Feedback Options for a Personal News Recommendation Tool

C. Bomhardt and W. Gaul

Abstract Recommendations can help to vanquish the information overload problem on the web. Several websites provide recommendations for their visitors. If users desire recommendations for sites without this service, they can use browsing agents that give recommendations. In order to obtain user profiles, common agents store interesting and/or uninteresting web pages, use them as training data for the construction of classifiers, and give recommendations for unseen web pages. Feedback via explicit rating is regarded as most reliable but exhausting method to obtain training data. We present three alternative feedback options (explicit, implicit, and hybrid) and evaluate the alternatives via SVMs. We show that feedback options that are less exhausting than explicit rating can be applied successfully.

1 Introduction

Information overload is a severe problem that influences web usage. It arises out of the sheer mass of available information on the web and can even worsen due to technical reasons. Welcome pages of several online news websites present lists of headlines and abstracts of articles. A user has to click on, e.g., a headline and wait until the web page containing the accompanying article is transferred and displayed. In order to read another article, users often have to navigate back to the welcome page. Compared to fast running over the pages of a classical newspaper, information gathering via online news websites can be time consuming, less intuitive, and uncomfortable. This problem is not limited to online newspapers but affects also other websites like online shops or portals. Several websites use personalization techniques to address this problem. According to Mobasher et al. (2000), web personalization

W. Gaul(✉)
Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe (TH), 76128 Karlsruhe, Germany, E-mail: wolfgang.gaul@wiwi.uni-karlsruhe.de

A. Okada et al. (eds.), *Cooperation in Classification and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organizations.
DOI: 10.1007/978-3-642-00668-5_9,

can be described as any action that makes the web experience of a user personalized to the user's taste? One can further distinguish between personalization by information filtering and personalization by information supplementing. Most online shops offer information filtering via a product search function (a user can supply keywords and the shop presents a filtered product list) and automatic information supplementing (the shop recommends additional or alternative products). Some online newspapers allow their users to subscribe to news of predefined categories. All these examples have in common that they are offered by website operators and are limited to websites that provide this kind of service. If a user desires recommendations on a website without such a service, additional tools like browsing agents are necessary.

2 Interest Profiles

Interest profiles of some kind are required by every browsing agent. Some agents model user interests with the help of user-supplied keywords. Unseen web pages are recommended, if they are similar enough to the specified keywords (filtering-based approach). Other agents model user interests with the help of collections of interesting and/or uninteresting web pages. Classifiers are trained on these collections and used to obtain recommendations for unseen web pages (machine-learning based approach). Training pages are needed for the machine-learning based approach. Typical agents therefor allow their users to explicitly rate the page currently shown. This is regarded as most reliable method to obtain feedback but exhausting for the user. A more convenient way is to monitor user behavior and thereby collect interesting and uninteresting web pages.

Browsing agents do exist for various applications. A good overview is given in Middleton (2001). NewsWeeder (Lang 1995) is an agent for Usenet newsgroups. It is realized as web based newsgroup client, uses explicit feedback in order to learn preferences, and compiles personalized news collections. WebMate (Chen and Sycara 1998) is a personal browsing agent that tracks interesting documents via explicit feedback. A new document is recommended if it is similar enough to an interesting reference document. Personal WebWatcher (PWW) (Mladenić 2001) accompanies a single user to become a specialist concerning the interests of the corresponding person. PWW records the URLs of pages read by the user and considers web pages shown as interesting. WebWatcher (Joachims et al. 1997) collects interest profiles of several users based on keywords. It recommends web pages that were interesting for other people with similar interests and relies on implicit (link followed) and explicit (users leaving the website can tell if their tour was successful or not) feedback. The browsing agent NewsRec (Bomhardt 2004) is specialized in recommendations for online newspapers and uses explicit feedback.

