Implicit Rating and Filtering

David M. Nichols

Computing Department, Lancaster University,
Lancaster, LA1 4YR, UK

dmn@comp.lancs.ac.uk

Abstract

Social filtering systems that use explicit ratings require a large number of ratings to remain viable. The effort involved for a user to rate a document may outweigh any benefit received, leading to a shortage of ratings. One approach to this problem is to use implicit ratings: where user actions are recorded and a rating is inferred from the recorded data. This paper discusses the costs and benefits of using implicit ratings for information filtering applications.

Introduction

The increasing availability of information in computer-readable form is changing the nature of information searching. The users of information retrieval (IR) systems are faced with two problems: the sheer number of documents and a greater variation in the quality of those documents. The increasing heterogeneity of documents (both in quality, form and media) means that there is a greater need than ever before for tools to aid users in filtering and selecting relevant documents.

Malone *et al.* (1987) describe three forms of information filtering: cognitive (or content), economic and social. Content-based filtering is dominant in IR (e.g. Foltz and Dumais (1992)) – typified by profiles based on keywords. Economic filtering will become increasingly important as digital cash, micro-payments and secure payment technologies emerge from research laboratories onto the Internet. The third form, social filtering, has moved on from the original description (of the importance of the identity of the sender of a message) to several research projects and a few actively-used systems. The social filtering these systems perform is largely based on explicit ratings – where users rate a document on a pre-defined scale.

The rating of resources to enable collaborative (or social) filtering poses several problems: use of appropriate scales, motivation and incentives for evaluators (Avery and Zeckhauser, 1997), biased evaluators (Palme, 1997), avoiding the free-riding problem, achieving a critical mass of users (Oard and Marchionini, 1996) etc. Several of these problems are related to the explicit rating of items.

A small amount of other work has been done on using implicit information (Oard and Marchionini, 1996) - where ratings are automatically inferred from a user's behaviour. This paper will discuss the potential and the problems with using such implicit sources as a basis for filtering and recommending. The evidence for the use of implicit ratings is reviewed and the various types of implicit data available to digital library systems is described.

Implicit and Explicit Ratings

The use of explicit ratings is common in everyday life; ranging from grading students' work to assessing competing consumer goods (see Alton-Scheidl *et al.* (1997) for a review). Although some forms of rating are made in free text form (e.g. book reviews) it is frequently the case that ratings are made on an agreed discrete scale (e.g. star ratings for restaurants, marks out of ten for films etc). Ratings made on these scales allow these judgements to be processed statistically to provide averages, ranges, distributions etc. Implementations of ratings for computerised systems have largely followed this explicit approach.

A central feature of explicit ratings is that the evaluator has to examine the item and assign it a value on the rating scale. This imposes a cognitive cost on the evaluator – this is not necessarily a bad thing; society expects our teachers to think about the grades they give to their students. The value of many forms of rating derives from this intellectual effort and provides the justification for the remuneration that accompanies many rated information streams.

> Expert annotations require effort and have economic value, so the marketplace will undoubtedly assign them a price.
>
> (Oard and Marchionini, 1996)

When explicit ratings are used in social filtering systems (where the ratings of other users are used to generate predictions) the costs and benefits are clearly represented at the interface. The act of rating alters a user's behaviour from their normal pattern of reading - similarly, the choice of which items to examine is altered by providing a rated list. Moreover the benefits of any individual user's ratings are experienced by the other users of the system. This separation of costs and benefits has been noted as being very important in the failure of social computing systems (Grudin, 1994). Unless the user perceives some benefit for participating in the system then they have an incentive for leaving. Even worse, if the link between rating and receiving rated items is not reinforced then users may have an incentive to cease rating but continue to read. In such a system this could result in a lack of any ratings at all (Avery and Zeckhauser, 1997).

The problems for social filtering systems in acquiring explicit ratings have led to speculation that implicit ratings (gathered from user behaviour) may be a solution:

> We believe an ideal solution is to improve the user interface to acquire implicit ratings by watching user behaviors. Implicit ratings include measures of interest such as whether the user read an article and, if so, how much time the user spent reading it.
>
> (Konstan *et al.*, 1997)

The main motivation for using implicit ratings is that it removes the cost to the evaluator of examining and rating the item. Whilst there remains a computational cost in storing and processing the implicit rating data this can be hidden from the user. In a networked environment it is usually difficult for the user to separate network latency from extra application processing. Although there are clearly limits to user tolerances the storage/transport of implicit data at the client end is not a computationally intensive task.

