STUDIA INFORMATICA
Nr 1(12) Systems and information technology 2009

Approaches to "Cold-Start" in recommender systems

Mieczysław A. Kłopotek[1,2]
[1] Institute of Computer Science, University of Podlasie, ul. Sienkiewicz 51, 08-110 Siedlce, Poland
[2] Institute of Computer Science, Polish Academy of Sciences, ul. Ordona 21, 01-237 Warsaw, Poland

Abstract. The paper explores the possibilities of handling cold start problems for recommenders associated with document-map based search engines.

Keywords. Recommender, cold start, document map

1 Introduction

Recommender systems can be viewed as a kind of decision support systems by subjects (human or software agents) that have to act under restricted resources (limited knowledge, limited processing capabilities, limited memory etc.). They are considered as an important way to take business advantage via providing potential buyers with proper information. Therefore much research is being done over years to improve the quality of recommendation.

A recommender, from the business point of view, may be considered as a type of advertisement. However, it is applied differently. Usually, a recommender is applied while the user is searching for something, either explicitly, or implicitly, and would welcome any help.

A typical recommending engine works as follows: (1) Identify items interesting for the user, (2) Discard not useful ones (as possessed already, too similar to earlier offers, or too expensive etc.), (3) Recommend a selected one (possibly with justification)

Usually, a recommender system integrates machine learning, decision support, information filtering and profiling subsystems. As the machine learning component plays always an important role, there is a grieving problem of lack of training data when massively new items are to be recommended.

A recommender is a must of modern commercial sales systems. But applications are known for identification of social relationships, of potential HTTP request (improvement of server performance), adaptive web navigation (suggesting sites/links of interest, based on user's earlier visits or collective behaviour), for assisting new users by modelling a Web site via really used links, suggestions of whole tours of Web sites etc.

Two major lines of recommender systems are those that are content-based and those that are society (usage) based. The first brand takes whatever descriptive information about the item is available to position the item against other items recommended earlier. The second brand does not require any item description as the usage by similar customers is used to predict item satisfaction.

The recommender system main part is frequently a machine learning component. The grieving issue for its usage is the lack of training data when massively new items are to be recommended While the second approach suffers heavily from the lack of item-related information while a new item (line) is introduced (and hence no meaningful recommendation can be done), the first approach usually also can be driven astray in case of lack of usage information. This kind of problems is referred to as "cold start problem".

This can severely affect the operation of the business model of the recommender. An advertisement recommender system may be for example asked to present the advertisements in such a way as to earn as much money as possible advertisement pays off if it is clicked. Missing information, leading to failed recommendations, may cause the system not to match fundamental economic constraints.

In this paper we take a closer look at some promising approaches in the literature to handle the "cold-start" problem.

2 Issues in cold-start and solution proposals

We are facing usually several types of cold-start problems:

- New user problem – occurs when new users do not have any ratings or purchases in their profile.
- New item problem – the system is able to recommend an item, until a considerable number of users has rated the new item.
- Non-transitive association – poor and over specific recommendations may occur when two items are never rated by same user or a user has different preferences from the rest of population
- Overfitting – occurs, when common information about users is used, while differences between them are ignored

In the past, various ways of handling these issues have been proposed, which may be split into the following categories:

- Merge content and usage information
- Create pseudo-usage-profiles for new items
- Manually cluster new items into existent categories
- Dimensionality reduction [LSI, etc.]
- Hidden variable model,

- Item-based versus user based usage investigation
- Attribute-aware item / user descriptions
- Optimization approaches

Merging content and usage may be realized in a number of different ways. In the so-called content-boosted collaborative filtering, collaborative filtering is applied to a modified data set. A data mining process is carried out to create “rules” predicting ratings of items based on content information. Then for the items, where the usage data is missing, missing ratings are derived from the content according to the learned rules. While the usage experience from real users is collected, the artificial ratings are over-weighted by them with system , moving towards traditional collaborative filtering.

