Problem Statement

Recommender Systems

DATA SUMMARY

MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota.

This data set consists of:

\* 100,000 ratings (1-5) from 943 users on 1682 movies.

\* Each user has rated at least 20 movies.

The data was collected through the MovieLens web site

(movielens.umn.edu) during the seven-month period from September 19th,1997 through April 22nd, 1998. This data has been cleaned up – users who had less than 20 ratings or did not have complete demographic information were removed from this data set. Detailed descriptions of the data file can be found at the end of this file.

DETAILED DESCRIPTIONS OF DATA FILES

Here are brief descriptions of the data.

u.data -- The full u data set, 100000 ratings by 943 users on 1682 items.

Each user has rated at least 20 movies. Users and items are

numbered consecutively from 1. The data is randomly

ordered. This is a tab separated list of

user id | item id | rating | timestamp.

The time stamps are unix seconds since

u.info -- The number of users, items, and ratings in the u data set.

u.item -- Information about the items (movies); this is a tab separated

list of

movie id | movie title | release date | video release date |

IMDb URL | unknown | Action | Adventure | Animation |

Children's | Comedy | Crime | Documentary | Drama | Fantasy |

Film-Noir | Horror | Musical | Mystery | Romance | Sci-Fi |

Thriller | War | Western |

The last 19 fields are the genres, a 1 indicates the movie

is of that genre, a 0 indicates it is not; movies can be in

several genres at once.

The movie ids are the ones used in the u.data data set.

**Steps Involved**

1. Data Gathering
2. Data Cleaning (Filling Null Values, removing Garbage Data )
3. Feature Training
4. Research Of Market : Models and Algorithms ( With Efficiency)
5. Inferences – Checking Accuracy
6. Results
7. Improve Results – (Comparing Other Models on the basis of Output)