As one of the main problems with obtaining explicit ratings is seen to be the acquisition costs (Oard and Marchionini, 1996) there should be a greater number of implicit ratings. Potentially, every user interaction with a system will generate implicit data – in fact we could move to a situation with too much data rather than the sparseness encountered by explicit rating approaches. Each implicit rating will probably contain less 'value' than an explicit rating but the appropriate cost-benefit trade-off for different types of implicit data will have to be determined empirically.

Acquiring Implicit Ratings

There are several types of implicit data that can, in principle, be captured and studied. (Stevens, 1992) uses three types of implicit data: read/ignored, saved/deleted and replied/not replied. (Morita and Shinoda, 1994) use reading duration in place of the read/ignore attribute. Table 1 shows the result of combining these forms with the types of usage data described in (Nichols, Twidale and Paice, 1997).

Action	Notes
Purchase (Price)	buys item
Assess	evaluates or recommends
Repeated Use (Number)	e.g. multiple check out stamps
Save / Print	saves document to personal storage
Delete	deletes an item
Refer	cites or otherwise refers to item
Reply (Time)	replies to item
Mark	add to a 'marked' or 'interesting' list
Examine / Read (Time)	looks at whole item
Consider (Time)	looks at abstract
Glimpse	sees title / surrogate in list
Associate	returns in search but never glimpses
Query	association of terms from queries

Table 1 Potential types of implicit rating information

Some of the data sources have additional information (e.g. a *Purchase* action has an associated *Price*) - these are indicated in parentheses. The actions are listed in an approximate ordering reflecting the importance of the type of data; it seems reasonable to conclude more from the purchase of an item rather than a simple inspection.

As Digital Libraries (DLs) and the Internet in general become a more commercial environment information providers will increasingly have *Purchase* information available. Elements of this style of investigation into users' purchase patterns are already being undertaken by business who provide 'loyalty cards'. Alongside the ostensible benefits to the customer the supermarket, for instance, gains data about the types and combinations of goods bought by consumers. These patterns can be used to inform marketing activities, e.g. at least one UK supermarket generates money-off vouchers for complementary and substitute goods at the checkout based on the type of goods bought. The extra information available from loyalty cards (or lifetime user IDs in a DL) can only reinforce this trend.

Although we are discussing implicit data it is also possible to gather implicit data from an explicit rating scenario. The *Assess* category distinguishes those events when a evaluator chooses *not* to rate an item when they could have done so. Hence this category would not contain any reference to the actual value of a rating only the fact that a rating had, or had not, occurred.

The *Repeated Use* category in Table 1 could really refer to any of the other categories of data. However, it has an appealing analogue with conventional library practice, that of date stamps in the back of a borrowed book. Items that a user wishes to preserve for some purpose are often *Saved* to personal file space or *Printed*. The *Delete* category will clearly only apply to certain types of information stream (e.g. Usenet News) and differs from the others in that it expresses a negative judgement.

The field of Library & Information Science (LIS) has examined the use of citations in considerable depth. The *Refer* category contains all those instances where one information item links to another item; this includes traditional academic citations as well as less formal links such as hyperlinks on Web pages or the threaded links between Usenet News articles. In some interactive information environments (e.g. Usenet) users can *Reply* to items they encounter; either back to the sender or via a public forum. The *Time* taken to compose this reply may also be available. In many environments a user will *Mark* certain items as being of particular interest so that they can easily return to them, e.g. Web browsers enable hotlists or bookmarks to be recorded.

The next three categories, *Examine*, *Consider* and *Glimpse*, all refer to the same action: the user reading a document (or document surrogate). Systems usually allow users to read a shortened or summary version of a document; bibliographic databases often have an abstract rather than the full-text of an article.

At the bottom of the list in Table 1 the action *Associate* refers to items which are closely connected to those that are examined, e.g. items in the second page of hits which is never reached by the user. The action *Query* refers to query terms which have been used by searchers and can then be reused by subsequent searchers who use related terms (Koenig, 1990).

The collection of these types of implicit data does not pose difficult technological problems: many information access tools could easily be modified to record most of the categories of data in Table 1. In addition, there is a considerable body of research in LIS on the closely related field of transaction log analysis, e.g. (Flaherty, 1993). Data acquired through transaction log analysis has been passed back to designers to refine their systems (typically through user interface modifications). In contrast, an implicit rating system directly accesses the data to modify the system – there is no human in the feedback loop.

