

Szymon CHOJNACKI, Mieczysław KŁOPOTEK  
Instytut Podstaw Informatyki Polskiej Akademii Nauk

## ENSEMBLE OF INCREMENTAL MULTILABEL CLASSIFIERS FOR TAG RECOMMENDING IN SOCIAL NETWORKS

**Summary.** The purpose of this article is to present profits and costs of enriching state of the art real life tag recommender system with incremental learning mechanisms. We describe modifications to a system that successfully participated in Online Task of ECML/PKDD Discovery Challenge 2009. The system's architecture follows an idea to construct hierarchical ensemble of simple classifiers, which was implemented in various ways by the systems with highest performance in the Challenge. The system is currently integrated as a web service with BibSonomy bookmarking portal and outperforms other algorithms in terms of effective latency. We focus on incremental learning techniques that improve quality of the system's recommendations, but do not raise maintainability, efficiency or reliability issues.

**Keywords:** Tag recommender, multilabel classification, social networks

## RODZINA PRZYROSTOWYCH KLASYFIKATORÓW WIELOETYKIETOWYCH DLA REKOMENDACJI OTAGOWAŃ W SIECIACH SPOŁECZNYCH

**Streszczenie.** Celem artykułu jest prezentacja korzyści i zagrożeń związanych ze wzbogaceniem systemu rekomendującego etagowania o mechanizmy uczenia przyrostowego. Opisujemy modyfikacje do systemu, który z powodzeniem brał udział w konkursie ECML/PKDD Discovery Challenge. Architektura systemu oparta jest na idei wspólnej dla systemów, które osiągnęły najwyższe oceny podczas konkursu i składa się z hierarchicznego łączenia wyników prostych klasyfikatorów. Opisany system jest obecnie zintegrowany z serwisem internetowym BibSonomy i osiąga najwyższe oceny ze względu na czas dostarczania rekomendacji dla użytkowników. W tym artykule skupiamy uwagę na zastosowaniu technik uczenia przyrostowego, które nie powodują znacznego obniżenia prędkości systemu lub obniżają jego niezawodność we wdrożeniu w środowisku badawczym.

**Słowa kluczowe:** rekomendacje etagowań, sieci społeczne

## 1. Introduction

Tagging digital resources is an important element of mass collaboration and knowledge sharing in social networks. Users of Web 2.0 portals participate in content creation of their favorite services by bookmarking photos, videos, web sites, blogs or other resources, ranking them and attaching labels. A list of portals that enable users to save (bookmark) favorite resources, share it, label and make public for other users is growing. Among most popular services are Delicious, BibSonomy, Digg, Reddit, CiteULike or Simpy. BibSonomy and CiteULike enable users to bookmark publications. Most of the systems have basic recommender system that suggests tags based on historical resource profile filtered by user's profile. However, empirical research show that this simple approach is not sufficient for real life data, which evolve in a fast pace and new resources are being tagged repeatedly. In such case content based techniques are essential and our system relies heavily on content of a resource.

The second section of this article describes related work. In the third section we describe the structure and content of datasets used in tag recommending tasks. We also show how the data are created by system-user interaction. In the fourth section we define measures used to evaluate performance of a tag recommending system. Following section contains details of our system. In the final section we point out the directions of further research.

## 2. Related work

The changes that we observe in the sociological aspect of the above described phenomenon are reflected by growing interest of scientific community in applying machine learning algorithms to social network analysis. Recommending tags in folksonomies can be seen as an extension to classic recommendation problem, in which bi-partite representation is extended to tri-partite graph data structure. However, techniques used in the former problem e.g. collaborative filtering were reapplied to our task. The task is both interesting and demanding, it requires simultaneous application of methods from natural language processing, knowledge engineering, graph theory, information retrieval, multilabel classification. Therefore several diverse approaches have been proposed. The authors of [10] look at the problem from ontological perspective. Association rules mining approach was verified in [5]. PageRank algorithm typical for searching engines was adjusted to our task in [8] and was named FolkRank. Attaching labels to photos in Flickr portal was improved by construction of co-occurrence graph in [11]. The task of finding best tags for a resource and

a user was optimized by unsupervised clustering techniques [2, 12] and supervised classification [1].