Creating pseudo-usage-profiles for new items works in a similar way, but the pseudo-profiles are derived by experts and they are assigned to new items also by skilled employees.

Another way to handle the issue is to have clusters of items in use (either done manually or automatically) and to assign manually the new items to existing clusters. In this case the initial pseudo-usage profile of an item would be the cluster average (cluster center).

Dimensionality reduction (based on techniques like Principal Component Analysis/PCA, Latent Sematic Indexing/LSI, Random Projection, Non-negative Matricx Decomposition and other) helps to handle the cold start in that fewer parameters need to be learned from data, so that with the same amount of information more reliable estimates can be performed.

Closely related is the probabilistic hidden variable model (similar to PLSA). A Bayesian network model: U->Z->I is assumed, implying satisfaction probability of the form: P(u,i)=SUM(z): P(u)*P(z|u)*P(i|z), where U – user, I – item, Z – topical area (hidden variable). The model can be enhanced with content information C, yielding P(u,c)=SUM(z): P(u)*P(z|u)*P(c|z). Expectation maximization (EM) models or naive Bayes approaches are used to compute the probability estimates.

Another way to mitigate the large areas of missing data is to consider item-based usage analysis instead of user based one. A typical sales enterprise has considerably more users than products offered, so that the items have considerably more users, that acquired them, than there are items purchased by an individual user. So instead of asking “which users, that purchased this product, are similar to this user?”, we can ask “how similar is this item in terms of purchase behavior to the items that this user already acquired?”.

Attribute-aware item /user descriptions is an attempt to escape the missing data problem on an individual item by considering items, users, or both as instances of obkjects from populations, where the individual behaviour is driven by features characterizing them (and not by “personal” characteristics). So any recommendation rules can be derived and applied at the level of attributes, instead of the item/user level.

Optimization approaches, that will be more closely described in the next section, may be considered as “active learning” methods. They attempt to balance exploration of the realm, learning and application of gained knowledge.

All of these approaches are applicable both to new items and new users problems.

3 Exploration versus exploitation

The issue that is of interest for anybody running advertisement business with a search engine is to match the fundamental economic constraint: present the advertisements in such a way as to earn as much money as possible. The advertisement pays off if it is clicked.

The problem as such is not new. A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani (2005) handled the problem of choice of advertisements under known click-through rates. Estimation of click-through rates themselves is a difficult problem, some work has been done by O. Madani and D. Decoste (2005) and N. Immorlica, K. Jain, M. Mahdian, and K. Talwar (2005). Given the many advertisement choices one has at any point of time, you run into an issue of optimizing the gains. N. Abe and A. Nakamura (1999) and P. Rusmevichientong and D. Williamson (2006) investigated this problem to some extent.

The most interesting approach to the issue seems to be the method of S. Pandey, C. Olston (2006). They reformulated the issue in terms of a multi-armed bandit problem (D. A. Berry and B. Fristedt 1985). Their approach is to seek a balance between exploration and exploitation. If an advertisement is not known to the public, it is promoted for showing even if the probability of bringing the customer satisfaction is low. The bigger experience the higher the certainty about the click-through rate and the more rational is the usage of any optimization technique to choose better earning presentation. The authors study additionally various kinds of budget constraints.

Note that the Dempster-Shafer Evidence theory may be perfectly applied under these circumstances. Click-through-probabilities may be amended with weight assigned to the universe, and decreasing universal weight may be treated as indicator of well-learned probabilities

4 BEATCA PRO Concepts

The methodology of Pandey and Olston, while in its original form suffering again from the scaling issues, seems to be quite attractive in case of recommendations related to document-map based approach, followed by our search engine BEATCA PRO (Kłopotek et al., 2007).

A map of document collection consists of a two-dimensional set of cells, to which documents are assigned in such a way that closeness on the map reflects closeness of the content. The map-based approach to search engine interfacing comprises two important features from the point of view of the target recommendation system: providing an overview over the whole collection of objects, and a very detailed clustering into groups of objects and their immediate (local) contexts.

