SOCIAL RECOMMENDATION ENGINES USING MACHINE LEARNING

Archit Prasad Kane
Roll no. 31206
TE 2
Guide: Dr. A. S. Ghotkar

CONTENTS

1

PROBLEM DEFINITION AND SCOPE

PROBLEM DEFINITION

Understand how social recommendation engines work with the help of machine learning.
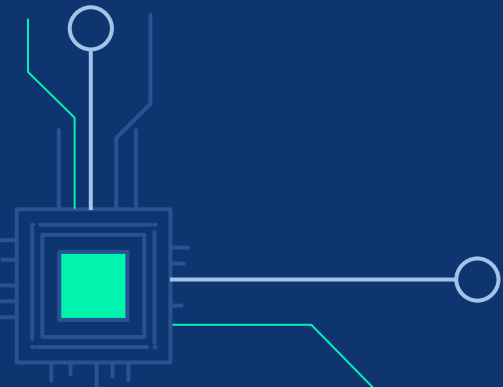
SCOPE

Extracting user-generated data and examining it to find out the users interests. The data includes text and images shared or interacted with by users. This derived data is used to recommend users or forums of similar interests to the user on the social graph.
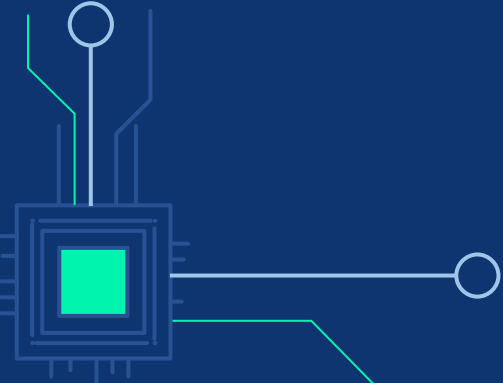
2

INTRODUCTION

INTRODUCTION

- The number of users using online social media platforms has increased over the past decade.
- This is coupled by the large number of social media communities.
- This change is also reflected in the types of messages sent.
- Users interact with users or communities.
- Interaction is based on their roles or topics of interest.
- These online connections are called social graphs.
- Connections happen by shared interests.
- Social Recommendation System Approaches:
  - Collaborative filtering - based on the opinions of the user.
  - Content-based filtering - based on the previous posts liked by the user.
  - Hybrid recommendations combine the above two specified methods.
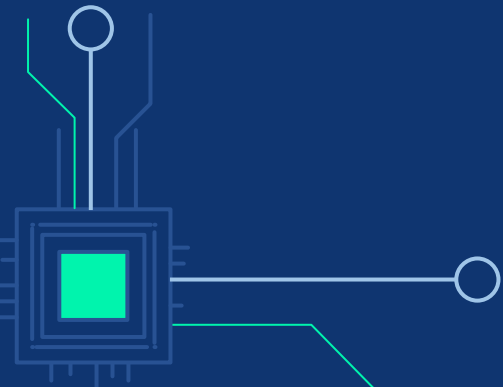
3

MOTIVATION

MOTIVATION

- The number of users using online social media platforms has increased over the past decade.
- Social Media has now transitioned from a simple text messaging platform to a richer multimedia experience.
- This lead to an increase in the amount of content that can be accessed by the user.
- Giving us the need to improve our social recommendation algorithms to provide more engagement to the user using machine learning.
- Content is tailored by looking into the user's past interests and activities. This helps to increase the engagement levels for the user.
- These techniques have been proved an effective alternative to the traditional approach which used user described tags or social graphs.

4

LITERATURE SURVEY

LITERATURE SURVEY

- Characterizing User Connections in Social Media through User Shared Images
  - Images analysed using Machine learning algorithms generate labels to identify the user's interests.
  - Irrespective of the source or medium of exchange between the users they share similar images of interest.
  - It shows better performance than the said friends of friends recommendation approach.
- Recommendation Framework for Online Social Networks
  - The algorithms needed to measure the strength of relationships, user activity and similarities are discussed.
  - The data required for analysis and recommending users is filtered through these algorithms.
  - These algorithms together are used to recommend new users on the social media platform.
- Characterizing User Behaviour and Information Propagation on a Social Multimedia Network
  - The roles among users are inferred from the behaviours on the social platform.
  - Using data extracted from user messages they can be clustered based on these behavioural patterns.
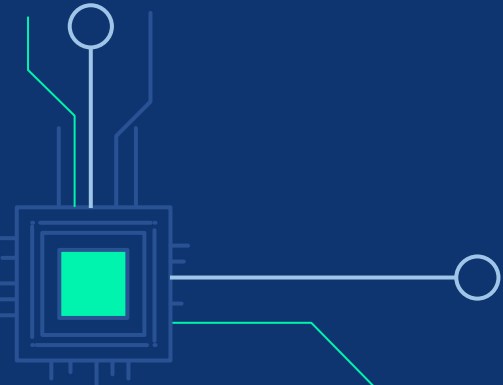
5

