Assignment 2 Instructions: Content-Based Recommenders

Overview

In this assignment, you will hand-create and use some content-based profiles. You'll go through a set of variations to see how certain features of the computation can introduce (or reduce) biases.

Instructions

The Data Set

First, download the Assignment 2 dataset (linked here as a spreadsheet). It contains a table of content attributes for 20 documents across 10 attributes. It also lists two users' evaluations of five documents each. For purposes of this assignment, we're treating content attributes as boolean (either an article is about a topic or it isn't) and we're treating evaluations as positive (liked it), negative (disliked it), or unknown (never saw it).

Part 1. Build and use a very basic profile

First, you will build a very simple profile of user preferences for attributes. In this profile, you'll count the total the number of positive and negative evaluations associated with each attribute, and create a profile with the total score for each attribute for each user. For example, user 1's score for "Family" will get a +1 from doc1 (positive evaluation) and a -1 from doc 19 (negative evaluation) for a total profile value of 0 (neutral). In contrast, user 2's score for Europe will be +3 (+1 each for doc2, doc4, and doc17).

You can compute the profiles and place them in the "User Profiles" section of the spreadsheet