



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

Towards more targeted recommendations in folksonomies

Mohamed Nader Jelassi
Engelbert Mephu Nguifo
Sadok Ben Yahia

29 mars 2015

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Folksonomy : What's that ?

Folk = The Users

FOLK



Done by people

Folk = People = Users

Taxonomy = Objects Hierarchy

TAXONOMY



Classifications of items

Classified by users

3 main actors

Triadic Context : Users-Tags-Resources



USERS



TAGS



RESOURCES

Main Issues

Information overload issue

- Lack of data organisation : How to choose?
- Altering recommendations quality : redundant resources, inactive users, rare tags, etc.
- Recommendations which are not pertinent and not personalized

Solution

Context-awareness

- Adding a fourth dimension (time, **user's profile**, etc.) to the folksonomy
- Collecting information about users preferences from their behavior or explicit rating
- Becoming aware of the user's profile in order to improve recommendations
- User's profile = Contextual information

Solution

- Additional information about users : the user's profile
 - a deeper understanding of users and their information seeking tasks
- Quadri-concepts to reduce the huge data
 - grouping under concepts maximal sets of users, tags, resources and profiles

⇒ PersoRec : a new personalized recommender system (Jelassi *et al.*, 2013)

What's new over PersoRec ?

Contributions

Cold Start Considering new users in folksonomies

Coverage and Scalability Improving the coverage of users and items /
Improving the response time

User Study Interaction with the system's users / Feedback of users
towards the system

Ranking score Classify recommendations in order to improve the
precision and recall of our system

Some Formal Notions

A V-Folksonomy

- A **v-folksonomy** is a set of tuples $\mathcal{F}_v = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{V}, Y)$
- \mathcal{U} , \mathcal{T} , \mathcal{R} and \mathcal{V} are finite sets which elements are called, respectively, **users**, **tags**, **resources** and **variables**
- a set of quadruples where $y = \{(u, t, r, v) \mid u \in \mathcal{U}, t \in \mathcal{T}, r \in \mathcal{R}, v \in \mathcal{V}\}$
- the user u has annotated the resource r using the tag t through the variable v (in the following, we choose the profile to model the variable v)

Some Formal Notions

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
- **Definition** : Each user of U (with the profiles of P) has tagged each resource of R with all tags from T .
 - **Property** : The Quadri-Concept is **maximal** : none of these sets can be extended without shrinking one of the other three dimensions.
 - **Why quadri-concepts ?** Quadri-concepts are a small representation of a folksonomy ($|QC| \ll |Y|$).
 - **Frequency** we can define minimum thresholds on each dimension
 - **Dedicated algorithms** QuadriCons (Jelassi et al., 2014), Data Peeler (Cerf et al., 2009).

Some Formal Notions

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**.

Plan

- 1 Introduction and Motivations
- 2 **Related Work**
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Works using similarity measures

Main Works

- (Diederich *et al.*, 2006)
 - The Recommendation relies on a similarity measure between users
 - Two users are similar if they share similar tags
- (Landia *et al.*, 2009)
 - Recommendation of tags
 - Relying on a similarity measure between users

Works relying on popularity

Main Works

- (Jäschke *et al.*, 2006)
 - Recommendation of tags based on the most used ones
- (De Meo *et al.*, 2010)
 - Recommendation of "*authoritative*" tags to enrich users' query

Hybrid Works using history tagging and similarity measures

Main Works

- (Hu *et al.*, 2011)
 - Recommendations based on both user's social contacts and already shared tags/resources
- (Basile *et al.*, 2007)
 - Tag recommendation based on both similarity between resources and already shared tags/resources

Works relying on user information

Main Works

- (Bellogin *et al.*, 2013)
 - Combining some strategies (e.g., content-based, collaborative filtering, and social) with user information to provide more valuable recommendations.
- (Kim *et al.*, 2011)
 - Finding *neighbors* using user's preferences for books and their feature information (i.e., profile) to generate personalized recommendations.
- (Qumsiyeh *et al.*, 2012)
 - A personalized recommendation system that relies on several users information as ratings and reviews of different multimedia items.