Implicit Rating Systems

There appears to have been little work done on implicit ratings, a recent survey (Oard and Marchionini, 1996) mentions only two sources: Morita & Shinoda and Stevens. There are two other major sources: the PHOAKS system (Hill and Terveen, 1996) and GroupLens (Konstan, *et al.*., 1997). The GroupLens project have reported the most interesting results with regard to time-based implicit data; they summarise the situation as:

> Our initial studies show that we can obtain substantially more ratings by using implicit ratings and that predictions based on time spent reading are nearly as accurate as predictions based on explicit numerical ratings. … Our results also provide large-scale confirmation of the work of Morita and Shinoda in finding the relationship between time and rating holds true without regard for the length of the article.
>
> (Konstan, *et al.*., 1997)

Both GroupLens and Morita & Shinoda judge regard time spent reading as a good candidate as a basis for filtering. There is however a difference between the two sets of experiments - the GroupLens data is derived from an explicit rating scenario whereas Morita & Shinoda use post-session rating.

The GroupLens experimental model (Model 1) is:

1. users evaluate news items
2. system collects implicit data
3. compare the explicit and implicit data

From these results they show the similarity of the two sources. The Morita & Shinoda model (Model 2) is:

1. users read news items
2. system collects implicit data
3. system predicts using implicit data
4. users return to items and evaluate
5. compare the explicit and implicit data

Neither model is perfect, in Model 1 it may be premature to use implicit data derived and verified from an explicit rating scenario. User behaviour may be significantly different when they are reading 'normally' - consequently the correlation between implicit data and user judgements may not be as strong or as reliable. In Model 2, the users make their evaluations on their second view of the items - when they have already seen the rest of the items. It seems likely that viewing the other items will alter their assessment of the earlier items.

Indeed, (Oard and Marchionini, 1996) make the general point as to how we can generalise these experimental results to real-world settings where users are distracted and interrupted. The test subjects in the Morita & Shinoda experiment were asked to read items continuously, a very different scenario from their usual news-reading habits.

A separate source of implicit data for Web users are the files of bookmarks, or hotlists, of Web pages; the *Mark* category from Table 1. The Siteseer system (Rucker and Polanco, 1997) uses the overlap between bookmark files to create *virtual communities of interest* and then recommends URLs pages from a users' *virtual neighbours*. The Group Asynchronous Browsing (Wittenburg *et al.*, 1995) uses a similar approach but aims to create an enhanced browsing structure rather than an explicit recommendation.

A simple example of using *Purchase*-based implicit data is currently in operation at Amazon.com (Amazon.com, 1997), the online bookstore. When the entry for a book is displayed other titles bought in conjunction with it are also shown, e.g. *The Design of Everyday Things* by Donald Norman produces:

Check out these titles! Readers who bought The Design of Everyday Things also bought:

• The Visual Display of Quantitative Information; Edward R. Tufte

• Visual Explanations : Images and Quantities, Evidence and Narrative; Edward R. Tufte

• Things That Make Us Smart : Defending Human Attributes in the Age of the Machine; Donald A. Norman

http://www.amazon.com/exec/obidos/ISBN=0385267746/5932-9389921-467955

The other major project using implicit data is PHOAKS; this system scans Usenet postings to find mentions of URLs which it takes as an implicit recommendation. This is an example of the *Refer* category from Table 1. The PHOAKS system has several heuristics to try to eliminate URLs that contain little value, e.g. those contained in signatures. It seems likely that similar 'pruning' techniques will be necessary to use implicit data efficiently.

Rating Scenarios

	Give Explicit Ratings	Give Implicit Ratings	Receive Predictions	Examples
1	–	–	–	normal Usenet reading
2	–	–	✓	freeloader, client
3	✓	–	–	rating service
4	✓	✓	–	rating service
5	✓	✓	✓	GroupLens
6	✓	–	✓	GroupLens
7	–	✓	–	implicit data provider only
8	–	✓	✓	implicit data provider only

Table 2 Scenarios for implicit and explicit rating

Table 2 shows the possible scenarios involving a combination of implicit rating, explicit rating and receiving predictions. Case 1 is the present situation, where most readers do not use ratings and do not receive predictions. In case 2 the user receives the benefit of predictions but does not contribute any ratings; such a user could be a freeloader or a client of a rating service – depending on whether they pay for the predictions. Cases 3 and 4 could describe the behaviour of such a rating service – where predictions are not important.

Cases 5 and 6 describe the users of a social filtering system such as GroupLens – giving ratings and receiving predictions. Case 7 describes the situation where a user allows their implicit data to be gathered but does not receive any predictions. Case 8 describes the scenario where the user does receive those predictions.

