# Recommender System Using Sentiment Analysis

Vidya Adekar[1], Sonali Acharya[2], Prathmesh Kedar[3], Nitin Borkar[4],
Dr.Prasad Halgaonkar[4]
Zeal College of Engineering and Research, Pune

## Abstract:

For security of the data as well as maintaining privacy over the internet, authentication is very important. Usually the password which is used by the user is small password, easy to memorize or password which can be guessed easily. People use personal information as their password for easy memorization. In this paper, we analysis the various passwords from the leaked dataset to research their personal information for finding the relation between them and the password. We use Probabilistic Context-Free Grammars (PCFG) method with semantic-rich method to propose Personal-PCFG method in which it will find if there is any correlation between password and personal information. This method will help us to crack the password much faster than any other method which increases the chances of successful password crack. To protect user's from this type of attacks we use distortion function. This paper also provide security by notifying authenticated user if someone is trying to access the account from another or trying to attempt their account by entering wrong password.

## Introduction

### 1.1 Detailed Problem Definition

Recommender system these days mostly follow the collaborative filtering techniques. These techniques generally find users with users with similar surfing patterns (interest) and recommend the products viewed by those users. There are various ways that can tell user's interest apart from surfing patterns. Existing methodologies fail to cover those ways. Thus, performance of system can still be improved in terms of finding interests of user.

### 1.2 Justification of the Problem

In the existing systems, the user with similar interests were recommended by same services or products. This was done by using Collaborative Filtering (CF) and Content-Based Filtering (CBF). But these recommendations were not much efficient and accurate. Also, more computation time was required. The previous system's recommendations were for a particular domain, but the proposed system takes into considerations various domains for recommendations.

### 1.3 Need for the new system

A recommender system is commonly associated with only one domain, while the recommender system described in this paper is able to generate recommendations

from different domains (movies and music). In addition to recommendations related with the specific domain, our system is able to recommend the web articles (unstructured text), relevant to the user that may belong to more than one category of interest.

### 1.4 Advances / Additions / Updating Previous Systems

In our project we are overcoming the drawbacks of previous systems. In our system, Natural Language Processing is being used to extract interest out of the post that is being posted by the user. This interest is taken into consideration for generating recommendation for that particular user. The system also shows the websites with different area of interest that the user has posted for previously.

### 1.5 Presently Available Systems for The Same

Currently recommender systems identify likes/dislikes/users' interests based on the products that a user surfs through. Then system finds users with similar interests. These users collectively referred as neighborhood. Users from same neighborhood have similar interests. Thus, products from their wish list, products they've bought in the past as well as the products that these users are looking forward to buy, can be recommended to other users and that's what these systems do.

### 1.6 Purpose of Your System

The main Objective of the system is to identify the interests of user from his/her social networking profiles and search products with the help of recommender system which collects information about product through RSS feeds, different and web crawling systems etc. Using that information, system recommends products that can be more relevant to user’s interests.

System will be able to generate recommendations from different domains (movies and music). System deduces user interests based on his activities and post in social network.

### 1.7 Organization of the Report

The report mainly focuses on the discussion about different aspects of evidence acquisition and the ways to address them. The first chapter deals with the introduction about the system defining the problem statement to which the system is focused on. It also discusses the scope of the system to which it works efficiently. It also contains briefing of various research papers is to be used. The proceeding chapter contains the detailed SRS of the project describing the user feasibility, functioning of various modules, time line chart and

process modeling. The third chapter contains mathematical model and feasibility analysis. It also contains all the use case diagrams related to the project.

