

ANALYSIS OF SOCIAL MEDIA DATA TO DETERMINE POSITIVE AND NEGATIVE INFLUENTIAL NODES IN THE NETWORK

SHUBHANSHU MISHRA (07MA2023)

Under the Guidance of:

Prof. Gloria Ng

Institute of Systems Science
National University of Singapore

&

Prof. Pawan Kumar

Department of Mathematics
IIT Kharagpur



A Common Social Media Interaction

Jason

500
connections

The latest model launched by Toyota has **fantastic** performance upgrade and **amazing** mileage. =)

20 mentions. **10** re-share.

Mark

1000
connections

I **hate** the new Toyota design, **pathetic** interiors.

15 mentions. **5** re-share.

Ashley

500
connections

Toyota, what a piece of **CRAP**.

80 mentions. **40** re-share.

Introduction

- Social Media is currently growing as a means of communication and opinion sharing by people.
- Everyday more than a million users utilize various social media platforms to connect, share ideas, views and learn.
- Every social media interaction can be divided into 2 types – POSITIVE interaction and NEGATIVE interaction.
- Each individual influences their network in a positive or negative manner, through each of their social media interaction.

Positive versus Negative

- Every interaction on social media is either positive or negative.
- **Positive interactions:**
 - They give a feel good effect
 - Talk about solutions
 - Appreciative in Nature
 - “Good”, “Awesome”, “Fantastic”, etc.
- **Negative Interactions:**
 - They are demotivational in nature
 - Argumentative, abusive, bullying sounding.
 - “Very Bad”, “unworthy”, “pathetic”

Problem Definition

- For an individual **[a]** in a given social media circle the objective is to determine the following
 - Total Positive influence – P_a
 - Total Negative influence – N_a
 - Overall influence – I_a
- Also based on the values above all the individuals in the network need to be ranked according to their positive influence and negative influence.

Sentiment Analysis

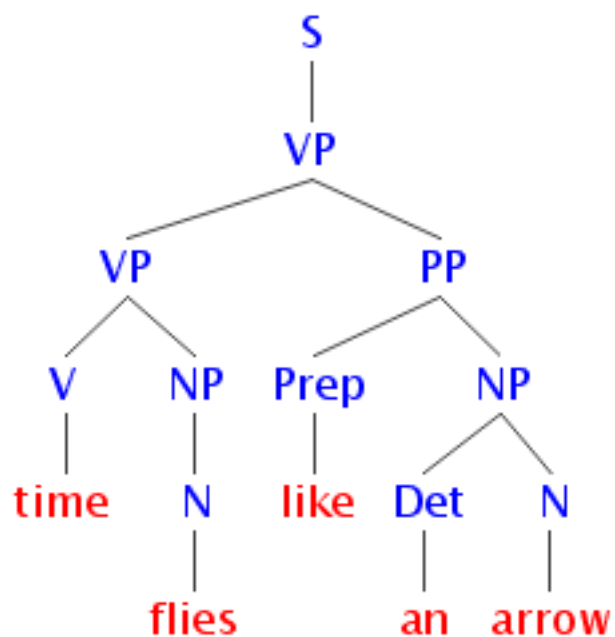
Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.

Fact: - "The painting was more expensive than a Monet"

Opinion/Sentiment: - "I honestly don't like Monet, Pollock is the better artist"

- The usage of sentiment analysis in this context is to classify an interaction as positive or negative.
- The process is to extract the important keywords and they sentence is classified based on the keyword type.

Process of Sentiment Analysis



S = subject

VP = Verb Phrase

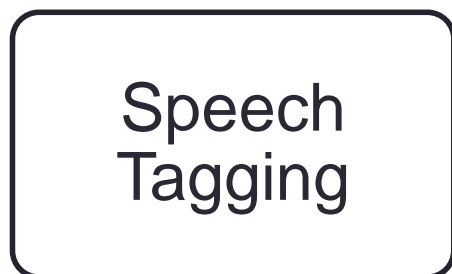
V = Verb

N = Noun

NP = Noun Phrase

PP = Preposition

Det = Determiner



Classifying Sentiments

- **Naive Bayes** (probabilistic classifier using Bayes theorem)
- **Maximum Entropy** (Uses probability distributions on the basis of partial knowledge)
- **Support vector machine** (Data is set as 2 vectors in an n-dimensional space) Pang et al. found the SVM to be the most accurate classifier (around 80%).