### 3. Datasets

The dataset used to train models consisted of three files. The files contained 1 401 104 tag assignments, 263 004 bookmark posts and 158 924 BibTeX posts. This data was a full snapshot of BibSonomy.org users' activity until the end of 31-12-2008. Test set consisted of posts bookmarked in the system in the first half of 2009. A print screen with an example of tag assignment process is presented in Fig. 1. In this example the website of BDAS (Bazy Danych: Aplikacje i Systemy) conference is labeled with one tag "bdas".

Fig. 1. Example of user-system interaction in BibSonomy

Rys. 1. Przykład interakcji użytkownik-system w BibSonomy

BibSonomy portal enables to bookmark also publications, which can be described by several additional fields (e.g. volume, chapter, edition, institution, publisher, journal, pages, author). The total number of distinct user, URLs, BibTeXs and tags amounts to 3 617, 235 328, 143 050 and 93 756 respectively.

This type of datastructure (containing assignments of tags by users to resources) is usually named Folksonomy. Formally, a Folksonomy is a tuple  $F := (U, T, R, Y)$ , where  $U$  (users),  $T$  (tags) and  $R$  (resources) are finite sets.  $Y$  stands for a ternary relation between them. If we define a set of tags attached by user  $u$  to resource  $r$  by  $T_{ur} := \{t \in T \mid (u, t, r) \in Y\}$ , then a post is  $(u, T_{ur}, r)$ .

## 4. Evaluation methodology

Tagging Primary evaluation measure for tag recommender systems are precision, recall and f1-measure. In case of multilabel classification they are defined slightly different than for classic one class or multi-class classification. For each post we compare tags assigned to it by a user with tags proposed by a recommender. In case of the challenge only first five tags were compared. Additionally, the tags were normalized by Java method. All the technical details of calculating evaluation measures can be found in [6]. These measures are used for offline evaluation, but can be easily extended for online evaluation [7]. In order to compute precision and recall we need to count the tags that appear both in user's final assignment and system's recommendation. Secondly we divide this amount by the number of user's tags to obtain recall (Equation 1), and by the number of system's tags to obtain precision (Equation 2). These values are summed for all posts and the f1-measure is obtained from those aggregates (Equation 3).

$$recall(T(u, r)) = \frac{|T_{ur} \cap T(u, r)|}{|T_{ur}|} \quad (1)$$

Precision and Recall differ by denominator:

$$precision(T(u, r)) = \frac{|T_{ur} \cap T(u, r)|}{|T(u, r)|} \quad (2)$$

F-measure represents a trade-off between both measures:

$$f1m = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (3)$$

## 5. Our algorithm

Our base algorithm is based on counting simple probabilities and merging the weighted lists. This idea is common for the most successful algorithms during Discovery Challenge 2009. The most successful implementation was proposed by Lipczak [9]. Our implementation outputs comparable f-measure results (at level of 18,8%), but is based on simpler structure of the hierarchical ensemble. We do not utilize information that is particularly difficult to maintain in incremental learning i.e. tag coocurance graph and word-tag coocurance graph. For the details of Lipczak's system please refer to [9].

