popularity
3. Hybrid Works using history tagging and similarity measures

Main Works

Main Works

- (Diederich *et al.*, 2006)
    - The Recommandation relies on a similarity measure between users
    - Two users are similar if they share similar tags
- (Lipczak, 2008)
    - Recommandation of tags by co-occurrence on other resources

Main Works

- (Landia *et al.*, 2009)
    - Recommandation of tags
    - Relying on a similarity measure between users
- (Hu *et al.*, 2011)
    - Recommandation combining both history tagging and social contacts

The surveyed approaches at a glance

	Multi-Mode	Coverage	Cold Start	Diversity
Diedrich *et al.*	No			✓
Landia	No			
Jaeschke *et al.*	No			✓
Lipzack	No			✓
De Meo *et al.*	No			?
Hu *et al.*	No			✓
Basile *et al.*	No			?
Our Approach	Yes	✓	✓	✓

The surveyed approaches at a glance

	Serendipity	Adaptivity	Scalability
Diedrich *et al.*	?		
Landia	?	✓	?
Jaeschke *et al.*	?		✓
Lipzack	?	✓	?
De Meo *et al.*	?	✓	?
Hu *et al.*	?	✓	?
Basile *et al.*	?	✓	?
Our Approach	✓	✓	✓

Contributions

Contributions

- Combining tagging history with users' profile
- Focus on tags and resources the most used in combination (quadri-concepts)
- A more specific result for our two-steps personalized recommender system

Plan

1. Introduction and Motivations

2. Related Works

3. PersoRec : the personalized recommender system

4. Results and Discussion

5. Conclusion and Perspectives

Two-steps personalized recommender system

Two-steps personalized recommender system

1. **Quadri-Concepts Extraction :** Bridging under quadratic structures the users, tags and resources the most used in combination(QuadriCons, Data Peeler)

2. **Personalized Recommandation :** An algorithm called PersoRec

Step 1 : Quadri-concepts

What is a quadri-concept ?

Quadri-concept is a quadratic structure (U,T,R,P)

1. U : the set of users
2. T : the set of tags
3. R : the set of resources
4. P : the set of profiles

- **We read :** Each user of U (with a profile P) has tagged each resource of R with all tags of T.
- The Quadri-Concept is **maximal** : none of these sets can be extended without shrinking one of the other three dimensions.
- Quadri-concepts are a small representation of a folksonomy which may contains thousands of quadruples in real-life datasets.

Step 1 : Quadri-concepts

Example of a quadri-concept

{john,peter,dana},
{papers,books,reviews},
{sciencedirect.com, springer.com},
{18-25 years old,student}
We read : **John**, **Peter** and **Dana**, three **students** aged between **18 and 25 years old** have shared the websites **sciencedirect.com**, **springer.com** via the tags **papers**, **books**, **reviews**.

Step 2 : PersoRec

Input

- $\mathcal{QC}$: A set of quadri-concepts
- (u,p) : the target user and its profile

Output

1. A set of proposed friends
2. A set of suggested tags
3. A set of recommended resources

Remark

- The step of quadri-concepts extraction is an **off-line** phase which is performed only **once**.
- The recommender system does not depend of the extraction of quadri-concepts at each recommendation

Plan

1. Introduction and Motivations

2. Related Works

3. PersoRec : the personalized recommender system

4. Results and Discussion

5. Conclusion and Perspectives

Datasets and Plan

Datasets

1. The MovieLens filmography dataset : 95580 tags applied to 10681 movies by 71567 users
2. The BookCrossing library dataset : 278858 users providing 1149780 ratings about 271379 books
   - Profiles : Gender, Age, Occupation, Country

Plan of our evaluation

1. Quality metrics : Precision, Recall and F1-Score
2. Social Evaluation
3. Recommender Systems Properties
4. User Study

Some Examples (MovieLens)