3 Feedback Options

Today's machine-learning based browsing agents typically offer explicit feedback (NewsWeeder, WebMate, NewsRec) or track shown web pages (PWW) and consider them interesting (implicit feedback). This kind of implicit feedback leads to mislabeled documents if users request uninteresting web pages, a rather common situation. Other implicit interest indicators proposed by Claypool et al. (2001), Kelly and Belkin (2004), Kelly and Teevan (2003) include display time, scrolling activity, mouse activity, keyboard activity, bookmarking, and saving or printing of web pages. Thereof, display time is, according to literature, the most promising. While Claypool et al. (2001) found display time to be a good indicator, it turned out to be unsuitable in Kelly and Belkin (2004). We expect display time to be inadequate for personal news recommendations due to three reasons: (1) most articles are short, thus, the variance concerning transfer delays can outweigh the variance concerning viewing time of a web page, (2) in contrast to the basic assumption that long display times indicate interestingness, very short articles ("new security update available") can be interesting, and (3) telephone calls or other external interruptions can lead to artificially long display times during regular usage outside of a laboratory. Additional problems of some of the mentioned indicators concern their availability w.r.t. regular browsers (e.g., keyboard activity) or that most classifiers require positive and negative training examples, but not every indicator identifies interesting AND uninteresting content (e.g., articles in web pages not printed are not necessarily uninteresting).

Our goal was to reduce the burden of explicit feedback and increase the number of correctly labeled training documents as compared to implicit feedback. The main problem of implicit feedback, as implemented by PWW, is that every web page shown is considered as interesting, even if it is uninteresting. This problem is mitigated by our hybrid feedback approach.

Hybrid feedback combines implicit and explicit feedback in such a way that implicit feedback is superimposed by explicit feedback because users can explicitly rate pages seen as uninteresting.

Hybrid feedback reduces the burden of rating articles. Mislabeled training examples can still occur if unseen pages that are automatically considered as uninteresting are indeed interesting. However, this situation should not occur too often in a news recommendation application, as it can be assumed that users of a special interest news website set value on reading all interesting articles.

Feedback options require information about whether an article was requested (yes/no) and/or which explicit user rating was assigned (none, +, −) in order to incorporate it into a set of training documents. All possible combinations together with the true rating and the corresponding assignments to one of the training sets are contained in Table 1. Six cases are possible. It is assumed that a rating, if given, is valid and matches the true rating (cases 3 and 4 of Table 1). Cases 1 and 2 describe situations where articles were seen but not rated. According to the true rating, in case 1 the corresponding article is interesting, in case 2 it is not. The implicit and hybrid feedback options both assign the corresponding articles to the set of

Table 1 Assignment of training documents for different feedback options

Case	Article requested?	User rating	*True rating*	Training set based on feedback option		
				Explicit	Implicit	Hybrid
1	Yes	None	+	None	+	+
2	Yes	None	−	None	+	+
3	Yes	+	+	+	+	+
4	Yes	−	−	−	+	−
5	No	None	+	None	−	−
6	No	None	−	None	−	−

interesting documents, explicit feedback doesn't assign. In case 3 the article was seen, obtained a positive rating, and is interesting. All feedback options assign this article to the set of interesting training documents. In case 4, the article was seen, obtained a negative rating and is uninteresting. Explicit and hybrid feedback assign it to the set of uninteresting training documents whereas implicit feedback assigns it to the set of interesting training documents. In cases 5 and 6 the corresponding articles were not seen and, thus, couldn't obtain a user rating. In case 5 the article is interesting, in case 6 the article is uninteresting for the user. Explicit feedback doesn't assign the articles to any training set; hybrid and implicit feedback assign both articles to the set of uninteresting training documents. As one can see from Table 1, explicit feedback collects two training documents compared to the 6 documents of implicit and hybrid feedback. Explicit feedback has no mislabeled training documents, implicit feedback 3 (50% of the cases shown in Table 1) and hybrid feedback 2 (33%). For implicit feedback, no user rating action is required to obtain the set of training documents as all requested articles are labeled as interesting and the rest as uninteresting. For explicit feedback, only articles that are requested can be rated. Here, 2 rating actions occur. Within hybrid feedback only 1 rating action (a user who knows how hybrid feedback works doesn't have to rate in case 3) is necessary. If a user correctly handles a system with hybrid feedback, the situation of case 2 should not occur (because the user knows that without an user rating a requested but uninteresting article will be assigned to the set of interesting training documents by hybrid feedback). Thus, the error rate of hybrid feedback drops down to 16% of the cases shown in Table 1. This comparison shows that hybrid feedback requires less user ratings compared to explicit feedback and can lead to more training documents. Compared to implicit feedback, it misassigns less documents.