### 5.1. Calculating probabilities

In the first step we calculate probabilities that a given post  $(u, T_{ur}, r)$  will be tagged with labels based on user's profile (equation 4), resource's profile (equation 5), title's and URL's content (equation 6). Probabilities for tokens extracted from post's URL (not calculated for BibTeX entries) are analogous to title's probabilities.

$$P(\text{tag} = t \mid \text{user} = u) = \frac{|Y \cap \{u\} \times \{t\} \times R|}{|Y \cap \{u\} \times T \times R|} \quad (4)$$

$$P(\text{tag} = t \mid \text{resource} = r) = \frac{|Y \cap U \times \{t\} \times \{r\}|}{|Y \cap U \times T \times \{r\}|} \quad (5)$$

$$P(\text{tag} = t \mid t \in r_{\text{title}}) = \frac{|Y \cap U \times \{t\} \times R|}{|\{t \in r_{\text{title}} \mid r \in R\}|} \quad (6)$$

### 5.2. Weights optimization

In the second step we merge the lists of recommended tags from title, URL and resource profile, by multiplying the probabilities by 0.9, 0.1 and 0.3 respectively and for the tags that appear in user's profile we add both probabilities (the latter is weighted by 0.9). The outcome is limited to the tags which final score is higher than an optimized 0.05 threshold.

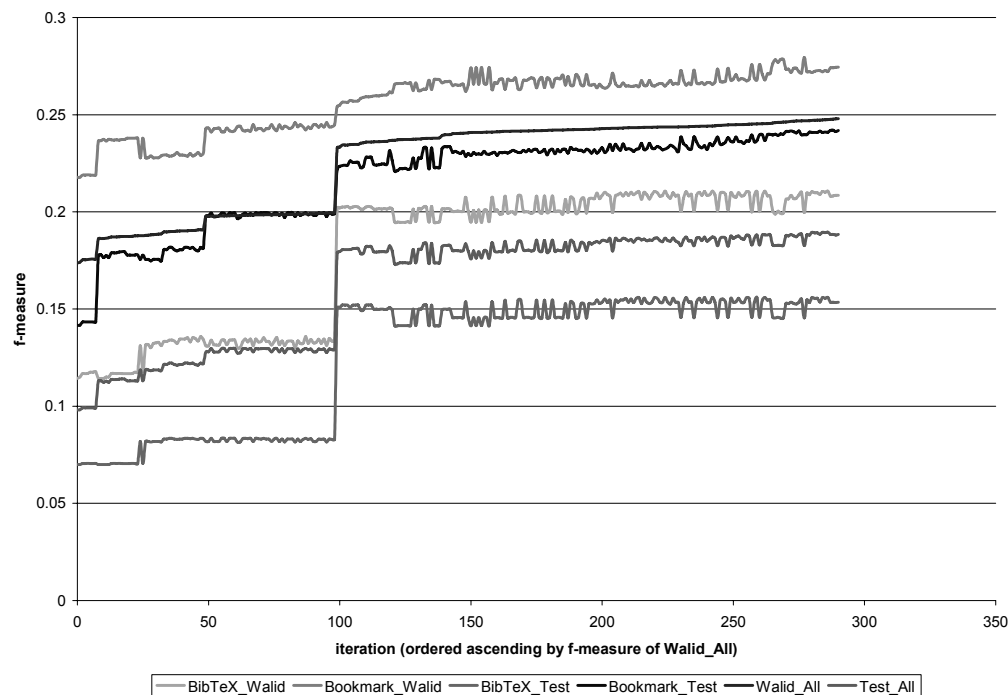


Fig. 2. Performance of recommender systems  
Rys. 2. Wyniki systemów rekomendujących

The weights used in our system were optimized by counting f-measure for a validation set (Figure 2). The validation set consists of posts bookmarked in BibSonomy three months before the first date of test set's minimal date. The recommender system used to predict tags for validation set was not trained on data contained in validation or test sets. The data were additionally cleaned from posts that were automatically loaded from other services to the BibSonomy. The most important observation is that we did not observe over-fitting during this optimization and results for different weights are correlated between validation and test sets. We can also see that it is more difficult to recommend tags for BibTeX posts than for websites. The difference of f-measure between validation and test sets are higher than 5%, which is due to cleaning of validation set. It is surprising that in most of weight's variants we obtain very high performance (we evaluated  $4*4*4*4=256$  variants). For each classifier we checked systems performance when the weight was set to 0.1, 0.3, 0.6 and 0.9.