Users	Tags	Resources	Profile
{bernadette, bridget, margaret62}	{classic, dialog, oscar}	{Star Wars, M.A.S.H, Rear Window}	Female, 46-73 years, retired
{mulder, scully, csmdavis	{bestmovie, cult}	{Usual Suspects, Silence of the Lambs, Sound of Music}	Male, 25-35 years, healthcare
{ross, anlucia, franela	{classic, oldmovie, quotes}	{Rear Window, Magician of OZ, Gone with the Wind}	Male, 36-45 years, Writer

Precision, Recall and F1-Score
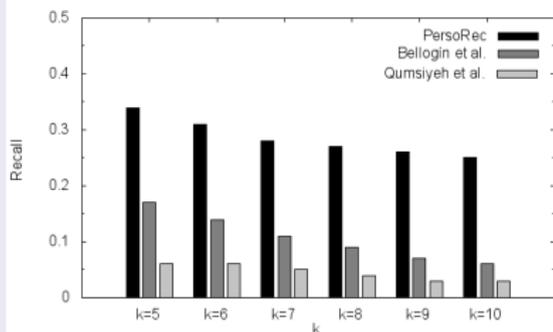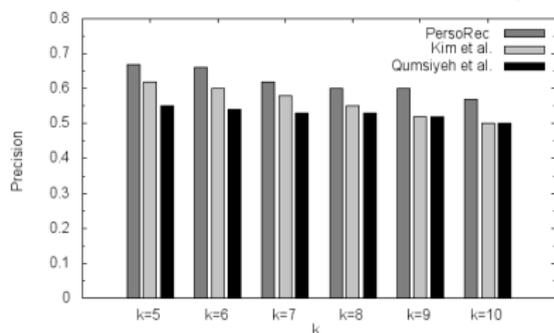
Precision, Recall and F1-Score

Explanations

- The use of quadri-concepts improves the recommendations by suggesting the closest tags and resources to users' needs.
- Related works focus on most used items (books, movies, tags)
- Quadri-concepts offer tags and resources that have been shared in common by a set of users with close profiles.

Social Evaluation

New Idea

We are looking at what happens **after** the recommendation step, *i.e.*, if the target user really enjoys the recommendation and if the users (friends) proposed to him/her are getting the same behavior.

Test Users

1. BookCrossing : *skinner* (38 years, New York, USA)
2. MovieLens : *Bruce* (47 years, Male, Educator)

Social Evaluation (BookCrossing)

Recommendations

- three books from the franchise *Harry Potter*
- four new friends : *snowh* (43 years, *Illinois*, *USA*), *char_dav* (54 years, *California*, *USA*), *emma* (40 years, *Oregon*, *USA*) and *henry90* (36 years, *Teheran*, *Iran*).

Conclusions

- His new friends also share all books of the franchise *Harry Potter*
- *skinner* and his recommended friends, have rated the recommended books with a rating equal to 9
  ⇒ They enjoyed the recommendations

Properties of PersoRec

User Space Coverage

- All users are covered : all users receive recommendations whenever they shared or no.
- Covered Profiles : 100% of genders (male and female), 100% of age categories, 100% of professions and 88% of countries

Cold Start

New Users Recommendations according to their profile

Properties of PersoRec

Serendipity

$$d(b, B) = \frac{1 + C_B - C_B.w(b)}{1 + C_B} \qquad (1)$$

- Serendipity is a measure of how surprising the recommendations are
- d measures the distance between recommended items and known items
- An average score of 44%

Properties of PersoRec

Diversity

$$d(b, B) = \frac{1 + C_B - C_B.w(b)}{1 + C_B} \qquad (2)$$

- Users prefer different recommendations : *e.g.*, 5 books from 5 authors instead of 5 books from the same one
- d measures the distance between recommended items
- An average score of 56%

Properties of PersoRec

Scalability

$minsupp\_u$ (MovieLens)	$|\mathcal{QC}|$	# Unique Users	Task 1 (sec)	Task 2 (sec)
14	795	634	$0,000426$	$0,004947$
12	1295	668	$0,000778$	$0,006169$
10	2430	718	$0,001741$	$0,009348$
8	5123	805	$0,004099$	$0,013556$
6	13461	865	$0,012717$	$0,023303$

- Quadri-Concepts = A small representation of the folksonomy
- # Unique Users = Number of recommendations
- An average response time of $0,002$ seconds for the recommendation of resources and around $0,008$ seconds for the user proposition task.