4 Web Page Classification

Due to space restrictions, we are limited to a brief overview of web page classification as used by NewsRec (see Bomhardt 2004 for further details). Typical web pages consist of HTML code containing the article text and formatting instructions, navigational elements, and advertisements (see Fig. 1). A web page should

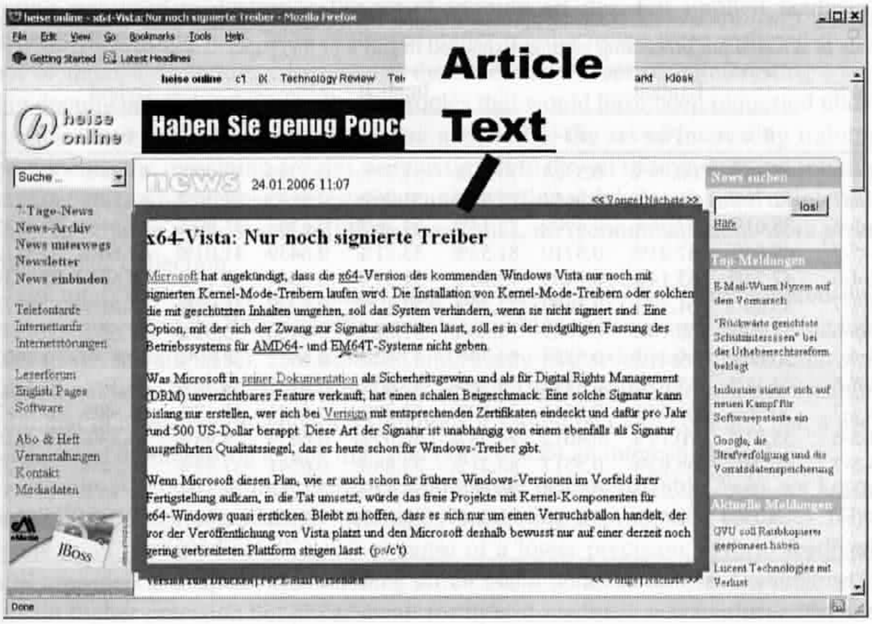

Fig. 1 A screenshot of the examined news website with highlighted article text

be classified based on its content without framing elements. In this context, the content is the article text. Its generic extraction is a problem on its own (Kushmerick et al. 1997). Thus, a wrapper was build manually. The extracted text still is contaminated with HTML tags like <*B*> that have to be removed. The remaining text is converted to lowercase and transformed to vector space model representation. As SVMs are used for classification, dimensionality reduction is not required. For every pre-processing step (*frequency transformation* [TF (1), LOG (2), BIN (3)], *term weighting* [NOWEIGHTS (4) or IDF (5)], and *normalization* [NONE (6), L2-NORM (7)]), one possibility has to be selected. We checked all possible combinations of these settings (see the first column of Table 2 for the application of different pre-processing settings). Further settings do exist but are not considered here as this work focuses on feedback options rather than pre-processing settings for text classification. Our pre-processing steps include settings recommended by Joachims (2002) for text classification with SVMs.

