

Content-Based Recommendation Systems

Michael J. Pazzani¹ and Daniel Billsus²

¹ Rutgers University, ASBIII, 3 Rutgers Plaza
New Brunswick, NJ 08901
pazzani@rutgers.edu

² FX Palo Alto Laboratory, Inc., 3400 Hillview Ave, Bldg. 4
Palo Alto, CA 94304
billsus@fxpal.com

Abstract. This chapter discusses content-based recommendation systems, i.e., systems that recommend an item to a user based upon a description of the item and a profile of the user's interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend. The profile is often created and updated automatically in response to feedback on the desirability of items that have been presented to the user.

10.1 Introduction

A common scenario for modern recommendation systems is a Web application with which a user interacts. Typically, a system presents a summary list of items to a user, and the user selects among the items to receive more details on an item or to interact with the item in some way. For example, online news sites present web pages with headlines (and occasionally story summaries) and allow the user to select a headline to read a story. E-commerce sites often present a page with a list of individual products and then allow the user to see more details about a selected product and purchase the product. Although the web server transmits HTML and the user sees a web page, the web server typically has a database of items and dynamically constructs web pages with a list of items. Because there are often many more items available in a database than would easily fit on a web page, it is necessary to select a subset of items to display to the user or to determine an order in which to display the items.

Content-based recommendation systems analyze item descriptions to identify items that are of particular interest to the user. Because the details of recommendation systems differ based on the representation of items, this chapter first discusses alternative item representations. Next, recommendation algorithms suited for each representation are discussed. The chapter concludes with a discussion of variants of the approaches,

the strengths and weaknesses of content-based recommendation systems, and directions for future research and development.

10.1.1 Item Representation

Items that can be recommended to the user are often stored in a database table. Table 10.1 shows a simple database with records (i.e., “rows”) that describe three restaurants. The column names (e.g., Cuisine or Service) are properties of restaurants. These properties are also called “attributes,” “characteristics,” “fields,” or “variables” in different publications. Each record contains a value for each attribute. A unique identifier, ID in Table 10.1, allows items with the same name to be distinguished and serves as a key to retrieve the other attributes of the record.

Table 10.1. A restaurant database

| ID | Name | Cuisine | Service | Cost |
|-------|----------------|---------|---------|--------|
| 10001 | Mike’s Pizza | Italian | Counter | Low |
| 10002 | Chris’s Cafe | French | Table | Medium |
| 10003 | Jacques Bistro | French | Table | High |

The database depicted in Table 10.1 could be used to drive a web site that lists and recommends restaurants. This is an example of structured data in which there is a small number of attributes, each item is described by the same set of attributes, and there is a known set of values that the attributes may have. In this case, many machine learning algorithms may be used to learn a user profile, or a menu interface can easily be created to allow a user to create a profile. The next section of this chapter discusses several approaches to creating a user profile from structured data.

Of course, a web page typically has more information than is shown in Table 10.1, such as a text description of the restaurant, a restaurant review, or even a menu. These may easily be stored as additional fields in the database and a web page can be created with templates to display the text fields (as well as the structured data). However, free text data creates a number of complications when learning a user profile. For example, a profile might indicate that there is an 80% probability that a particular user would like a French restaurant. This might be added to the profile because a user gave a positive review of four out of five French restaurants. However, unrestricted text fields are typically unique and there would be no opportunity to provide feedback on five restaurants described as “A charming café with attentive staff overlooking the river.”

An extreme example of unstructured data may occur in news articles. Table 10.2 shows an example of a part of a news article. The entire article can be treated as a large unrestricted text field.

Table 10.2. Part of a newspaper article

Lawmakers Fine-Tuning Energy Plan
SACRAMENTO, Calif. ~ With California's energy reserves remaining all but depleted, lawmakers prepared to work through the weekend fine-tuning a plan Gov. Gray Davis says will put the state in the power business for "a long time to come." The proposal involves partially taking over California's two largest utilities and signing long-term contracts of up to 10 years to buy electricity from wholesalers.