### 5.3. Incremental learning

The simplest idea in incremental learning is to learn a model from a scratch when new data arrive. This approach is represented by a green and red lines in Figure 3. Green recommender was learned with data starting 3 months prior to the start of this evaluation, blue recommender was learned with all data available prior to the start. The green algorithm gives average 17.8% f-measure during 6-months evaluation period, the blue algorithm gives higher average f-measure of 18.8% (the best result during ECML/PKDD Discovery Challenge 2009 for this data was 18.7% [9]). This result is consistent with our intuition, however it is not a practical result in our setting, mostly because restarting a web service rises reliability and maintainability issues.

Our second incremental approach, shown by a red line in Figure 3, gives an average f-measure of 20.65%. This approach is based on online updating statistics for resource's and user's profiles after every bookmarking. It is worth mentioning that when we only inserted new profiles for resources and users we obtained 20.0% f-measure. The decrease of accuracy is not remarkable, but only inserting new profiles is much faster than updating old ones, and does not influence synchronization aspects of giving recommendations in multithreaded environment. The results that we present do not contain updating probabilities for tokens in post's title or post's URL. It is interesting that when we update probabilities for resource and user profiles separately, slightly higher improvement is obtained for a resources profile incremental learning.

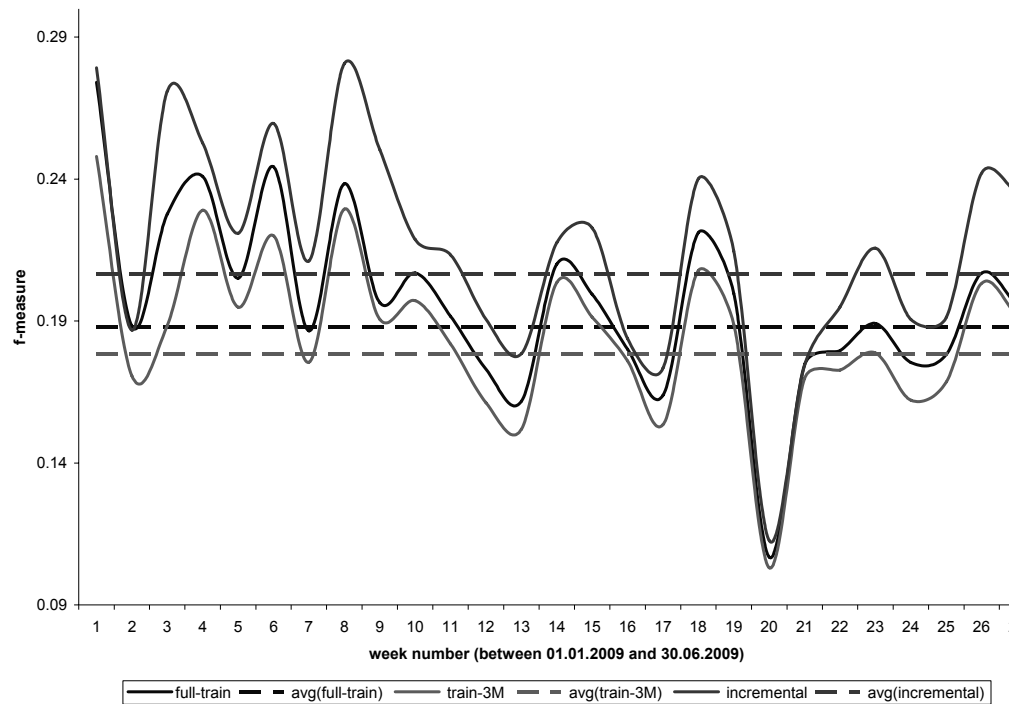


Fig. 3. Incremental approaches to tag recommendation

Rys. 3. Przyrostowe podejście do rekomendacji otagowań

In the following subsection we present in detail recommendation process for a single post.