With a strongly parameterized map creation process, the user of BEATCA can accommodate map generation to his particular needs, or even generate multiple maps covering different aspects of document collection. The overall complexity of the map creation process, resulting in long run times, as well as the need to avoid "revolutionary" changes of the image of the whole document collection, require an incremental process of accommodation of new incoming documents into the collection.

Within the BEATCA project we have devoted much effort to enable such a gradual growth. We investigated vertical (new topics) and horizontal (new documents on current topics) growth of document collection and its effects on the map formation capability of the system. To ensure intrinsic incremental formation of the map, all the computation-intense stages involved in the process of map formation (crawling, indexing, GNG clustering, SOM clustering) need to be reformulated in terms of incremental growth.

In general, our goal is to develop a recommender systems, based on combined paradigms of content-based filtering as well as on collective filtering principle, to exploit possible synergic effects. Such effects may emerge when:

- both methodologies are merged,
- system is able to model joint, integrated recommendation of passive and active objects (i.e. clients and products), and not only passive objects pointed by active ones,
- recommendations are based on visual system, which helps to explain and justify a recommendation.

Application of joint methodology is possible if available data contain information on recommended objects as well as relations between recommended and recommending objects. Such information is present, e.g. in WWW documents, where individual html pages have not only textual context, but also hyperlinks between them. From logs saved on a particular host one can obtain so-called click-stream of users surfing from one page to another, and some additional data such as voluntarily filled-in questionnaires. Among other examples are libraries, book stores, or any shop (including e-shops), where products can be described by a set of attributes (e.g. advertisement leaflet) and users can be identified by some ID cards (e.g. loyalty program participation cards). Similarly, for some services (e.g. concerning education or health), both pointed(passive) and pointing(active) objects are described by attributes.

By an *integrated recommendation* we mean recommendation such as “People interested in <characteristics of people> are buying also book <title>” (instead of typical recommendation in form: “People interested in <title> are buying also book <title>”). Thus, integrated recommendation requires that system has an ability to generalize features describing characteristics of active objects (i.e.users or clients).

Recommendation with a visual explanation and justification is a completely new approach, based on creation of two-dimensional, navigational map of objects. Such a map yields a possibility to present an identified area of user's interests together with surrounding context, i.e. main directions of his/her future activities.

Note that for a number of reasons, users of search engines require supporting tools like recommender systems. Lists of documents are usually, too long, documents of interest appear too far away by click distance. One may be tempted to improve search engines to adapt to potential user needs. But this is in our opinion the wrong way to go. Search machinery should be kept universal and not adapting too quickly to user's needs, as the search process should be repeatable (the user may want to find something later again or may have apruptly different needs). Instead, an additional recommending subsystem should be added, reacting to user profile more promptly, and the user may, or may not use the recommendation results.

So from our point of view each document has its value for the user, either in the context of his query, and / or of his profile and / or of his history (documents viewed recently may be of no value). On the other hand each document has its chance to be seen by the user (e.g. top documents on the retrieved list, documents in a short clicking distance from them). We suggest to recommend those with high residual value (value minus chance to be seen).

On the other hand, a useful recommendation would be one that forecasts the interest of the user to come. When having sequences of searches (search engine usage statistics) then one can predict from one query what will be the next one. The recommendation can be based on retrieving documents before hand.

Of course the detailed prediction of next query would be possibly error-prone, so that detection of the next context of interest may be more interesting (map, cell, neural gas neighbour etc.).

5 BEATCA PRO Cold Start Issues

So we have to do hear also with a kind of multi-armed bandit problem, where one of possible ways of promoting documents could be used. Each one requires, however, a learning phase. The more complicated problem, the longer the learning phase may be.

The advantage of the document collection map is that we can manage the overall learning complexity, splitting the learning process into9 less complex sub-problems.