METHODOLOGY

- Data Collection
- Data Analysis
- Recommendation Process
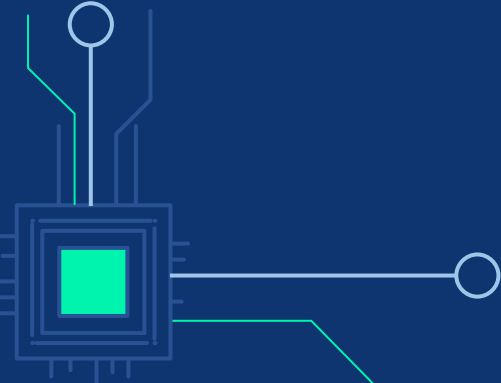
5.1

DATA COLLECTION

DATA COLLECTION

Static Data

Demographics

Language
Location
Age
Gender
Occupation

Interests

Books
Movies
Cycling

Dynamic Data

User To User

User To Forum

- Posts interacted with
- User Chat
- Media files shared
- Frequency of user communications

5.2

DATA ANALYSIS

DATA ANALYSIS

Text Based Data Analysis:

- Find out Topics Of Interest
- Rapid Automatic Keyword Extraction
  - a list of stop words ('and', 'of', 'the')
  - partition the text into an array of keywords
  - create a graph of word co-occurrence and calculate the word score.
  - Word Score Metrics:
    - Word Frequency,
    - Word Degree
    - Ratio of the degree to the frequency
    - words that occur often
    - words that occur the most

Image Based Data Analysis:

- Find out Topics Of Interest
- Neural network
- Extract Tags
- Assigned as topic of interest.

Conscientiousness: Frequency of responding to messages.

Extraversion: No of times conversations are initiated, analysing the images posted or tagged in.

Agreeableness: Strength of the social graph links.

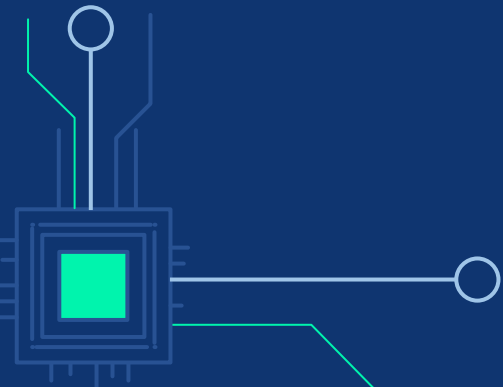Neuroticism: Analysing the chats sent.

DATA ANALYSIS

- K-means algorithm to extract entity grouping based on their interests.
- We then examine user profiles within the clusters.
- The following are the user features that have been identified:
  - Avg. no. of messages/broadcasts per day
  - Number of friends
  - A common time for messages/broadcasts
  - Diversity of topics within messages
  - Diversity of topics within posts
- The feature values are normalized to 0 - 1. This is done because each feature value have different units.
- We then apply the k-means algorithm to divide the users into clusters based on similar behaviour.
- Each user within the feature cluster goes through demographic and behavioural profiles to identify common characteristics between the users.
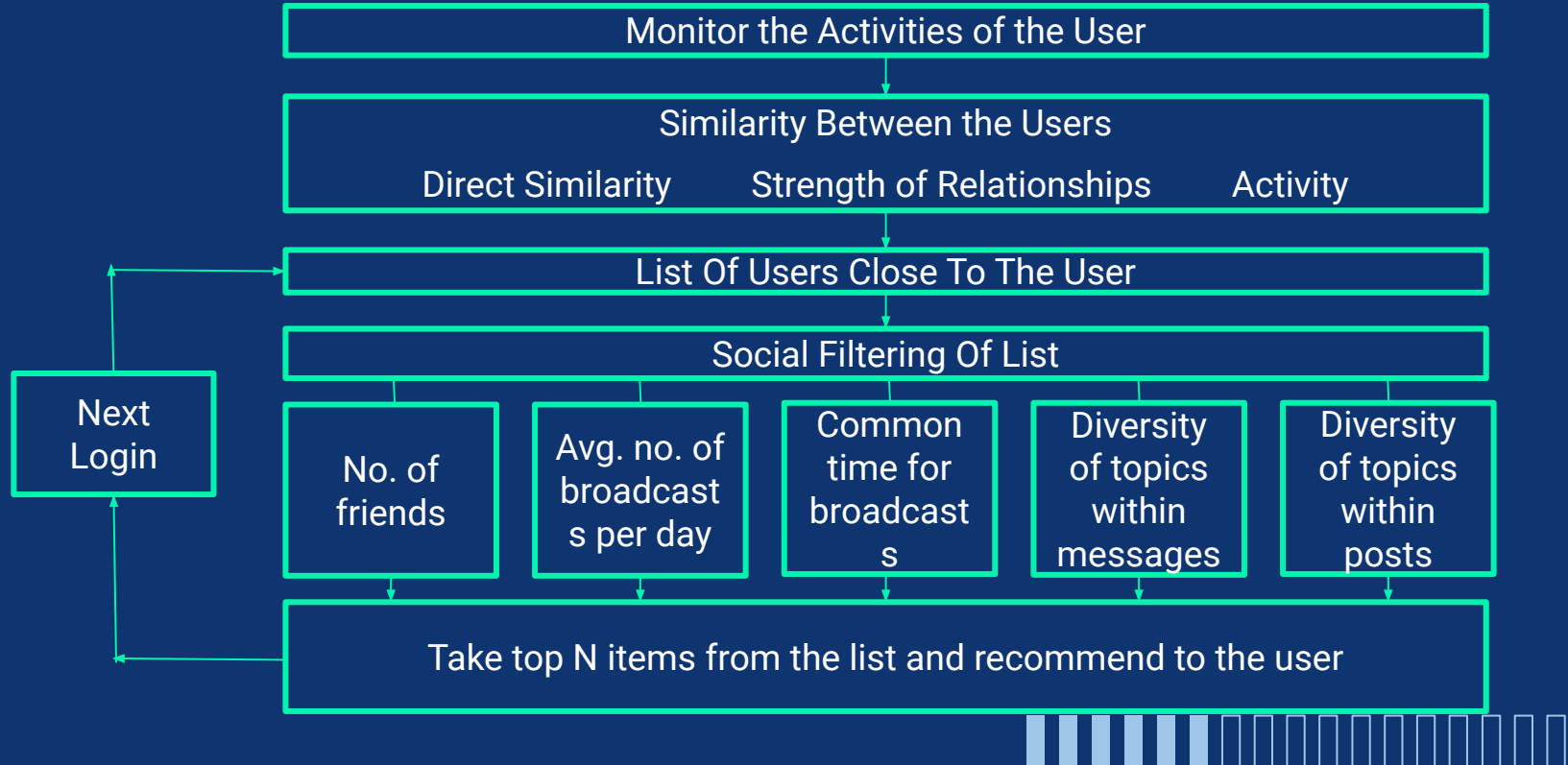
RECOMMENDATION PROCESS