#### 5.4. Example

We consider a post bookmarked by a user with 3 022 identifier on 01.10.2008. The user attached 10 tags to the website (Table 1) i.e. code, extensions, firefox, form, html, moteur, rechercher, script, visualiser and webdeveloper. The resource was not bookmarked by any user earlier. The title consisted of nine words/tokens in English, but extended description was in French. The tokens extracted from URL, title and extended title are given in Table 2. The probabilities of being a tag for those tokens are also given in Table 2. Lists of suggested tags are merged with weighting and filtered by tags used by a user earlier (e.g. best, extension, outil, flickr). The final set of recommended tags consist of five tags with highest scores i.e. firefox, extension, google, outil, recherché. There are three common tags in recommender's output and user's selection, which gives precision of 60%, recall of 30% and f-measure of 40%.

Table 1

## Website bookmarked in BibSonomy portal

Date	Wed Oct 01 03:58:13 CEST 2008
User Id	3022
URL	<code>http://www.google.fr/search?num=100&amp;hl=fr&amp;safe=off&amp;rlz=1B3GGGL_frFR276FR277&amp;q=Convert+POST+to+GET+in+forms+firefox&amp;btnG=Rechercher&amp;meta=cr%3DcountryFR</code>
Title	Convert POST to GET in forms firefox - Recherche Google
Extended Title	Je cherche à savoir quelle données passent en post après le submit d'un form.. Y'a t'il une extension de firefox qui permette de visualiser les données en post ? - - WebDeveloper Ensuite sur la page de ton formulaire, Forms>Convert Form Method> ...
Websites's previous tags	None
Tags attached to the website	code extension firefox form html moteur rechercher script visualiser webdeveloper
Number of attached tags	10

Table 2

## Recommendations of the system

Tokens extracted from Title	convert, post, to, toget, get, in, forms, firefox, recherche, google, je, savoir, quelle, données, en, le, submit, d, dun, un, form, y, a, ...
Tokens extracted from URL	http, www, google, fr, search, num, 100, hl, safe, off, q, convert, post, to, toget, get, in, forms, firefox, rechercher, meta, cr
Tags from title	firefox 0.69 google 0.47 extension 0.46 recherche 0.15 formulaire 0.13 url 0.11 variables 0.11 forms 0.1 convert 0.08
Tags from url	firefox 0.69 google 0.47 search 0.31 forms 0.1 convert 0.08
User's previous tags	best 0.47 extension 0.39 outil 0.36 flickr 0.35 howto 0.35 tutoriel 0.3 firefox 0.27 ... fun 0.14 recherche 0.14 tag 0.14 utilitaire 0.14 google 0.14
Tags recommended	firefox 0.92 extension 0.84 google 0.57 outil 0.4 recherche 0.29
Tags attached to the website	code extension firefox form html moteur rechercher script visualiser webdeveloper
Matching tags	firefox extension recherché
Precision	0.6
Recall	0.3
F-measure	0.4



## 6. Conclusion

In this article we described results of improving state of the art recommender system with simple incremental learning mechanisms. The f-measure for evaluation period of incremental algorithm is significantly higher than of an algorithm learned once at the beginning of the evaluation period. These results are consistent with our intuition. It is important to mention that we updated single classifiers and left the ensemble structure unchanged during evaluation period. In the future we are planning to verify the impact on performance of more advanced incremental learning techniques (e.g. learning to forget and temporal activation of single parameters).

