Towards an Experimental News User Community as Infrastructure for Recommendation Research

Joseph A. Konstan [1], Robin Burke [2] and Edward C. Malthouse [3]

[1] *University of Minnesota, 200 Union Street SE, 4-192 KHKH, Minneapolis, MN, USA*
[2] *University of Colorado, 1045 18th Street, UCB 315, Boulder, CO, USA*
[3] *Northwestern University, 1845 Sheridan Road, Evanston, IL, USA*

Abstract

While substantial advances have been made in recommender systems -- both in general and for news -- using datasets, offline analyses, and one-shot experiments, longitudinal studies of real users remain the gold standard, and the only way to effectively measure the impact of recommender system designs (algorithmic and otherwise) on long-term user experience and behavior. While such infrastructure exists for studies within some individual organizations, the extensive cost and effort to build the systems, content streams, and user base make it prohibitive for most researchers to conduct such studies.

We propose to develop shared research infrastructure for the research community, and have received funding to gather community input on requirements, resources, and research goals for such an infrastructure. If the full infrastructure proposal is funded, it would result in recruiting a community of thousands of users who agree to use a news delivery application within which various researchers would be install and conduct experiments. In this short paper we outline what we have heard and learned so far and present a set of questions to be directed to INRA attendees to gather their feedback at the workshop.

Keywords
Recommender Systems, Research Infrastructure, News Recommendation

1. Introduction

The field of Recommender Systems has made great strides in its first 25 years (since the 1994 publication of the first paper reporting on an automated collaborative filtering system). These personalization tools are now an integral part of commerce, information dissemination, education, and myriad other applications. Algorithmic improvements have moved the field from nearest-neighbor correlational algorithms to matrix-based latent factor methods, optimization techniques, neural network approaches, and online learning techniques that can recommend more efficiently and effectively with reduced data densities. And the prevalence of large datasets has facilitated offline algorithmic research and helped grow the field substantially.

At the same time, the field of Recommender Systems research is struggling today in ways we could not have anticipated two decades ago. The early days of recommender systems were innovative ones where a large fraction of researchers built systems and experimented by delivering predictions or recommendations to users. In that era we saw substantial advances in areas such as interaction models for recommendation, interfaces for eliciting preferences, evaluating diverse recommendation objectives, and other human-centered questions around recommendation. Today only very few research recommender systems remain (e.g., the 22-year-old MovieLens system, which is both narrow and not

INRA'21: 9th International Workshop on News Recommendation and Analytics, September 25, 2021, Amsterdam, Netherlands
EMAIL: konstan@umn.edu (A. 1); Robin.Burke@colorado.edu (A. 2); ecm@northwestern.edu (A. 3)
ORCID: 0000-0002-7788-2748 (A. 1); 0000-0001-5766-6434 (A. 2); 0000-0001-7077-0172 (A. 3)


open to broader research community), and the effort required to build new ones is beyond the scope that can be supported by a single team or a single research grant.

This lack of infrastructure is hurting the field substantially. The vast majority of recommender systems research papers and projects no longer can address the hard questions of interaction, long-term usage, and impact on humans. Instead, the field presents a wide range of algorithmic advances, nearly all tested on the same sets of collected offline data. We don't disparage offline evaluation, but we note that it has a particularly salient limitation in a field like recommender systems where the goal of these systems is usually to present users with attractive options they would not have found themselves, not merely to "recover missing data" or "recommend what users would have discovered on their own."

Recommender systems have gained new importance as they serve as the engines determining the news and information people discover through news aggregators such as Google News or social media feeds such as Facebook or Twitter. Those of us who helped invent these technologies very much want to help ensure that they can be harnessed to help guide news consumers not only to "what they like" but to a meaningful understanding of the world around them.