Limits and Goals

Limits

- 1 Huge data addressed \Rightarrow may alter recommendation quality
- 2 Most of works are limited to the $\langle \text{user}, \text{tag}, \text{resource} \rangle$ information

Goals

- 1 Use Quadri-concepts to reduce the huge data with a minimal loss of information
 - linking maximal sets of users, tags, resources and profiles
- 2 Use User's profile to personalize the recommendations

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system**
- 4 Results and Discussion
- 5 Conclusion and Perspectives

The FolkRecommender Algorithm

Input

- 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

The FolkRecommender Algorithm

Pseudo Code

u is an old user

- 1 filtering out its shared tags and resources
- 2 proposing friends having the same profile
- 3 suggesting tags not used yet by u but used by users with the same profile on the resource r
- 4 recommending resources not shared yet by u but shared by users having the same profile

u is a new user

- 1 proposing friends having the same profile
- 2 suggesting tags used by users with the same profile on the resource r
- 3 recommending resources shared by users having the same profile

Ranking score

A new ranking score to classify recommendations

$$rec_score(r_i, v) = \frac{|u_i|}{|UU|} / \exists t_i \exists r_i \exists v_i \in v, (u_i, t_i, r_i, v_i) \in \mathcal{F}_v$$

The FolkRecommender Algorithm

Challenges

- Cold Start and user space coverage : we consider new users which have shared anything yet
- Diversity and Novelty of recommendations : we filter out tags and resources already shared by each user

Plan

- 1 Introduction and Motivations
- 2 Related Work
- 3 A new personalized recommender system
- 4 Results and Discussion**
- 5 Conclusion and Perspectives

Datasets and Plan

Datasets

- The MovieLens filmography dataset : 95580 tags applied to 10681 movies by 71567 users
- (user, **tag**, **movie**, profile)

| User | Tag | Resource | Profile |
|----------------|-----------------------|----------------------|---------------------|
| <i>Mulder</i> | <i>action</i> | <i>X-Files</i> | <i>student</i> |
| <i>Mulder</i> | <i>sciencefiction</i> | <i>X-Files</i> | <i>25 years old</i> |
| <i>Scully</i> | <i>adventure</i> | <i>Jurassic Park</i> | <i>professor</i> |
| <i>Scully</i> | <i>bestmovie</i> | <i>Jurassic Park</i> | <i>female</i> |
| <i>Skinner</i> | <i>thriller</i> | <i>Carrie</i> | <i>Canada</i> |
| ⋮ | ⋮ | ⋮ | ⋮ |

Table : A screenshot of the MovieLens dataset.

Datasets and Plan

Datasets

- 1 The BookCrossing library dataset : 278858 users providing 1149780 ratings about 271379 books
- 2 (user, **rating**, **book**, profile)

| User | Rating | Resource | Profile |
|---------|--------|---------------------------------|--------------|
| Malcolm | 10 | Mohammed : The Prophet of Islam | student |
| Kate | 8 | I got you under my skin | actress |
| Kate | 8 | I got you under my skin | 38 years old |
| Columbo | 9 | Mouth of Madness | 26 years old |
| Columbo | 9 | Mouth of Madness | Spain |
| ⋮ | ⋮ | ⋮ | ⋮ |

Table : A screenshot of the BookCrossing dataset.

Some Examples of quadri-concepts (MovieLens)

| Users | Tags | Resources | Profile |
|--------------------------------|-----------------------------------|---|--|
| {saloua, yasmine, wafa} | {classic, dialog, oscar} | {Star Wars, Magician of OZ, Rear Window} | {Female, 46-73 years, retired} |
| {mulder, scully, krycek} | {bestmovie, cult} | {Usual Suspects, Silence of the Lambs, X-Files} | {Male , 25-35 years, healthcare} |
| {ross, anlucia, franela} | {classic, oldmovie, quotes} | {Rear Window, Magician of OZ, Gone with the Wind} | {Man, 36-45 years, Writer} |

Validation protocol

- 5-fold cross-validation : Both MovieLens and BookCrossing datasets were split into two-sub datasets : the first sub dataset, containing random 80% of users, was used as **training set** while the second one, containing the remaining users (*i.e.*, random 20% of users), was retained as the validation data for tests (*i.e.*, the **test set**).

Precision

MovieLens

| k | FolkRecommender | Bellogin <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|------------------------|------------------------|----------|
| 6 | 0, 76 | 0, 40 | 0, 27 | 0, 39 |
| 7 | 0, 73 | 0, 37 | 0, 27 | 0, 35 |
| 8 | 0, 65 | 0, 35 | 0, 25 | 0, 36 |
| 9 | 0, 63 | 0, 33 | 0, 24 | 0, 37 |
| 10 | 0, 62 | 0, 32 | 0, 23 | 0, 35 |

BookCrossing

| k | FolkRecommender | Kim <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|-------------------|------------------------|--------------|
| 6 | 0, 72 | 0, 58 | 0, 55 | 0, 66 |
| 7 | 0, 63 | 0, 54 | 0, 50 | 0, 60 |
| 8 | 0, 56 | 0, 52 | 0, 47 | 0, 56 |
| 9 | 0, 51 | 0, 49 | 0, 45 | 0, 50 |
| 10 | 0, 47 | 0, 47 | 0, 42 | 0, 46 |