Methodology

- **Influence Group:**

- Followers of the User
- Profiles the User is following
- Individuals in the follower list in the same city as the user
- Individuals mentioned by the user in tweets

- **Influence of a profile is:**

- $f_i: (m + rt)/n$

-

- Where:
- f_i : influence of user in the circle
- m : Total mentions of the user
- rt : total re-tweets of the user's content by people
- n : total number of individuals in the circle

Methodology

- **For Defining Influence we have considered the following arrays:**
 - Array containing the influence value in each circle defined: A_i
 - Array containing weights of each influence circle: A_w
 - Array containing value of n for each circle: A_n
- **The total absolute influence (I_a) of an individual throughout our network can be found using:**

$$I_a = \frac{\sum_{j=1}^N A_i[j] * A_n[j] * A_w[j]}{\sum_{j=1}^N A_n[j] * A_w[j]}$$

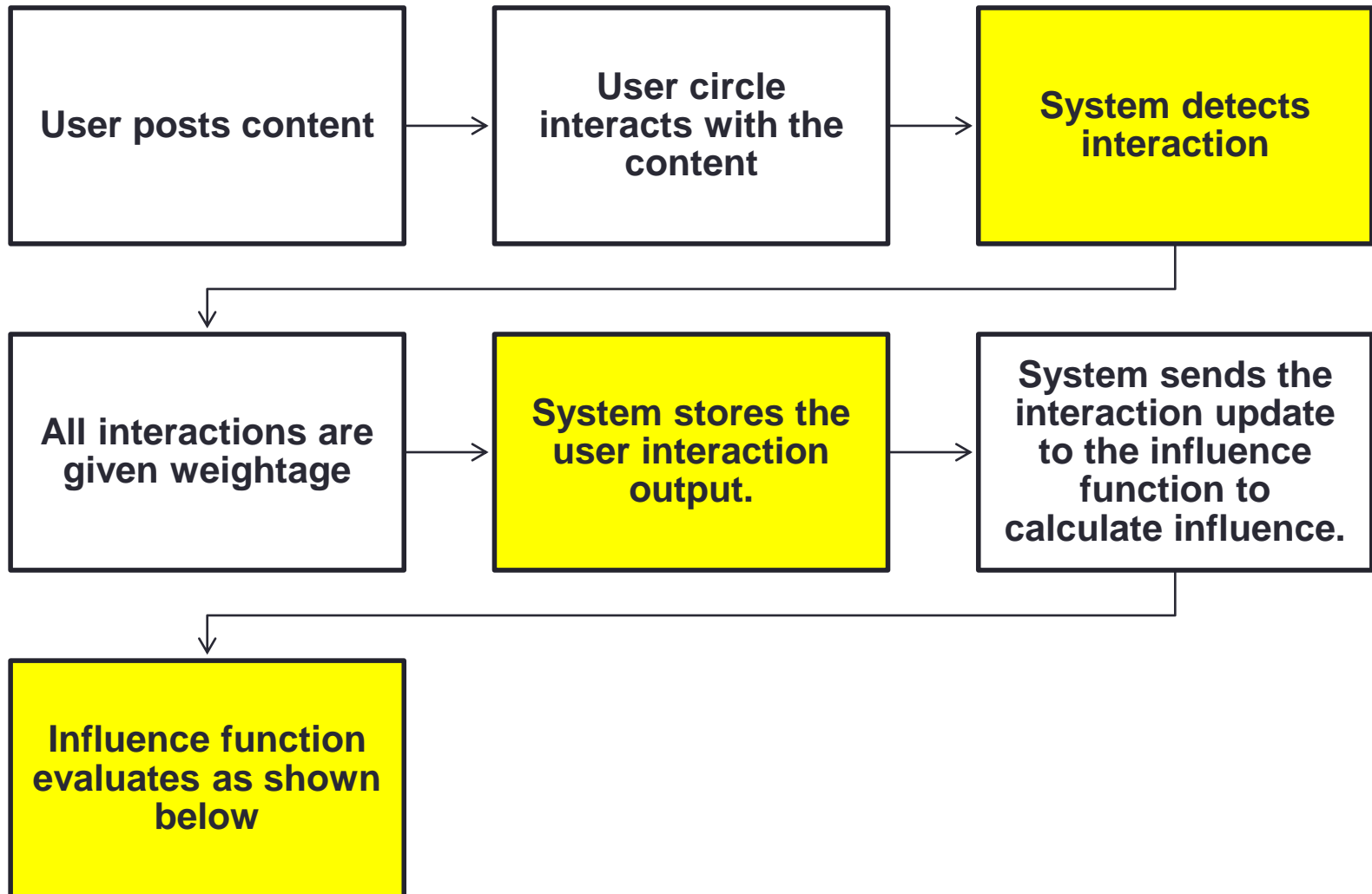
Methodology

- Overall Influence of user on the network (I_o) which will be given by the formula:

