

Research Resources for Recommender Systems

Joseph A. Konstan and John Riedl

GroupLens Research Project
Department of Computer Science and Engineering
University of Minnesota
{konstan,riedl}@cs.umn.edu

Introduction

The recommender system research community has become a well-identified and active research community built around the goal of understanding, building, and improving recommender systems. A recurring theme since the first workshop in Berkeley (March 1996) has been the development of substantial research infrastructure—typically publicly-accessible recommendation services—as a tool to conduct research. The result has been a variety of innovative recommender systems, including the Bellcore Video Recommender, M.I.T.'s Ringo system, Stanford's Fab, and our own GroupLens project. This plethora of systems helped the field achieve wider visibility, particularly through Varian and Resnick's special issue of *Communications of the ACM* (March, 1997), and has spawned several commercial recommender system projects and companies.

Reflecting back on the Berkeley workshop and the AAI workshop in Madison (July 1998), we see a structural challenge facing recommender systems researchers. Conducting an experiment to test a hypothesis relating to recommender systems requires too much experimental set-up and too much lag time. Despite the research and commercial successes in the field, we still lack the technology, tools, testbeds and data sets needed to efficiently carry out research. The regular messages on the collaborative filtering mailing list seeking research recommender engines (as opposed to commercial ones which are typically too expensive and too inflexible) and the remarkable usage of the EachMovie data archive (the only widely available data set collected from a running recommender system) attest to the unfilled demand for research infrastructure.

We propose to assemble, build, and maintain this infrastructure for the recommender system research community. This position paper reviews the research challenge, outlines our initial vision of such a research resource, identifies the components that we currently have implemented or under development, and suggests steps for realizing this resource.

Note to organizers and workshop participants. Our goal in submitting this position paper is to solicit feedback from the community and evolve a community consensus about the role, contents, and structure of the research resource. We would propose allocating a block of time during the workshop for brainstorming, and a second block of time for integrating and prioritizing ideas and organizing people to start the process. Ideally, we'd like a block of an hour or more on Sunday to present the idea briefly and brainstorm, and then some additional time on Monday to discuss the results after we've had an opportunity to synthesize them and distribute them to participants.

Challenges in Performing Innovative Recommender System Research

To simplify the discussion of research resources, we divide recommender system research into two categories:

- **Off-line research** seeks to learn from data that has already been collected. The EachMovie data set, in particular, has been used for substantial off-line research into recommendation algorithms. Other types of off-line research include retrospective studies of system usage, such as our own look at the start-up behavior of EachMovie users (in comparison to the start-up behavior of our MovieLens users).
- **On-line research** seeks to learn from experimentation with and/or observation of a live recommender system. For example, a researcher trying to determine whether users are happier with a recommender system that provides seven vs. five rating points, might choose to carry out

an experiment in which users are given different interfaces, their behaviors are observed, and perhaps they are interviewed. Similarly, most research on applying recommendation technology to new applications or communities, and most research on radically new interfaces such as community awareness interfaces, requires on-line experimentation.

We recognize, of course, that this taxonomy is imperfect and incomplete. We do find it useful, however, for thinking about research resources.

The biggest challenge for off-line research is the availability of data sets. At present, the recommender systems community has a single widely-available data set—the EachMovie data set. (The machine learning database at U.C. Irvine has extensive data collections, but they are largely focused on machine learning in domains other than opinion and have no information on observational ratings.) For much off-line research, the next challenge is the availability of recommendation engines that are flexible enough to customize for specific experiments. The remaining challenge is to develop a set of experiment control and data analysis tools, designed to generate community-accepted quality and performance metrics, that can be used to efficiently test hypotheses.

On-line research depends more on users than on data. The single most valuable resource for most researchers would be an already-existing recommender system complete with a set of active users ready and willing to take part in experiments. Where this is not feasible, construction of such a site would be greatly eased by the availability of a recommendation engine and of site creation and customization tools.

Conclusion: We need a community research resource

The recommender systems research community is being held back by the lack of research resources. We propose that the GroupLens Research group at the University of Minnesota take on the task of organizing such a resource. In the rest of this paper we describe some of the resource features we envision, the steps we've taken so far to develop these resources for our own use, and ideas about creating the community resource. We recognize that two things are critical to this resource:

Cooperation. We can take the lead in organizing this work, but we need four things from the research community. First, we need your ideas. We hope to use this workshop, in part, to start a community-wide dialogue on the content, structure, and management of such a resource. Second, we need your active participation. We need active and visible members of the research community to help direct and promote this effort. We will discuss formal roles such as a steering committee, but also need people to undertake equally important informal roles. Third, we need your data. It is central to the notion of a community resource that researchers with data allow that data to be stored in or referred to by the resource. We recognize that release of some data sets may need to be delayed, and that others may need to be sanitized before publication, but the collection of data is critical. We also include the collection of bibliographies, link lists, and other research tools. Fourth, we need your support. As we attempt to secure funding to make this resource a reality, we need the support of the recommender systems community.

Funding. Creation and maintenance of this resource will require staff and computing resources. We intend to seek funding from Government sources, as well as from interested research organizations. The support of the research community will be instrumental in demonstrating the justification for funding this resource.

