



Collaborative Filter Service

Table of Contents

1. WHAT IS A COLLABORATIVE FILTER	3
2. WHAT IS THE SERVICE OFFERING	3
2.1. WHAT ARE THE DETAILS OF THE IMPLEMENTATIONS?	4
3. HOW DO I GET STARTED.....	6
4. WHAT ARE THE BEST PRACTICES OF USING THE COLLABORATIVE FILTER.....	6
5. IS THERE A CHARGE TO THE SERVICE.....	6
6. WHAT IS THE STATUS OF IMPLEMENTATION	6

1. What is a Collaborative Filter

Collaborative filtering (CF) is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including sensing and monitoring data - such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data - such as financial service institutions that integrate many financial sources; or in electronic commerce and web 2.0 applications where the focus is on user data, etc. The remainder of this discussion focuses on collaborative filtering for user data, although some of the methods and approaches may apply to the other major applications as well.

The method of making automatic predictions (filtering) about the interests of a user by collecting [taste](#) information from many users (collaborating). The underlying assumption of CF approach is that those who agreed in the past tend to agree again in the future. For example, a collaborative filtering or [recommendation system](#) for [music](#) tastes could make predictions about which music a [user](#) should like given a partial list of that user's tastes (likes or dislikes). Note that these predictions are specific to the user, but use information gleaned from many users. This differs from the more simple approach of giving an [average](#) (non-specific) score for each item of interest, for example based on its number of [votes](#).

2. What is the Service Offering

The Offering is an implementation of the Collaborative Filter described in the Book – “Programming Collective Intelligence”. The implementation is a Java Implementation. The Service implements the following:

For the discussion below consider the following scenario as the backdrop to the discussion.

Scenario: An Internet Music Radio(WOW Music) has users that listen to songs. On listening to each song the user will rate the song as

- a) Like It
- b) Not Rated – played the full song
- c) Not Rated – Clicked on next song after listening to the song for 5 seconds
- d) Hate It

WOW Music intends to use a collaborative filter from webscale solutions to recommend songs to its users based on the principals of collective intelligence and specifically – collaborative filtering.

2.1. What are the details of the implementations?

In the Scenario describe above. WOW music goes the webscale solutions website and configures it Collaborative Filter , in the following manner.

The screenshot shows a dialog box titled "Interaction Definition : PICTURE Rating". It contains the following fields and options:

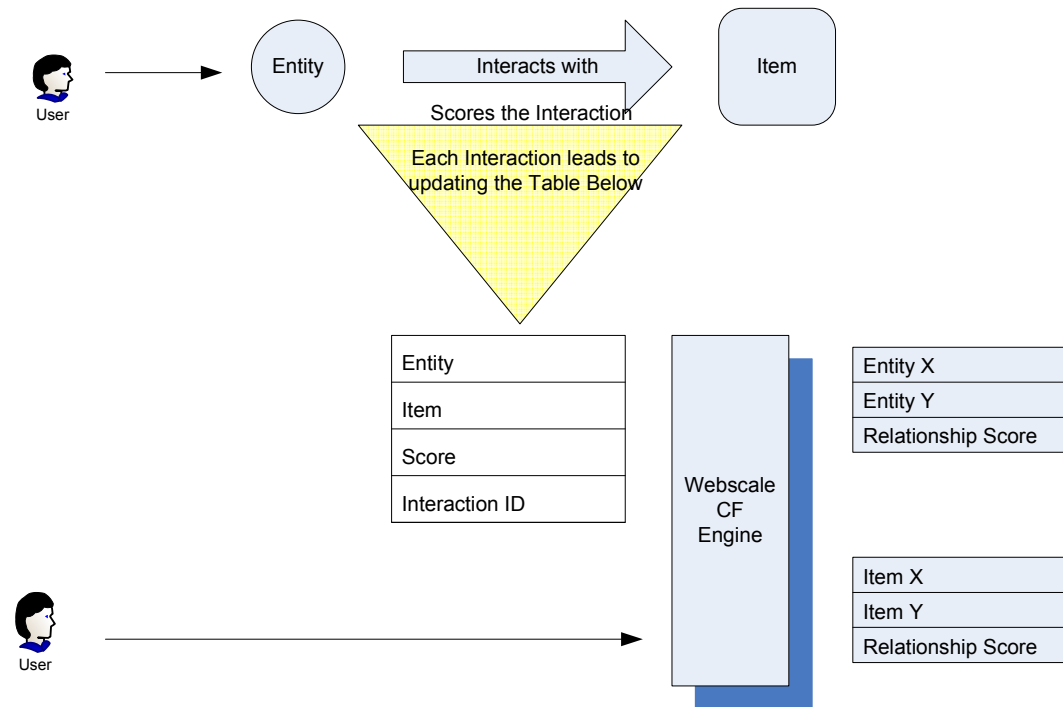
- Interaction Name: User - Song Interaction
- Interaction Description: A User rates a song on a scale of 0-4
1- Hate the song , 2 - Pass , 3- Like , 4- Really like
- Relationships: Item Map Entity Map
- Algorithm: Pearson Coefficient Euclidean Distance
- Max Score: 4
- Min Score: 0
- Persistence Mechanism: Big Table SimpleDB RDBMS