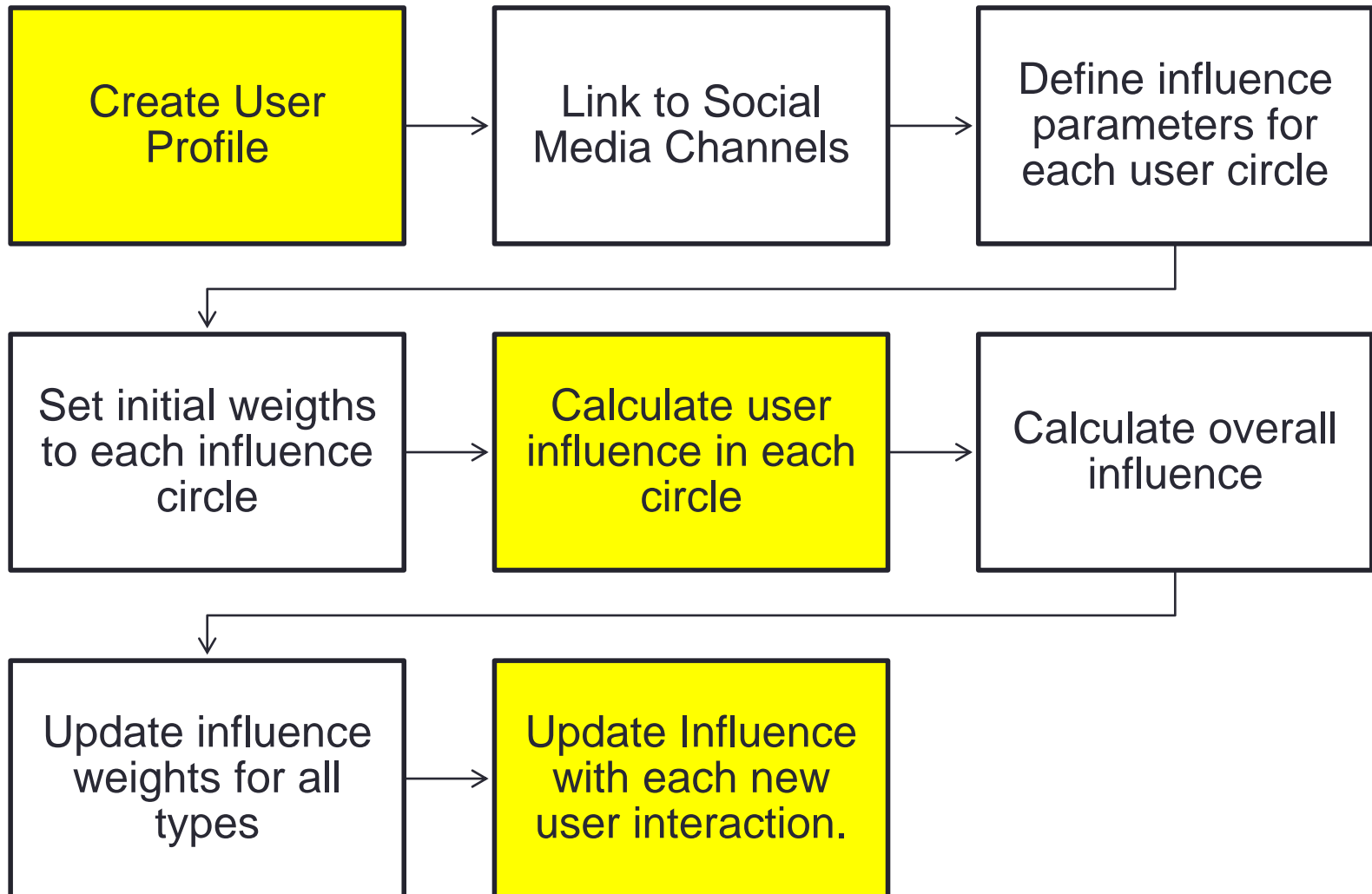
$$I_o = \frac{I_a - I_{a,min}}{I_{a,max} - I_{a,min}}$$

- $I_{a,max}$: Maximum absolute influence in the network
- $I_{a,min}$: Minimum absolute influence in the network

User Interaction



Flow Chart



Solving the given case

Name	Type of Action	Total Impression	Total Followers	Influence Value (f_a)
Jason	Positive	$20+10 = 30$	500	$30/500 = 0.06$
Mark	Negative	$15+5= 20$	1000	0.02
Ashley	Negative	$80+40= 120$	500	0.24

Positive Influence Ranks	Overall Influence Ranks
Jason [0.06]	Ashley [0.24]
Mark [0.02]	Jason [0.06]
Ashley [0.24]	Mark [0.02]

Solving the given case

- Jason has highest positive influence in his network
- Mark has a negative influence but his influence is very weak in the network
- Ashley has a very high negative influence in her network
- Ashley has the highest influence amongst all in the network
- Overall influence doesn't depend on positive or negative influence

Applications

Identifying top consumers to market to

Deriving inference on public sentiments on political and social issues

Improving gamification platforms for business

Designing better Human Resources and Recruitment tools for organizations

Possible improvements

Add functionality for Facebook and other social media channels

Define weightage functions for improving weights as different types of posts can have different influence

Add functionality for considering more impressions and interactions like comments, likes

Data can be analyzed in terms of specific terms or values like: trending topics, recent happenings and for each type there can be an influence metric.

Overall influence for each type of parameter can be also listed down like keywords, location, and age group.

QnA

Thank You =)