Any social filtering system will have users who can be located within these different scenarios but a successful system will have to maintain appropriate ratios between their users. A system with too many freeloaders from case 2 will soon cease to be viable.

Conclusion

The limited evidence available suggests that implicit ratings have great potential but their effectiveness remains unproven. As with many technologies implicit rating may first be combined with existing rating systems to form a hybrid system. One approach is to use implicit data as a check on explicit ratings, e.g. if an evaluator is explicitly rating an item then there should be some corresponding implicit data to confirm that she has actually examined it. If there is no evidence to suggest that the evaluator has examined an item then perhaps their rating should be ignored, or reduced in importance. Conversely, an evaluation with a relatively long 'examine time' may be increased in importance.

The existing systems which capture implicit data (such as Web servers) have generated some concerns amongst the general population of users about privacy. Although we can discuss the possibilities of using implicit data – systems need to be 'socially' accepted in order to be successful. This is especially true of social filtering systems - whose very power comes from a wide take-up of different users.

Acknowledgements

I would like to thank Mike Twidale, Paul Rayson and Chris Paice for stimulating discussions about the topics in this paper.

References

Alton-Scheidl, R., Schumutzer, R., Sint, P.P. and Tscherteu, G. (1997), Voting and rating in Web4Groups, in Alton-Scheidl, R., Schumutzer, R., Sint, P.P. and Tscherteu, G. (Eds.), *Voting, Rating, Annotation: Web4Groups and other projects: approaches and first experiences*, Vienna, Austria: Oldenbourg, 13-103.

Amazon.com (1997), *Amazon.com*, http://www.amazon.com,

Avery, C. and Zeckhauser, R. (1997), Recommender systems for evaluating computer messages, *Communications of the ACM*, 40(3), 88-89.

Flaherty, P. (1993), Transaction logging systems: a descriptive summary, *Library Hi Tech*, 11(2), 67-78.

Foltz, P.W. and Dumais, S.T. (1992), Personalized information delivery: an analysis of information filtering methods, *Communications of the ACM*, 35(12), 51-60.

Grudin, J. (1994), Groupware and social dynamics: eight challenges for developers, *Communications of the ACM*, 37(1), 92-105.

Hill, W. and Terveen, L. (1996), Using frequency-of-mention in public conversations for social filtering, *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'96)*, Cambridge, MA, ACM Press, 106-12.

Koenig, M.E.D. (1990), Linking library users: a culture change in librarianship, *American Libraries*, 21(9), 844-9.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J. (1997), Applying collaborative filtering to usenet news, *Communications of the ACM*, 40(3), 77-87.

Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, M.D. (1987), Intelligent information sharing systems, *Communications of the ACM*, 30(5), 390-402.

Morita, M. and Shinoda, Y. (1994), Information filtering based on user behaviour analysis and best match text retrieval, *Proceedings of the 17th ACM Annual International Conference on Research and Development in Information Retrieval (SIGIR'94)*, Dublin, Ireland, Springer-Verlag, 272-81.

Nichols, D.M., Twidale, M.B. and Paice, C.D. (1997), *Recommendation and Usage in the Digital Library*, Technical Report CSEG/2/97, Computing Department, Lancaster University.

Oard, D.W. and Marchionini, G. (1996), *A Conceptual Framework for Text Filtering*, Technical Report CAR-TR-830, Human Computer Interaction Laboratory, University of Maryland at College Park.

Palme, J. (1997), Choices in the Implementation of Rating, in Alton-Scheidl, R., Schumutzer, R., Sint, P.P. and Tscherteu, G. (Eds.), *Voting, Rating, Annotation: Web4Groups and other projects: approaches and first experiences*, Vienna, Austria: Oldenbourg, 147-62.

Rucker, J. and Polanco, M.J. (1997), Personalized navigation for the web, *Communications of the ACM*, 40(3), 73-5.

Stevens, C. (1992), *Knowledge-Based Assistance for Accessing Large, Poorly Structured Information Spaces*, PhD Thesis, Department of Computer Science, University of Colorado.

Wittenburg, K., Das, D., Hill, W.C. and Stead, L. (1995), Group asynchronous browsing on the World Wide Web, *Proceedings of the Fourth International World Wide Web Conference*, Boston, MA, O'Reilly & Associates, 51-62.

Cite as:

Nichols, D.M. (1998) Implicit rating and filtering, *Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering*, Budapest, Hungary: ERCIM. 31-36.

(note: the Workshop was held in November 1997 but the Proceedings were not published until 1998)