Buttons: Save, Revert

Once an Interaction Type is defined , WOW music will now start sending JSON Requests on the lines of

```
Sample JSON Request - For Data push
{
  "entityKey": "USER_ID_IN_WOW_DB",
  "itemKey": "SONG_ID_IN_WOW_DB",
  "interactionId": 3,
  "custKey": 34232,
  "passKey": "SDUIWIDNCNCIEISS",
  "transactionscore": 2.0
}
```

Please look at the diagram below that explains the interactions



After several requests have been sent to the server(> 1000 for an interaction type). WOW music can now start making calls to the Collaborative Filter to get Relationship Scores between Items and Entities.

The methods available for this are

```

Java Interface for Collaborative Filter
1. public List<RelatedEntity> findClosestEntitites(Entity e , int limit , int
mimShareCount , InteractionType forType );

2. public List<RelatedItem> findClosestItems(Item e , int limit , int
mimShareCount , InteractionType forType );

3. public List<Item> getRecommendedItemsBasedOnEntity(Entity e , int limit ,
int mimShareCount , InteractionType forType);
    
```

All the three methods are available as JSON Requests.

1. findClosestEntites – This returns the Entities closest to a particular give Entity
2. findClosestItems – This returns the Items closest to a particular give Item
3. getRecommendedItemsBasedOnEntity – This returns recommended items based on the an Entity.

3. How do I get Started

- a) Visit the website at <http://www.webscalesolutions.com>
- b) Click on Register Link and register
- c) Once you register you can go to the console screen shown below and configure your service

4. What are the best practices of using the Collaborative filter

The best place to start is – “Programming Collective Intelligence” by Toby. The book to some was the motivation behind this service.

If you need further information and a training session – please contact info@webscalesolutions.com

5. Is there a charge to the service

The service is in beta right now and there is not charge. If you planning to dump in large data. Please contact info@webscalesolutions.com before you do that.

6. What is the status of implementation

- a) The RDBMS implementation is live and runs on Sun’s Project Caroline and postgres
- b) Simple DB is the next item on priority list and we hope to have the service live by October end / early November
- c) BigTable implementation on Google AppEngine will be the third item on the list

I do intend to implement other patterns like Clustering and more. If you have use case and need an implementation earlier. Please contact info@webscalesolutions.com

7. Any final comments

As evident from this document and the completeness of this document. I have a long way to go before I can truly call this as an industrial strength service. So , if you really have a need and want to get started – PLEASE CONTACT ME at info@webscalesolutions.com and I would be very willing to get you started.