## Problem Definition:

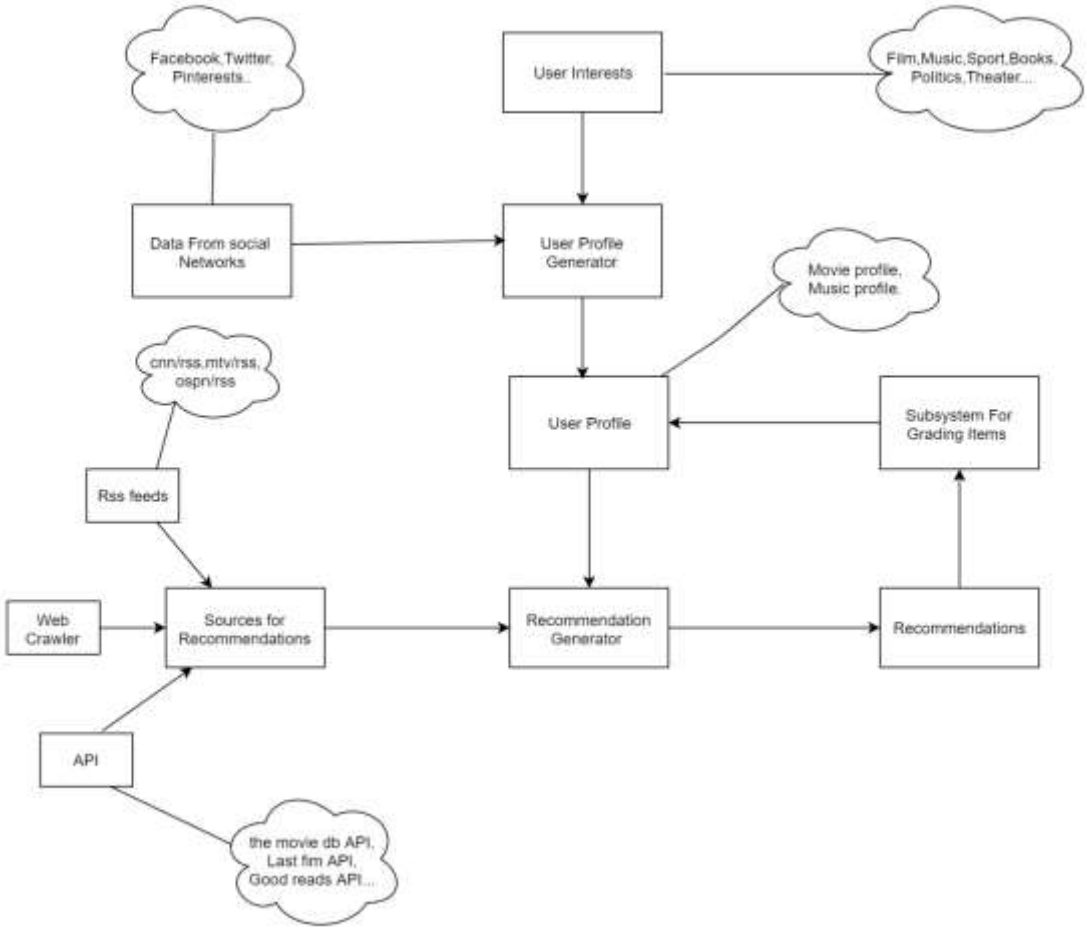
Recommender system these days mostly follow the collaborative filtering techniques. These techniques generally find users with users with similar surfing patterns (interest) and recommend the products viewed by those users. There are various ways that can tell user's interest apart from surfing patterns. Existing methodologies fail to cover those ways. Thus, performance of system can still be improved in terms of finding interests of user.

## Objectives

- To fetch websites related to various areas of interest.
- To extract interest from the user’s updated status.
- To recommend website links to the user according to his/her area of interest that is being analyzed by the system on the basis of user generated post.

## System design:

## SYSTEM ARCHITECTURE




## Related Works

1 Amazon.com Recommendations Item-to-Item Collaborative Filtering

Introduction: To be able to react to change in user's data quickly with the help of item-to-item collaborative filtering.

- Disadvantage :-

  1. System wasn’t implemented for targeted marketing.

2. An Algorithm for Suffix Stripping Program

Introduction:- To improve searching results/searching performance in IR systems by removing suffixes.

Disadvantage: - Systems suffers when dealing with complicated keywords

3 The Browsemaps: Collaborative Filtering at LinkedIn

Introduction:-

To analyze and present overview about collaborative filtering technique used in LinkedIn.com

4 Collaborative Filtering Recommender Systems

Introduction:-

To discusses a wide variety of the choices available and their implications, aiming to provide both practitioners and researchers with an introduction to the important issues underlying recommenders and current best practices for addressing these issues.

5 Collaborative Inference of Sentiments from Texts

Introduction:

To improve collaborative filtering technique by finding similar users from finding similarities in their text reviews.

Disadvantages: More similarity measures also needed to be added in order to achieve good performance.

6 Evaluating Recommendation Systems

Introduction :

To review the process of evaluating a recommendation system.

7 Text Categorization With Support Vector Machines: Learning with Many Relevant Features

Introduction: To user SVM for text classification and summarization

Disadvantages: System was built for text classification and yet to implemented in E-commerce application.

8 The Netflix Recommender System: Algorithms, Business Value,and Innovation

Introduction : To discuss and describe purposes of various recommendation algorithms used in NetFlix recommender system.

9 Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions

Introduction: To present an overview of content-based, collaborative, and hybrid recommendation approaches

Disadvantages: There's an need of improvement in modeling of users and items.

10 WTF: The Who to Follow Service at Twitter

Introduction: To study and provide an overview about Twitter's recommendation system.

## DESIGN OF THE STUDY

Propose Algorithm :-

**Step 1:** The admin will be registered to the database of the system.

**Step 2:** When the admin will login to the system, he/she will pass all the websites of different area of interest to the crawler so that internal links of each website link will be fetched upto depth one.

**Step 3:** Each link will be assigned a unique id so that the links will be further recommended to the user according to his/her area of interest.

**Step 4:** The user will be registered to the database of the system.

**Step 5:** As the user will login, he/she will post a status.

**Step 6:** The status will undergo Natural Language Processing so as to extract the user interest from the status posted.