2. The Domain: News Aggregation

We have chosen the domain of news aggregation--services that gather news from many sources and then, through a recommender system and interface, present that news to individuals with some degree of personalization. We find news aggregation to be a particularly apt domain of study for five reasons--both technical and related to broader impact:

News supports recommendation based on a full range of content analysis, metadata-analysis, and collaborative approaches while also having high temporal and contextual relevance. Today's interesting recommender systems questions often bring together multiple forms of analysis to understand user needs and preferences, to analyze items in relation to those needs and preferences, and to match recommendations to appropriate opportunities. News supports the full range of these approaches and therefore in turn helps bring together the content-processing (NLP and IR) parts of the community with the traditional (explicit or implicit) rating-based collaborative filtering parts.

News has a complex popularity and consumption curve based on locality, interests, and other factors. We often make simplifying assumptions about a Zipf distribution of popularity for items in a community of users (e.g., most movies are seen by very few users, but a few such as Star Wars or Titanic are blockbusters). News-reading has several layers of complexity that make it an interesting domain for recommendation and content-processing algorithms. News has locality—an article on a mayoral debate or park plan in Boulder, CO may enjoy wide readership locally, but few readers in Amsterdam, Beijing, or Chicago are likely to be interested. It has topicality—American football fans, and perhaps even many sports fans, likely follow the college football playoffs, but few others may have (other than those connected to schools involved). It has temporality and even perishability in that stories can become uninteresting or obsolete quickly. It also has multiple sources and different perspectives which may mean that many people are reading articles about immigration or politics, but they may be getting wildly different content with little common ground. These properties make news substantially more complicated than well-studied domains such as movies and music, and in turn make it an excellent domain for future algorithm and interface experimentation.

News has long-term use integrated into its users daily lives. One of the appealing factors with news today is how well-integrated its consumption is into the lives of its consumers. The era where news readers would subscribe to one or more daily papers, reading them in the morning (or in some cities, also in the afternoon) is mostly gone. Instead users integrate news into their day with feeds on their mobile phones, tablets, and computers. News reading combines recommendation, explicit profiles, search, and other elements into an activity that requires content and interaction that fulfill without crowding out the rest of a user's activities.

News recommendation blends elements of learning preferences with elements of shaping them. One of the interesting challenges for recommender systems, and for computing technology in society more broadly, is to strike a balance between simply giving people what they want today and helping them realize

a broader goal (and perhaps a greater good). News exhibits this challenge at all levels. News publishers struggle between the costly challenge of providing important, informative content (that may not attract nearly enough attention to warrant its cost) and providing click-bait articles or "fluff." Aggregators have the same challenge--to what extent do they insist on presenting important news about war, politics, or the economy? To what extent do they accept and honor reader preferences for sports or celebrity information only? And how do they apply the same concepts to points of view? All of the major commercial aggregators have structures that attempt to balance topics of interest with ones of importance (a concept that goes back to the Krakatoa Chronicle, the first Web-based newspaper prototype). This challenge supports a wide range of research we feel is both intellectually interesting and socially important.

News consumption, recommendation, concerns about filter bubbles and echo chambers are a socially-important problem in many societies today. Stated simply, news today is a political battleground (though news has a long history of being political and partisan). At times it is difficult to recognize that sources from different "sides" are even covering the same story. Without suggesting that we have any solution to this challenge, we believe that thoughtful recommender systems research can help provide tools that can help. We believe that research questions about credibility and trust, awareness of multiple points of view, transparency, agency, and other key issues can not only provide interesting research results, but important ones for society.

3. Envisioned Research Infrastructure

The envisioned infrastructure is new, though it will integrate with and incorporate some existing software elements that have been developed in prior research. At a high level, the infrastructure is a recommender-driven news aggregator application, used by thousands of consenting users, and available for experimentation by academic and non-profit researchers. We divide this infrastructure into four components:

News Aggregator System and Application. This first component includes the software and server systems needed to gather news metadata from diverse publishers (e.g., titles, links, indexing data), to build and maintain user profiles, to customize news delivery to users based on those profiles, and application interfaces to deliver and read news (including on mobile devices), to sign up and initialize profiles, and to carry out other user-centered tasks. We are fortunate in that the basic structures of both aggregation and delivery have been shown to work in both research and commercial systems; our challenge is to build a sufficiently flexible system as simply as possible to support the rest of the infrastructure.

