# Content Recommendation for Semantic Stream

## ABSTRACT

## Keywords

## 1. INTRODUCTION

Developing efficient and accurate personalized recommendation systems of news-worthy content has increasingly become a focus of today's media industry. On the one hand search engines have enabled users to locate articles or multimedia content of any topic within a few keystrokes. On the other hand, users are constantly overwhelmed by the amount of information generated everyday, much of which is irrelevant to their personal interest. Thus we are seeing a resurgence of interest in directory based content discovery systems, such as the infinite news streams pioneered at yahoo.

One of the earliest prototypes of personalized ranking system is based on a simple linear matching model of user interest profile and document topic profile. This crude approach has proved effective in promoting highly relevant articles, videos or slideshows to each individual user, even when we have only accumulated partial information about the user. Unfortunately there are several drawbacks to this simple model. First the dot product between the user profile and document profile is not necessarily a natural one, since both profiles are calibrated separately, with no natural complementary "units" relating the two. An immediate remedy is to learn a non-Euclidean inner product with suitable objective functions such as click through rate or dwell-time. Second, while the linear model presents a ranking order of documents based on user's implicit or declared interest, it nonetheless does not take into account generic information such as user's age, gender, time of the day, day of the week, etc, which could be crucial to users' preference among the top documents presented. These are the actual content that the user sees, which clearly requires more careful optimization than the entire content pool.

A second challenge goes back to the initial selection of content itself. While general news streams certainly occupy the central stage in terms of breadth of user coverage, there tends to be less depth and personalization opportunity involved in their content. Users who are interested in a specific category of news such as sports or finance, or even narrower ones like American football, are naturally interested in dedicated streams just for those topics. Indeed this is how one accumulates domain-specific expertise as opposed to common knowledge. Thus we are interested in constructing appropriate filters conforming to prescribed semantic topics.

In this paper we will address the second problem with a carefully designed linear classifier framework based on a combination of user click feedback and editorial label. We compare the performance of the classifier approach with the baseline TFIDF (token-frequency inverse-document-frequency) approach.

## 2. RELATED WORK