UPPSALA UNIVERSITET

Data Mining Project

Identifying Coordinated Accounts on Social Media through Hidden Influence and Group Behaviors

Salman Khan, Esraa, Sakib

UPPSALA UNIVERSITET

The problem

- **Disinformation** campaigns on social media, involving coordinated activities from malicious accounts towards **manipulating public opinion**, have become increasingly prevalent.

- Existing approaches to detect coordinated accounts either **make very strict assumptions** about coordinated behaviors, or **require part of the malicious** accounts in the coordinated group to be revealed in order to detect the rest

Solution

Generative model, **AMDN-HAGE** (Attentive Mixture Density Network with Hidden Account Group Estimation)

Jointly models account activities and hidden group behaviors based on **Temporal Point Processes** (TPP) and **Gaussian Mixture Model** (GMM),

To capture inherent characteristics of coordination which is, accounts that coordinate must strongly influence each other's activities, and collectively ap- pear anomalous from normal accounts.

UPPSALA
UNIVERSITET

Assumptions

Characteristics of coordination:

Strong hidden influence. If accounts coordinate to amplify social media posts ,there should be a strong hidden (latent) influence between their activities.

Compared to normal accounts, the **number of coordinated accounts** is quite **small**

Highly concerted activities. The collective behaviors of coordinated accounts should be collectively anomalous, from other normal accounts on the network with less organized activity patterns

What is AMDN-HAGE?

learn the latent interactions

Highly concerted activities

Model the distribution of future activities conditioned on past activities of all accounts

By **Neural Temporal Point Processes (NTPP)**

Jointly capture collective anomalous behavior by simultaneously learning the group membership of accounts.

By **Gaussian Mixture Model (GMM).**

Task Definition

Activity traces: a sequence of events ordered in time, which can be formulated as

Cs = [(u1,t1),(u2,t2),(u3,t3),⋯(un,tn)] account ui at time ti.

Hidden Account Group : supposing that there are N groups in the account set U , we can define a membership function M :$U \rightarrow$ {1,⋯ ,N}, which projects each account ui to its group index.

AMDN-HAGE Architecture

consists of two components:

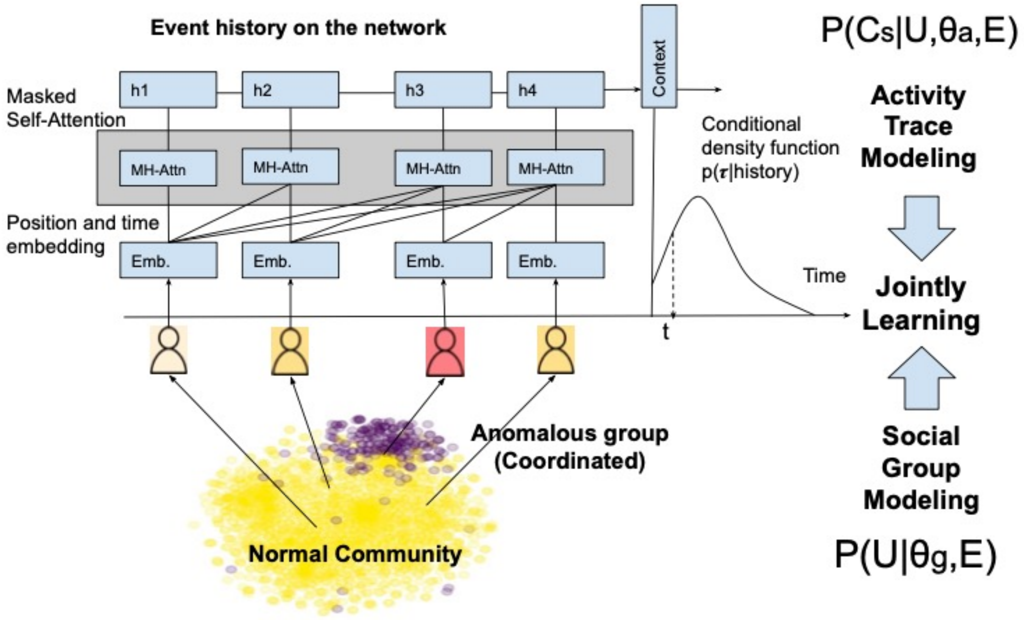
Attentive Mixture Density Network (AMDN) that models observed activity traces as a temporal point process

Hidden Account Group Estimation (HAGE) component that models account groups as mixture of multiple distributions.

Both share the **account embedding layer** and reflect the complete generative process

that the **accounts are first drawn from multiple hidden groups** and then **interact with each other so that activity traces are observed.**

Using the observed activity traces, we can learn the generative model by **maximizing the likelihood function** of **the joint model**, and acquire not only account embedding but also a activity trace model and group membership function.



AMDN-HAGE Model

Joint likelihood function can be written as:

$$\log p(Cs, U; \theta g, \theta a, E) = \log p(Cs|U; \theta g, \theta a, E) + \log p(U; \theta g, \theta, E)$$

$$= \log p\,(Cs\,|U\,;\,\theta a,\,E) + \log p\,(U\,;\,\theta g,\,E)$$

Probability density that the activity traces are observed given a known account set.