Experiment-Support Infrastructure. In addition to the system itself, we need a set of tools to manage and implement experiments (closely integrated with the above). These include mechanisms to implement experimental conditions (i.e., to override algorithms and/or interfaces), mechanisms to assign users to various experimental conditions for particular durations, mechanisms to log both what is presented to users and their subsequent actions, mechanisms to deliver surveys to users at appropriate times, and mechanisms to perform analyses for researchers comparing different experimental and control groups on various metrics. In addition to software, there is an issue of human subjects review, and developing templates and agreements to simplify the review of each experiment (so that whenever possible an experiment can be reviewed by a single institution's IRB).

Community of Users. The most unusual and important need within this infrastructure is the recruitment and retention of a community of users for this news recommender. We still need to determine whether this will be a small set of regional user bases (to achieve greater local density) or a broader global base, but our goal is to maintain an active user base of at least 2000 users—a sufficient number to support dozens of experiments per year—and to grow that user base as the use of the infrastructure increases. Part of the design stage also includes designing a base level of informed consent (for non-deceptive studies and for baseline measurements) to facilitate relevant research.

Community Governance Infrastructure. As a community resource, this infrastructure also needs a governance model including controlling key issues such as: (a) review and approval of requests to allocate

users to experiments; (b) prioritization of enhancements and new developments; and (c) development of a business model to make the resource sustainable for research after grant funding support concludes.

4. Project Status, Request for Input, and Next Steps

We have received a planning grant from the US National Science Foundation (grant CNS-2016397) to gather community input and establish feasibility and plans for this infrastructure. We have conducted initial consultations at an advertised open gathering at RecSys 2020 and through email solicitations for research questions that researchers would wish to address using the infrastructure. While NSF support is primarily focused on supporting US researchers, we recognize that successful design (and in some cases key questions) require a broader international perspective.

Request for Input. We have brought this work to INRA with the goal of gathering input from both prospective users of this infrastructure and prospective partners in its construction and/or operation. While we are eager to receive input on a wide range of issues, there are five key questions that we hope will guide most of the input. We welcome input both at INRA and out-of-band (including by email to any of the authors or to the email list news-recsys@umn.edu).

1. **Please identify research questions that you would use this infrastructure to study.** Most useful would be about 1/3 of a page identifying the question, why it is important, why it needs the infrastructure, and what specific requirements it might present for the infrastructure (e.g., need for in-line surveys; need to log time-spent-reading by article, etc.).
2. **Please identify specific technical requirements you feel would be important for the infrastructure for it to support your work.** These could include information about the aggregator interfaces themselves, the content, the user population, or (perhaps most interestingly) the elements of the system that can be customized for certain users in an experiment and the factors that can be measured and reported in experiments. If there are features of the experimental framework you feel are important, please tell us.
3. **Please make recommendations about the governance of the infrastructure that would make it more useful for your work.** How should scarce resources (users) be allocated to experiments? What structure should be in place to ensure that researchers have an incentive to recruit new users to the system? How should human subjects review work? Or other thoughts about governance.
4. **Please identify resources (including collaborators) that you feel could make this more feasible or more successful.** Is there an existing aggregator you know who would partner with us to provide content? Particularly talented developers working in this space? Existing software or other infrastructure? Models of other shared resources you think we should look at? We're interested in hearing whatever suggestions you have.
5. **Finally, we'd like your thoughts on long-term financial sustainability.** NSF funds the creation of research resources, but expects them to become self-sustaining over time. We're interested in thought about the degree to which researchers could afford to pay for their experimental use (perhaps based on the number of users and length of the experiment). We're also interested in hearing about other models for financial support.

We are grateful for any input and feedback you can provide. We will be working towards a January deadline to apply for funds under the NSF CISE Community Research Infrastructure program.