The first issue in document recommendation is the complexity of document description: a document is in fact a feature vector with thousands of dimensions (terms), which usually occur sparsely. With document collection multiple maps we can add a couple of new attributes ("attribute-aware training") and relationships, like common / different map [fuzzy] membership, common/different context [fuzzy] membership, common / different map cell [fuzzy] membership. With our histogram-based document group representation (Ciesielski and Kłopotek, 2007), we can further say if a document is typical or non-typical for its group Last not least groups of documents are labelled with keywords, which may be treated as special attributes.

The contexts themselves provide with another way of handling learning scale. A local context, that is the group of similar documents (Ciesielski et al. 2007), may make a number of terms irrelevant, as common or rare in a given context, so that a significant reduction in the dimensionality is achieved.

Last not least only typical documents may be first considered in the training process, just ignoring the non-typical ones.

Typicity of a web page, of the query accountable for within a context (not just similarity).

Another simplification in the learning process is that with insufficient data, one can refrain from recommending a document and recommend something easier: either a map, or a context, or a map cell, or a query expansion, or a document author (domain etc.). Then, if document recommendation is necessary, one can revert to the notion of typicity. Generally, one can apply the principle that the less we know about the user, the less detailed recommendation is targeted at.

In a symmetric way, we can profile the user at different levels of conceptual similarity. Instead of targeting a document, we can create a user profile at the level of choice between maps, between contexts on a sinle global map, between cells with in a context, between various document authors (domains etc.), or even observing whether the user tends to choose typical or non-typical documents. With a clustered query stream, the profiling of a user against the query stream is also an option.

6 Final remarks

The methodology of Pandey and Olston of exploration and exploitation seems to be a promising way to handle the cold-start problem of recommendation systems. It can be quite surely used at any place where we can define an optimization problem with clear target function.

With search engines in general, we have a simplified task to execute because the users are exploring out of their own interest. So we need to trace their activities only.

With document map based search engines a structure on the space of documents is imposed which may be exploited to accelerate the learning phase for recommender systems and allows for mixing of exploration and exploitation at different levels of detail of information about user interests.

Acknowledgement

This research has been partially financed from the science budget for the years 2007-2009 as research Project N516 015 32/1906 ***"Rosnąca mapa przestrzeni obiektów jako nowy paradygmat dla systemów rekomendujących".***

References

1. N. Abe and A. Nakamura. Learning to Optimally Schedule Internet Banner Advertisements. In ICML, 1999.

2. D.A. Berry and B. Fristedt. Bandit Problems: Sequential Allocation of Experiments. Chapman and Hall, London, 1985.
3. K. Ciesielski and M. Kłopotek: Towards Adaptive Web Mining: Histograms and Contexts in Text Data Clustering . IN: M.R. Berthold and J. Shawe-Taylor *Intelligent Data Analysis (IDA-2007)*., Springer-Verlag, LNCS 4723, Ljubljana, Slovenia 2007, 284-295
4. K. Ciesielski and M.A. Kłopotek and S. T. Wierzchoń: Contextual adaptive clustering with personalisation. IN: Z. Raś and D. Zighed and S. TsumatoProc. 3rd Int. ECML/PKDD Workshop on Mining Complex Data., Warszawa, Poland 2007, 128-138,
5. N. Immorlica, K. Jain, M. Mahdian, and K. Talwar. Click Fraud Resistant Methods for Learning Click-Through Rates. In WINE, 2005.
6. M. Kłopotek and S. Wierzchoń and K. Ciesielski and M. Dramiński and D. Czerski: *Conceptual Maps of Document Collections in Internet and Intranet. Coping with the Technological Challenge*. IPI PAN Publishing House, Warszawa, Poland 2007.
7. O. Madani and D. Decoste. Contextual Recommender Problems. In Proceedings of the 1st International Workshop on Utility-based Data Mining, 2005.
8. A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. AdWords and Generalized On-line Matching. In FOCS, 2005.
9. S. Pandey, C. Olston: Handling Advertisements of Unknown Quality in Search Advertising. NIPS 2006.
10. P. Rusmevichientong and D. Williamson. An Adaptive Algorithm for Selecting Profitable Keywords for Search-Based Advertising Services. In EC, 2006.