5 Evaluation Method

We selected recall (rec), precision (prec), and F1 as evaluation measures for web page classification. It should be mentioned that recall can be calculated here due to the known total number of news articles on a given website. This is not true

Table 2 Comparison of the feedback options. Column TWS (term weighting-scheme) contains the code of selected pre-processing settings explained in part 4 of this paper

TWS	Explicit			Implicit			Hybrid		
	Rec	Prec	F1	Rec	Prec	F1	Rec	Prec	F1
1-4-6	47.24%	60.99%	0.5324	70.50%	51.31%	0.5939	39.09%	61.28%	0.4773
1-4-7	52.76%	61.11%	0.5663	77.94%	51.26%	0.6185	44.36%	63.14%	0.5211
1-5-6	38.61%	67.36%	0.4909	73.14%	54.76%	0.6263	31.89%	67.51%	0.4332
1-5-7	49.64%	67.21%	0.5710	81.53%	53.21%	0.6439	41.01%	67.86%	0.5112
2-4-6	47.24%	63.14%	0.5405	72.90%	51.61%	0.6044	38.13%	62.60%	0.4739
2-4-7	53.00%	61.39%	0.5689	77.46%	52.10%	0.6230	45.08%	63.51%	0.5273
2-5-6	37.89%	68.10%	0.4869	75.06%	57.01%	0.6480	33.09%	72.63%	0.4547
2-5-7	50.60%	69.41%	0.5853	82.73%	53.82%	0.6522	38.85%	69.83%	0.4992
3-4-6	44.12%	63.67%	0.5212	74.34%	52.90%	0.6181	38.37%	66.39%	0.4863
3-4-7	50.36%	61.58%	0.5541	76.50%	51.29%	0.6141	43.17%	64.98%	0.5187
3-5-6	35.01%	70.19%	0.4672	74.58%	56.65%	0.6439	29.98%	73.53%	0.4259
3-5-7	46.04%	68.82%	0.5517	83.21%	53.88%	0.6541	37.89%	70.22%	0.4922
Mean*	46.04%	65.25%	0.5364	76.66%	53.32%	0.6284	38.41%	66.96%	0.4851
Stand. dev.*	0.0599	0.0358	0.0377	0.0407	0.0200	0.0197	0.0478	0.0401	0.0336

*The mean and standard deviation values for the columns were incorporated on request of a reviewer

for other information retrieval tasks where the total number of relevant documents can be unknown. A typical user of NewsRec would read (and rate) a bunch of articles, train a classifier on the rated articles, obtain recommendations, and read (and rate) the next bunch of articles, train an updated classifier on all rated articles,.... If the true class labels of all articles are known, usage can be evaluated as follows: train a classifier on the first bunch of articles (training sets assigned by feedback option), compare predicted class labels with the true labels of the next bunch, train a new classifier on the first two bunchs of articles (training sets assigned by feedback option), evaluate it on the third bunch,.... Recall and precision for all evaluated bunchs are microaveraged; F1 is calculated on microaveraged recall and precision. On the examined website, about 50 new articles are presented per day. Thus, we set the size of a bunch to 50. Articles were sorted according to their order of appearance; the oldest articles were assigned to bunch 1.

6 Empirical Results

NewsRec had to be extented to allow for the comparison of explicit, implicit, and hybrid feedback options. An IT professional had to read and explicitly rate 1,185 articles of the Heise newsticker (http://www.heise.de/ct). In addition to the rating, the user had to submit whether (s)he would have requested the article during regular usage. This approach allowed us to evaluate the various feedback options on the same dataset, thus leading to comparable results. For explicit feedback, the true rating was used to determine the set of training articles. For implicit feedback, articles that would have been requested under regular usage were assigned to the set of interesting training documents, the others to the set of uninteresting training documents. For hybrid feedback, articles that would have been requested under regular usage and rated interesting were assigned to the set of interesting training documents, the remaining articles were assigned to the set of uninteresting training documents (assuming that requested but uninteresting articles were rated uninteresting). In order to prevent self fulfilling prophecies, no recommendations were given during data collection.

