

1. Introduction

1.1. Background

Recommendation engine is most commonly used in retail business or ecommerce websites to increase sales and generate revenue out of it. A recommendation engine is a system that suggests products, services, information to users based on analysis of data. Notwithstanding, the recommendation can derive from a variety of factors such as the history of the user and the behaviour of similar users. Recommendation systems are quickly becoming the primary way for users to expose to the whole digital world through the lens of their experiences, behaviours, preferences and interests. And in a world of information density and product overload, a recommendation engine provides an efficient way for companies to provide consumers with personalised information and solutions. Most of the recommendation engine algorithms use content based or collaborative filtering (User – User/Item - Item) or Hybrid approach (both). And in all these approaches, rating is used as a feedback variable. As we know some of the customers give their feedback as a rating and some as a comment. In the current approach most of the companies are only using rating as a feedback variable because it is in numeric format, and to run any algorithm, the data need to be in numeric format. People don't use comments as a feedback variable because it is in a non-numeric format. But using NLP technique we can convert this non numeric data into numeric format, prior to which we can use this in our recommendation collaborative filtering approach.

1.2. Objective

In our research paper, the main objective is to use NLP techniques to convert the unstructured data into structured data to enrich our dataset and use that data into recommendation engine algorithm to improve the performance our algorithm, which in return will improve the accuracy of our algorithm and will predict correct recommendation to our customers so they can buy that products.

1.3. Scope, Purpose, Applicability

1.3.1. Scope

Firstly, we will need unstructured data like comments, and then we will apply Natural language processing technique to convert unstructured data like comments into structured data on the scale of 0 to 5. Then we have to fill the missing values in rating variables, from the data which we converted into structured data and use it in our recommendation engine algorithm for better prediction.

1.3.2. Purpose

Purpose is to improve the performance and accuracy of recommendation engine algorithms to sell more products and earn more profit by giving better service.

1.3.3. Applicability

This Paper is applicable to any organization, which is collecting unstructured data like comments and are not using their recommendation engine algorithm to enrich data.

2. Description of the problems/topics

- Explain the problem and state why it matters?
 - Most common problem which comes up when building a recommendation engine is 'High percentage of Data Sparsity'. Sparsity and density are terms used to describe the percentage of cells in a database table that are not populated and populated, respectively. The sum of the sparsity and density should equal 100%. A table that is 10% dense has 10% of its cells populated with non-zero values. It is therefore 90% sparse – meaning that 90% of its cells are either not filled with data or are zeros. This is a huge problem in building recommendation engines and needs to be solved. Because if we recommend the wrong product to customers they might not get proper service, which may lead to customer loss.
- Describe how things should work?
 - Now to reduce the percentage of Data Sparsity and to increase the percentage of Data density, we have to use the data which we already collect but we don't use it in our recommendation algorithm. Here, in this paper we will be using unstructured data like comments which the people in industry ideally don't use for building recommendation engines.
- Propose a solution and Explain the benefits of your proposed solution(s)?
 - To solve the problem of High data sparsity, and to increase the percentage of data density, we are using NLP techniques to convert the unstructured data into structured data. We are going to normalize the comments of customers on the scale of 0-5. Wherever the rating is missing, we will use the normalized values which we have generated using NLP techniques.

- Conclude by summarizing the problem and solution?
 - Solving the problem of Data Sparsity in the current recommendation engine using NLP techniques for improving the performance and accuracy of algorithms.

3. **Status of the research/knowledge in the field and literature review**

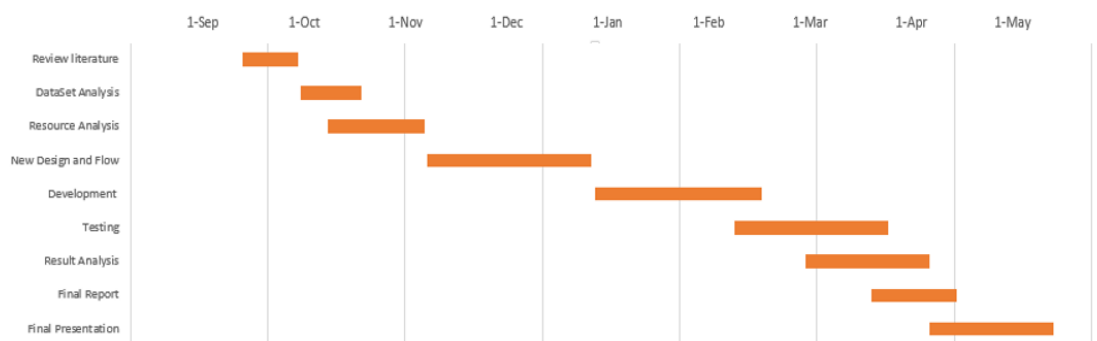
- We have referred these research paper:
 - Scienstein: A Research Paper Recommender System
 - A unified context-free grammar and n-gram model for spoken language processing
 - Collaborative Document Evaluation: An Alternative Approach to Classic Peer Review.
- Comparative study of existing system
 - Below are the advantages and disadvantages of recommendation engine algorithms.
 - Advantages
 - Simple to understand and implement the algorithm, if rating of most of the customers is available as a feedback for their product
 - No feature selection needed in collaborative filtering method, which is widely used in recommendation engines.
 - Minimum domain knowledge required.
 - Disadvantages
 - Data Sparsity is a major drawback which gives less accurate results.
 - Cold start problem.
 - In this paper we are going to solve the first disadvantages of the recommendation engine by embedding NLP technique in the recommendation engine algorithm.

4. **Description of the methodology/approach**

- Methodology

	Item i_1	Item i_2	Item i_3	Item i_4	Item i_5
Target user User u_1	5	7	5	7	?
Users most similar to the target user User u_2	5	7	5	7	9
User u_3	5	7	5	7	9
User u_4	6	6	6	6	5
User u_5	6	6	6	6	5

- Planning
 - Gantt Chart



- Flow chart diagram