An Envisioned Community Research Resource

Data Sets

We envision the resource hosting a wide variety of recommender system data sets, including ratings data, interest measures (e.g., click or revisit counts), and usage traces. We hope many of these data sets will be freely available for research use. Others, such as the EachMovie data set, may have restrictions placed on their use, but still be widely available. The resource would link to the owner's site for restricted data sets.

Finally, some data sets may be sanitized before contribution, since they would otherwise contain personal data.

In addition to raw data sets, it will be useful to maintain a set of benchmark data sets that have been well-characterized to help support comparative research. These benchmark data sets may be extracts from larger data sets.

Thus far, the GroupLens Research project has an ongoing MovieLens data set, with both ratings data and usage history. We also have several carefully extracted subsets of data from MovieLens and from the EachMovie data sets that have proven useful for algorithm research. It will be necessary to negotiate with Compaq to freely distribute EachMovie extracts, but other data sets can be made publicly available.

By combining our current data sets with contributed data sets from the community, we can better support not only off-line research but also comparative research.

Recommendation Engine

To support both off-line retrospective analysis and exploratory on-line experiments, the community needs a flexible, customizable recommendation engine. Commercial recommendation engines are well beyond the budget of most researchers, and they often depend on undocumented and unchangeable algorithms optimized for performance rather than control. As a trade-off, researchers are able to accept much lower performance, since our projects rarely experience hundreds or thousands of simultaneous uses. In essence, we need cheap and flexible, and we can sacrifice low latency and high throughput.

Such an engine could be made available in two forms: as a service that could accept jobs remotely (or serve live requests remotely), or as a software package that could be installed at the researcher's site. In either case, there would need to be easy-to-use interfaces for customizing the engine and for using it.

The GroupLens Research project has developed a small research collaborative filtering recommendation engine on top of the Oracle DBMS. We have been using this engine for the past six months as part of our off-line recommender system research, and are just starting small on-line experiments with it. One elegant feature of DBLens, as we've termed it, is that any or all parts of the recommender system algorithm can be reprogrammed using PL/SQL. We are investigating ways to make DBLens suitable and available for wider use.

Analysis Tools and Metrics

The recommender systems research community has explored a variety of metrics for evaluating and comparing recommender system performance. Among those metrics used in prior work are: prediction accuracy metrics (e.g., mean absolute error, root mean squared error), decision-support metrics (e.g., ROC sensitivity, precision/recall performance, reversal rates), coverage measures, and others. There are also problems for which metrics and benchmarks are still being defined, including task-performance metrics for users working together with recommender systems (as opposed to recommender system performance by itself).

The community is properly evolving its experimental design and metrics through workshops, publication, and review. The current proliferation of metrics, however, places a burden on researchers. Many of these metrics are not computed by common statistical packages, and many researchers are unfamiliar with statistical tests using them. Accordingly, we expect it would be useful to have a reusable set of scripts for analyzing data and generating metrics. It would be even more useful to have scripts that generate graphs (or data that can be easily plotted) for comparison, analysis, and publication.

The GroupLens Research group has been developing such analysis tools for our own use. We have a set of scripts that take data in our format and generate a wide range of metrics. These tools, and other similar tools within the community, can be collected, documented, and converted to use common file formats. The community resource would take responsibility for validating contributed tools and for developing tools that are not already in existence.

Making Recommender Systems Available for On-Line Research

One exciting opportunity for a community resource is the maintenance of live recommender systems that can be used by researchers to conduct on-line experiments. If a set of well-trafficked recommender systems were available, researchers could "borrow" a set of users and divert them to an experimental version of the system. The nature of the borrowing (e.g., specific volunteering, general volunteering, etc.) and the form of borrowing (e.g., user impressions vs. user for time period, etc.) would need to be determined based on the integrity of the system and the needs of the research.

To make user-borrowing a possibility, the community would need a set of recommender systems, mechanisms for diverting certain users to experimental systems, standards of data collection, and mechanisms to ensure the safety of human subjects and the long-term integrity of the site. While this experimental model needs refinement, it presents an exciting opportunity that is enabled by the Internet and the web.

While the GroupLens Research group does not yet have infrastructure for diverting users to experimental sites, we are developing this infrastructure to facilitate our MovieLens research. We can envision maintaining a collection of well-used recommender systems (not necessarily all housed centrally) on which researchers could experiment. A steering committee could set guidelines for research usage (including institutional review) and could manage the integrity of the site.

Other Resources

Other recommender systems resources currently exist. We do not propose replacing these, rather we propose to link to these resources or to host them, depending on the preference of the maintainer. Among other resources of interest are: lists of recommender systems, recommender systems bibliographies, directories of researchers, calls for papers, and e-mail lists.

Steps Forward

We see this workshop as the first step in a process. Our first goal is to obtain feedback from and brainstorm ideas with the recommender systems research community. Depending on the feedback we receive, we anticipate the following steps:

- Short Term.** Identify existing resources, continue development of our resources, identify interested researchers (possible steering committee), develop more detailed proposal/plan.
- Medium Term.** Seek funding to develop/maintain resource (government and/or industry), form steering committee, develop web site with resource collection.
- Long Term.** Document and maintain collection, hold periodic training sessions and workshops, facilitate low-overhead research (students should be able to experiment with RS analysis in course projects), grow the RS field.