In total, 449 articles were interesting for the user (37%). Four hundred and twenty-five out of the 449 interesting articles would have been requested during regular usage (recall: 94%). Two hundred and twenty-five articles that were requested turned out to be uninteresting (precision: 65%). These results confirm our assumption that in news recommendation situations it should be more common that a user looks at an uninteresting article than that (s)he misses an interesting one.

Results for the examined feedback options are printed in Table 2. As we know, implicit feedback considers more articles interesting than explicit feedback. This leads to an increased recall at the expense of a lower precision. Hybrid feedback still considers not rated but interesting articles as uninteresting. The expectation to obtain higher precision but lower recall for hybrid feedback was confirmed by our empirical results.

In terms of F1, implicit feedback won the competition on this dataset due to the high recall values obtained.

For our application, however, precision is the more interesting measure. Here, hybrid feedback is best while the implicit counterpart only gets the last position.

Another possible cause for the results could be that the sets of training documents obtained through explicit feedback may lead to overfitted classifiers whereas hybrid feedback reminds the user that (s)he should rate a requested but uninteresting article as uninteresting.

7 Conclusions

Explicit feedback is considered as most reliable feedback option, but it is unpopular due to the required additional effort to rate all requested documents. Implicit feedback requires no additional effort at all, but it is considered as unreliable because it mislabels requested articles that are uninteresting and unrequested ones that are interesting. We combined both approaches within the so-called hybrid feedback option in order to reduce the burden of explicit feedback and lower the number of mislabeled training documents obtained by implicit feedback.

Obviously, implicit feedback leads to the largest absolute and relative numbers of positive training documents. The classifiers react accordingly. Implicit feedback leads to the highest recall, followed by explicit and hybrid feedback for our data. However, in terms of precision, the order is reversed for the same reasons.

The increase of recall for implicit feedback can outweigh the decrease in precision (as in the underlying example) as far as highest F1 values are concerned.

With respect to our application, hybrid feedback turns out to be the best choice: users, if in doubt, tend to look at articles. Thus, precision is the most important measure and hybrid feedback, which optimizes this measure, requires less effort than explicit feedback.

References

BOMHARDT, C. (2004): Newsrec, a SVM-driven personal recommendation system for news websites. In *WI 04: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 545–548.

CHEN, L. and SYCARA, K. (1998): Webmate: A personal agent for browsing and searching. In *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, AGENTS 98*, 132–139.

CLAYPOOL, M., LE, P., WASED, M., and BROWN, D. (2001): Implicit interest indicators. In *IUI 01: Proceedings of the 6th international conference on Intelligent user interfaces*, 33–40.

JOACHIMS, T. (2002): Learning to classify text using support vector machines. Kluwer, Dordrecht.

JOACHIMS, T., FREITAG, D., and MITCHELL, T. (1997): WebWatcher: A tour guide for the world wide web. In *Proceedings of IJCAI97*, 770–775.

KELLY, D. and BELKIN, N. (2004): Display time as implicit feedback: understanding task effects. In *SIGIR 04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 377–384.

KELLY, D. and TEEVAN, J. (2003): Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum, 37(2)*:18–28.

KUSHMERICK, N., WELD, D., and DOORENBOS, R. (1997): Wrapper induction for information extraction. In *Intl. Joint Conference on Artificial Intelligence (IJCAI)*, 729–737.

LANG, K. (1995): NewsWeeder: Learning to filter netnews. *ICML*, 331–339.

MIDDLETON, S. (2001): Interface agents: A review of the field. *Technical Report ECSTR-IAM01-001, University of Southampton.*

MLADENIĆ, D. (2001): Using text learning to help web browsing. In *Proceedings of the 9th International Conference on Human–Computer Interaction.*

MOBASHER, B., COOLEY, R., and SRIVASTAVA, J. (2000): Automatic personalization based on web usage mining. *Communications of the ACM, 43(8)*, 142–151.