**Step 7:** The links sorted by the admin with respect to the area of interest will be recommended to the user.

## TOOLS USED

- JDK 1.8 or higher version
- Eclipse Mars or higher version
- MySQL 5.7 or higher version
- Tomcat 8 or higher version

## Software Requirement:

- Operating System : Windows 8 and above..
- Application Server :Tomcat5.0/6.X
- Language :Java
- Front End :HTML, JSP
- Database :MySQL

**Hardware Requirement:** The hardware design of the system includes designing the hardware units and the interface between those units.

- Processor - Pentium –III/IV
- RAM - 4 GB (min)
- Hard Disk - 50 GB

## Statistical Technique Used

We have used the crawler technique and Natural Language Processing for our project. These techniques allow us to efficiently fetch the links and also extract the user's area of interest based upon the status posted

by the user. The system efficiently provides recommendation to the user according to his/her area of interest.

## Experiment Result

The system efficiently recommends websites to the users that are having different area of interests. The user's interest is analyzed from the status updated by the user by using Natural Language Processing. The interests found are being used to fetch those websites only that have the corresponding area of interest.

## Future scope:

In future we plan for developing a system with maximum different areas of interests so that user with any interest will get recommendations. Also, in future, we will develop a flexible system which can work with real time social media applications.

## Conclusion:

The system that we have developed can efficiently suggest the user different website links according to the user's area of interest. To determine the user's area of interest we are using Natural Language Processing where we are successfully able to extract nouns from the user generated post and thus we can find the user's area of interest. Using crawler the website links with various interest extract the internal links of the corresponding website.

## Reference:

[1] T. Pinckney, "eBay NYC - Graph-based Recommendation Engine on Cassandra," 30 July 2013. [Online]. Available: https://www.hakkalabs.co/articles/tom-pickney-discusses-ebaysrecommendation-engine.

[2] G. Linden, B. Smith i J. York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering,"2003.

[3] . A. Gomez i N. Hunt, "The Netflix Recommender System: Algorithms, Business Value, and Innovation," ACM Trans. Manage. Inf. Syst., 2016.

[4] C. Mims, "MIT Technology Review - How iTunes Genius Really Works," 2010. [Online]. Available: https://www.technologyreview.com/s/419198/how-itunes-geniusreally- works/.

[5] L. Xiang, "Hulu Tech blog - Hulu's Recommendation System," 19 September 2011. [Online]. Available: http://tech.hulu.com/blog/2011/09/19/recommendationsystem.
html.

[6] L. Wu, "Browsemap: Collaborative Filtering At LinkedIn," 23 October 2014. [Online]. Available: https://engineering.linkedin.com/recommendersystems/
browsemap-collaborative-filtering-linkedin.

[7] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang i R. Zadeh, "WTF: The Who to Follow Service at Twitter," u Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil, 2013.

[8] S. K. Rogers, "Item-to-item Recommendations at Pinterest," u Proceedings of the 10th ACM Conference on Recommender Systems, Boston, Massachusetts, USA, 2016.

[9] J. Moore, "Building a recommendation engine, foursquare style," 22 March 2011. [Online]. Available: https://engineering.foursquare.com/building-a-recommendationengine-foursquare-style-4df6dc23ea15#.tk20r8sd0.

[10] "Mendeley," [Online]. Available: https://www.mendeley.com/.

[11] "Google Scholar," [Online]. Available: https://scholar.google.ca/.

[12] "PubChase," [Online]. Available: https://www.pubchase.com/.

[13] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2005.

[14] B. Schafer, D. Frankowski, J. Herlocker and S. Sen, "Collaborative Filtering Recommender Systems," Lecture Notes In Computer Science, vol. 4321, pp. 291-324, 2007.

[15] "Cosine similarity, Pearson correlation, and OLS coefficients," 13 march 2010. [Online]. Available: https://brenocon.com/blog/2012/03/cosine-similarity-pearsoncorrelation- and-ols-coefficients/.

[16] D. Gupta, M. Digiovanni , H. Narita and K. Goldberg, "Jester 2.0 (demonstration abstract): collaborative filtering to retrieve jokes," in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, Berkeley, California, USA, 1999.

[17] . Joachims, “Text Categorization With Support Vector Machines: Learning with Many Relevant Features,” in European Conference on Machine Learning, Chemnitz, Germany , 1998.

[18] . M. Porter, An Algorithm for Suffix Stripping, Program, 1980, pp. 130-137.

[19] Y. Seroussi, I. Zukerman and F. Bohnert, “Collaborative Inference of Sentiments from Texts,” in User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, 2010.

[20] F. Esposito, S. Ferilli, T. Basile and N. Di Mauro, “Machine Learning for Digital Document Processing: from Layout Analysis to Metadata Extraction,,” in Machine Learning in Document Analysis and Recognition, 2008, pp. 105-138