User Study

Test Subjects

1. (*Nidhal*, Male, 30 years, Academic assistant, Tunisia)
2. (*Imen*, Female, 26 years, Student, Tunisia)
3. (*Roxane*, Female, 27 years, Educator, France)
4. (*Raymond*, Male, 58 years, Retired, Belgium)
5. (*Wassim*, Male, 24 years, Engineer, Canada)
6. (*Quentin*, Male, 28 years, Optician, France)

The Study

1. **Quality of the recommendation** : rating the recommended items
2. **Resource Recommendation** : Selecting resources
3. **Tag Suggestion** : Selecting tags
4. **User Proposition** : Selecting friends

User Study : Some Conclusions

- Users enjoy the serendipity, diversity and quality of recommendations
- Users rated well recommendations (average note of $3, 5$)

- Need of an online track of shared tags and resources in order to more understand users' needs

Plan

1. Introduction and Motivations

2. Related Works

3. PersoRec : the personalized recommender system

4. Results and Discussion

5. Conclusion and Perspectives

Conclusion

Conclusion

- Consideration of the users' profile information
- Combining Quadri-Concepts and users' profile
  $\Rightarrow$ More targeted recommendations in folksonomies
- Good feedback of users

Perspectives

Perspectives

- An online system : an online tracking of users' tags and resources
- An incremental algorithm to take into account new tags and resources

Incrementality

u_1	t_1	r_1	p_1
u_2	t_1	r_1	p_1
u_2	t_1	r_2	p_1
u_3	t_1	r_3	p_1
u_4	t_5	r_1	p_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$

A new quadruple : $(u_3, t_1, r_1\ p_1)$ How it can be "incrusted" ?

Incrementality

First Idea

u_1	t_1	r_1	p_1
u_2	t_1	r_1	p_1
u_3	t_1	r_1	p_1
u_2	t_1	r_2	p_1
u_3	t_1	r_3	p_1
u_4	t_5	r_1	p_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Put the new quadruple directly into the folksonomy
- Keeping the sort of quadruples

- Requires the regeneration of quadri-generators

Incrementality

Second Idea

u_1	t_1	r_1	p_1
u_2	t_1	r_1	p_1
u_2	t_1	r_2	p_1
u_3	t_1	r_3	p_1
u_4	t_5	r_1	p_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_3	t_1	r_1	p_1

- Put the new quadruple directly into the set of already extracted quadri-concepts

u_1, u_2	t_1, t_2	r_1	p_1
u_1, u_3	t_2	r_1	p_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Check the quadri-concept's property
- Which quadri-concept to choose ?
- What if no quadri-concept satisfied the quadruple ?
- Erasing some quadri-concepts afterwards ?

Incrementality

Second Idea

u_1	t_1	r_1	p_1
u_2	t_1	r_1	p_1
u_2	t_1	r_2	p_1
u_3	t_1	r_3	p_1
u_4	t_5	r_1	p_2
⋮	⋮	⋮	⋮
u_3	t_1	r_1	p_1

- Put the new quadruple directly into the set of already extracted quadri-concepts

u_1, u_2, u_3 $\quad$ t_1, t_2(?)	r_1	p_1
u_1(?), u_3 $\quad$ t_1, t_2	r_1	p_1
⋮ $\quad$ ⋮	⋮	⋮

- Check the quadri-concept's property
- Which quadri-concept to choose ?
- What if no quadri-concept satisfied the quadruple ?
- Erasing some quadri-concepts afterwards ?

That's all folks !