Monitor the Activities of the User

Similarity Between the Users

Direct Similarity   Strength of Relationships   Activity

List Of Users Close To The User

Social Filtering Of List

| Next Login | No. of friends | Avg. no. of broadcasts per day | Common time for broadcasts | Diversity of topics within messages | Diversity of topics within posts |

Take top N items from the list and recommend to the user

RECOMMENDATION PROCESS

- The users are clustered separately into groups by using clustering algorithms. These groups are made based on the user's static data.
- Depending on the aim of the network the connectivity social filtering is used to improve the connections.
- This is accomplished with the help of final similarity function:
  - $r(x - y) = A * s(x, y) + B * c(x, y) + G * a(y) + D * sr(y)$
  - $s(x, y)$ – the comparison of demography and interests of x and y.
  - $c(x, y)$ – social behaviour of the users in relationships.
  - $a(y)$ – the activity of the user y
  - $sr(y)$ – the strength of the relationship between x and y.
- A,B,G,D are the importance coefficients which are assigned values from 0 -1.

6

CONCLUSION

This seminar examined how the social recommendation system works and the variety of data required along with the techniques of analysing and suggesting users and forums.
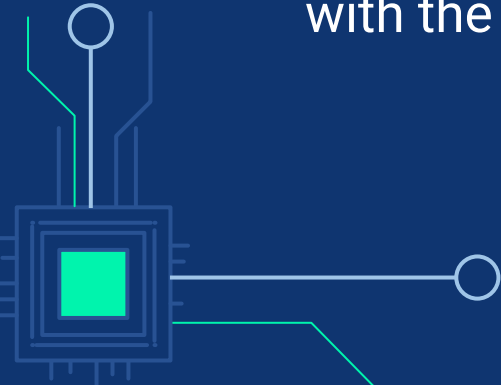
REFERENCES

- M. Cheung, J. She and N. Wang, "Characterizing User Connections in Social Media through User-Shared Images," in IEEE Transactions on Big Data, vol. 4, no. 4, pp. 447-458, 1 Dec. 2018, doi: 10.1109/TBDATA.2017.2762719.

- https://medium.datadriveninvestor.com/rake-rapid-automatic-keywordextraction-algorithm-f4ec17b2886c

- Kazienko, Przemys law Musial, Katarzyna. (2006). Recommendation Framework for Online Social Networks. 10.1007/3-540-33880-2-12.

- M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software," SIGKDD Explorations, vol. 11, no. 1, pp. 10–18, 2009.

- F. T. O'Donovan et al., "Characterizing user behavior and information propagation on a social multimedia network," 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2013, pp. 1-6, doi: 10.1109/ICMEW.2013.6618395.

THANK YOU