Recall

MovieLens

| k | FolkRecommender | Bellogin <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|------------------------|------------------------|----------|
| 6 | 0, 54 | 0, 20 | 0, 09 | 0, 32 |
| 7 | 0, 51 | 0, 16 | 0, 08 | 0, 29 |
| 8 | 0, 48 | 0, 14 | 0, 07 | 0, 28 |
| 9 | 0, 46 | 0, 11 | 0, 06 | 0, 27 |
| 10 | 0, 40 | 0, 10 | 0, 06 | 0, 26 |

BookCrossing

| k | FolkRecommender | Kim <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|-------------------|------------------------|----------|
| 6 | 0, 34 | 0, 12 | 0, 09 | 0, 30 |
| 7 | 0, 34 | 0, 11 | 0, 08 | 0, 28 |
| 8 | 0, 34 | 0, 11 | 0, 07 | 0, 26 |
| 9 | 0, 34 | 0, 09 | 0, 06 | 0, 24 |
| 10 | 0, 27 | 0, 08 | 0, 06 | 0, 21 |

F1-Score

MovieLens

| k | FolkRecommender | Bellogin <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|------------------------|------------------------|----------|
| 6 | 0, 57 | 0, 20 | 0, 09 | 0, 35 |
| 7 | 0, 56 | 0, 16 | 0, 08 | 0, 31 |
| 8 | 0, 54 | 0, 14 | 0, 07 | 0, 31 |
| 9 | 0, 56 | 0, 11 | 0, 06 | 0, 31 |
| 10 | 0, 52 | 0, 10 | 0, 06 | 0, 29 |

BookCrossing

| k | FolkRecommender | Kim <i>et al.</i> | Qumsiyeh <i>et al.</i> | PersoRec |
|-----|-----------------|-------------------|------------------------|----------|
| 6 | 0.48 | 0.20 | 0.15 | 0, 41 |
| 7 | 0.45 | 0.18 | 0.14 | 0, 38 |
| 8 | 0.40 | 0.18 | 0.13 | 0, 35 |
| 9 | 0.41 | 0.15 | 0.12 | 0, 32 |
| 10 | 0.39 | 0.13 | 0.12 | 0, 28 |

Evaluation metrics

Explanations

- related works rely on most used items (books, movies, tags)
- user's profile : useful information that improve recommendations
- quadri-concepts advantage : recommending the most shared tags and resources

Properties of FolkRecommender

User Space Coverage

- All users are covered : all users receive recommendations whenever they shared or not.
- Covered Profiles : 100% of genders (male and female), 100% of age categories, 100% of professions and 88% of countries

Cold Start : New Users

- Recommendations based on the user's profile first (as stereotypes)
- All users receive recommendations despite they did not share anything yet.

Properties of FolkRecommender

Item Space Coverage

- 13,62% of the MovieLens dataset's resources are covered (i.e., recommended by FolkRecommender)
- 71,62% of the BookCrossing dataset's resources are covered (i.e., recommended by FolkRecommender)

In general, the item space is quite well covered which shows that quadri-concepts are representative structures of the *v-folksonomy* with minimal loss of information.

Properties of FolkRecommender

Scalability

| | <i>minsupp-u</i> | $ QC $ | # UU | Task1 (ms) | Task2 (ms) |
|-----------------------|------------------|--------|-------|------------|------------|
| (MovieLens) | 20 | 221 | 526 | 0,1 | 2,6 |
| | 16 | 500 | 605 | 0,2 | 3,9 |
| | 12 | 1295 | 668 | 0,7 | 6,1 |
| | 8 | 5123 | 805 | 4,0 | 13,5 |
| | 6 | 13461 | 865 | 12,7 | 23,3 |
| (BookCrossing) | 30 | 553 | 6789 | 0,9 | 149,8 |
| | 20 | 1486 | 9092 | 4,9 | 296,9 |
| | 16 | 2638 | 10397 | 13,0 | 415,5 |
| | 12 | 5698 | 12239 | 45,0 | 542,3 |

- Quadri-Concepts = small representation of a v-folksonomy
- # UU = # unique users = number of recommendations
- An average response time of 0,002 seconds for the recommendation of resources (Task 1) and around 0,008 seconds for the user proposition task (Task 2).

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 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 Work
- 3 A new personalized recommender system
- 4 Results and Discussion
- 5 Conclusion and Perspectives

Conclusion

Contributions

- Consideration of the user's profile
- More targeted recommendations
- Proprieties analyze of the recommender system

Limits

- The first recommendation stills dependent of the quadri-concept's extraction
- Our recommender system rely on *static* quadri-concepts which not evolve through time
⇒ needing an incremental method to update the set of quadri-concepts in order to propose the most recent recommendation

Perspectives

- Extension of the 4th dimension to other variables : time, connection history, etc.
- Online system : an online tracking of users' tags and resources
- Incremental algorithm in order to track the dynamic updates of folksonomies

That's all folks!
Your Questions?