Collaborative Filtering for Web Marketing Efforts

Dan R. Greening

Chief Technology Officer, LikeMinds, Inc.
457 Bryant Street
San Francisco, CA 94107
greening@likeminds.com
(415) 284-6966

Abstract

Recommender systems can improve consumer response to ecommerce sites. Large commercial sites make extraordinary resource demands on systems and databases. This paper describes a case study involving Columbia House, a large consumer direct marketing firm. Columbia House required rapid response time with high traffic, good recommendations at site-opening, and the ability to recommend new titles as they became available. LikeMinds developed a parallel collaborative filtering recommender that provided nearly linear speedup, making recommendations in less than 30 milliseconds on a single processor. A technique called "composite archetypes" helped seed the database with information from legacy transaction databases, making good recommendations at the time of the site opening. Another technique called "objective archetypes" allowed new titles to be recommended using only categorical information.

Introduction

Collaborative filtering is a real-time personalization technique. It leverages similarities between people to make recommendations, and can be more accurate than mechanical techniques, such as rule-based systems or Bayesian methods.

LikeMinds uses collaborative filtering technology to improve marketing efforts for direct marketers: accurate personalization tends to drive up revenues on commercial web sites. LikeMinds developer John Hey created the first general collaborative filtering system in 1987. It was originally deployed in video recommender kiosks around Boston. Since that time, web commercialization has placed increasing demands on the technique.

Commercial Needs

Columbia House was one of the first large consumer retailers to move to the web. Columbia House markets music, videos and CDROMs directly to consumers as a loyalty club, originally through a direct mail marketing channel. Music customers receive 12 titles free for signing up, promising to purchase 6 more titles at regular price.

Columbia House has historically used rule-based personalization in its business, hiring a small army of analysts to assign each new album to one or more listening preferences. The first Columbia House web site used this personalization method. However, with broad categories, such as "alternative rock," rule-based personalization was not very personal.

Columbia House decided to create a new site that would make more individualized recommendations to visitors. It chose LikeMinds collaborative filtering technology for this effort.

The Columbia House site required real-time performance under heavy usage. Columbia House has the second highest number of ecommerce transactions of all web sites, and offers thousands of products. It had an existing relational database that provided customer and product information, and wanted to store ratings and intermediate values in its database.

For the first time, a major consumer products company wanted to integrate collaborative filtering into every aspect of a site, making performance a critical issue. Whenever a customer searched for albums, the site would state which albums were recommended. Nearly every page view demanded personalization. The recommended albums were to be different for each customer: segmentation, which increases speed at the expense of accuracy, could not be used to speed the process.

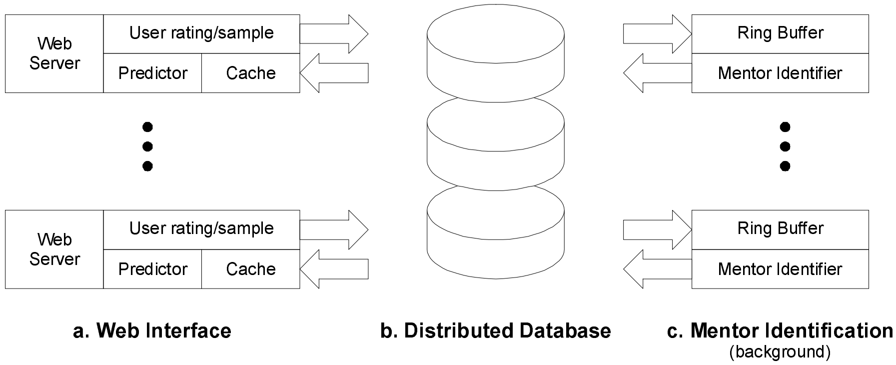

Figure 1. Parallelism in Preference Server 2.1

Parallel Collaborative Filtering

To satisfy these constraints, LikeMinds created a parallel processing model with a high degree of scaleability, as shown in Figure 1. Database locking was virtually eliminated by separating mentor matching processes from prediction vector construction processes. Multithreading was used throughout to keep long database requests from interfering with short prediction computations. Columbia House initially deployed the system on four networked UNIX processors. Later, LikeMinds moved its own demonstration site, MovieCritic, to a symmetric multiprocessor running UNIX. Performance scaling is close to linear in both cases.

Leveraging Legacy Data

Columbia House wanted to make good recommendations as soon as the site opened to the public. Like many direct marketers, Columbia House had a wealth of information about the past buying behavior of its customers. LikeMinds built a system to cluster these customers and aggregated them into "rating vectors" for hypothetical consumers, called "composite archetypes." LikeMinds used these archetypes to seed the rating database, successfully recommending albums from day one.

Books

	The Island of the Colorblind	A Leg to Stand On	Awakenings	An Anthropologist on Mars	Migraine	Paul Broca: Explorer of the Brain	James Parkinson	In Search of Churchill		**Hated It Item**	**Relevance weight**
Virtual Users											
VU author=Oliver Sacks	13	13	13	13	13					1	0.9
VU publisher=Harper		13	13							1	0.4
VU year=1992					13	13				1	0.2
VU subject=neurology	13	13	13	13	13	13	13			1	0.6
VU subject=biography					13	13	13	13		1	0.6
Real Users											
Smith			12		10					1	
Jones								8		1	
Wesson			12							1	

Mentors

	Smith	Jones	Wesson	VU author=Oliver Sacks	VU publisher=Harper	VU year=1992	VU subject=neurology	VU subject=biography
Smith			10	12.2	9.5	7	12.2	7
Jones								5.5
Wesson	10			9.5	9.5		9.5	

Figure 2. Generating Recommendations from Objective Factors

Columbia House wanted to recommend new products immediately. Straight collaborative filtering cannot recommend a product until someone rates it. To resolve this, LikeMinds used a variation of rule-based personalization, shown in Figure 2, to adaptively recommend products by category, artist, or listening type, creating hypothetical consumers who like everything satisfying a selection criteria. We call these hypothetical consumers "objective archetypes". A system seeded with objective archetypes adapts as users rate or buy new products, moving from rule-based predictions to more accurate behavior based predictions.

Although it may seem that collaborative filtering has matured, LikeMinds customers continue to make unforeseen demands on the technology, motivating innovative solutions.