## BIBLIOGRAPHY

1. Basile P., Gendarmi D., Lanubile F., Semeraro G.: Recommending smart tags in a social bookmarking system. In Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), 2007, s. 22÷29.
2. Chojnacki S.: Tag Recommendations based on Tracking Social Bookmarking Services. Proceedings of the ECML-PKDD Discovery Challenge Workshop, Bled, Slovenia, 2009.
3. Eisterlehner F., Hotho A., Jäschke R.: ECML PKDD Discovery Challenge 2009. CEUR-WS.org, 2009.
4. Golder S. A., Huberman B. A.: Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198, 2006.
5. Heymann P., Ramage D., Garcia-Molina H.: Social tag prediction. In SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, , New York, NY, USA, ACM, 2008, s. 531÷538.
6. Jäschke R., et. al.: Tag Recommendations in Social Bookmarking Systems. *AI Communications*, Amsterdam: IOS Press, Vol. 21, No. 4, 2008, s. 231÷247.
7. Jäschke R., et. al.: Testing and Evaluating Tag Recommenders in a Live System. Workshop on Knowledge Discovery, Data Mining, and Machine Learning, 2009, s. 44÷51.
8. Jäschke R., et. al.: Tag Recommendations in Folksonomies. *LNCS*, 4702:506, 2007.
9. Lipczak M., et. al.: Tag sources for recommendation in collaborative tagging systems. Proceedings of the ECML-PKDD Discovery Challenge Workshop, Bled, Slovenia, 2009.
10. Macgregor G., McCulloch E.: Collaborative tagging as a knowledge organization and resource discovery tool. *Library Review*, 55(5), 2006, s. 291÷300.

11. Sigurbjornsson B., Zwol R.: Flickr tag recommendation based on collective knowledge. WWW '08: Proceeding of the 17th international conference on World Wide Web, , New York, NY, USA, ACM, 2008, s. 327÷336.
12. Song Y., et. al.: Realtime automatic tag recommendation. SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, , New York, NY, USA, ACM, 2008, s. 515÷522.

Recenzenci: Dr inż. Michał Kozielski  
Prof. dr hab. inż. Alicja Wakulicz-Deja

Wpłynęło do Redakcji 31 grudnia 2009 r.

## Omówienie

W artykule opisano został system rekomendujący w czasie rzeczywistym spersonalizowane etykiety dla publikacji i stron internetowych zapisywanych przez użytkowników serwisu społecznościowego BibSonomy.

Czynność nadawania etykiet ulubionym zasobom cyfrowym (*ang. bookmarking and tagging*) staje się coraz bardziej popularna wśród użytkowników internetu. Etykiety spełniają podobną rolę jak słowa kluczowe przy artykule naukowym i pozwalają usprawnić procesy wyszukiwania i porządkowania informacji. Szczególnie duża wartość dodana związana jest z nadawaniem etykiet zasobom innym niż tekstowe (np. zdjęcia, utwory muzyczne, filmy). Jednak także dla zasobów tekstowych (np. strony internetowe, publikacje, blogi) opisane metody pozwalają poprawić jakość serwisów internetowych.

Opisujemy system, który proponuje spersonalizowane etykiety użytkownikom serwisu społecznościowego do zarządzania wiedzą. Do najbardziej popularnych serwisów o takim przeznaczeniu należą: Delicious, BibSonomy, Digg, Reddit, CiteULike oraz Simpy. W odróżnieniu od klasycznych etykiet występujących w internecie w postaci chmurki (*ang. tag clouds*) serwisy te pozwalają na spersonalizowanie tworzonych etykiet.

W wyniku przeprowadzonych symulacji pokazaliśmy, że wdrożenie mechanizmów uczenia przyrostowego do systemu składającego się z rodziny prostych klasyfikatorów pozwala na uzyskanie rekomendacji istotnie lepszych (ze względu na f-miary) niż najlepsze systemy uczące się na danych z obciążonym zakresem czasowym. Dodatkowo wskazaliśmy, które elementy złożonego klasyfikatora powinny być na bieżąco aktualizowane, a którym może zostać nadany niższy priorytet.

**Address**

Szymon CHOJNACKI: Instytut Podstaw Informatyki Polskiej Akademii Nauk,  
ul. J. K. Ordona 21, 01-237 Warszawa, Polska, sch@ipipan.waw.pl .