Probability density that we observe the account set drawn from the latent hidden social groups

Attentive Mixture Density Network (AMDN)

Temporal Point Process (TPP): stochastic process whose realization is a sequence of discrete events in continuous time

history of events H_t **= {(**ui**,**ti**)|**ti **<** t**,**ui **∈ U}**

Neural Temporal Point Process (TPP): Neural Network Encoder and Decoder

Encoder: For interpretable influence of past event on future events, we encode the event sequence with masked self-attention

Decoder: Conditional probability density function. With the encoded event history (context vector), the event decoder (learnable conditional density function $p\ (\tau\ |H\tau\ )$) is used to generate the distribution of the next event time conditioned on the history.

While we can choose any functional form for $p\ (\tau\ |H\tau\ )$, the only condition is that it should be a valid PDF (non-negative, and integrate to 1 over $\tau \in$ R+).

UPPSALA UNIVERSITET

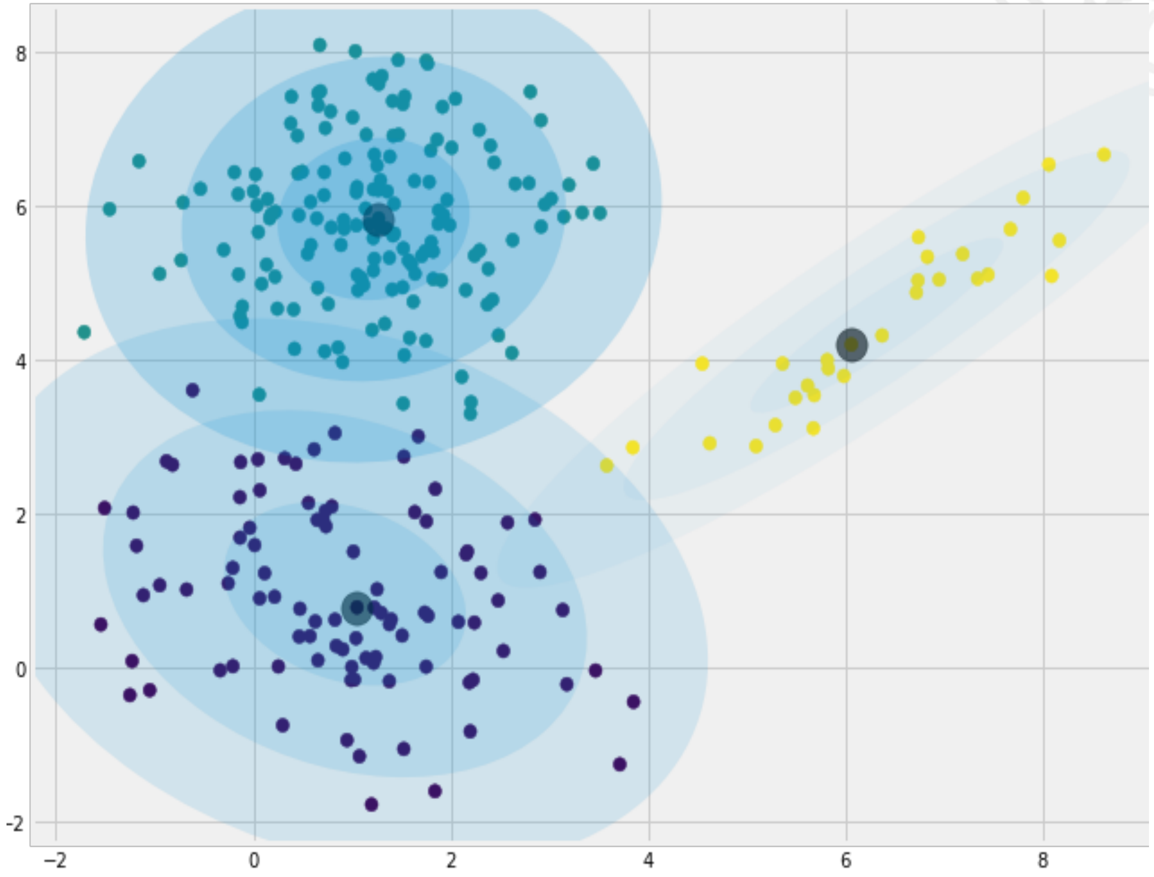
Attentive Mixture Density Network (AMDN)

Encoder and decoder architecture models activities with likelihood $p\ (Cs\ |U\ ;\ \theta a,\ E)$ factorized as:

$$\log p(C_s|U;\theta_a, E) = \sum_{i=1}^{L} \left[\log p_{\theta_a,E}(t_i|H_{t_i}) + \log p_{\theta_a,E}(u_i|H_{t_i})\right]$$

Modeling Hidden Groups(HAGE)

- Gaussian Multivariate Distributions (GMM)
- Hidden Group Estimation.
- Difference from general Gaussian mixture models.



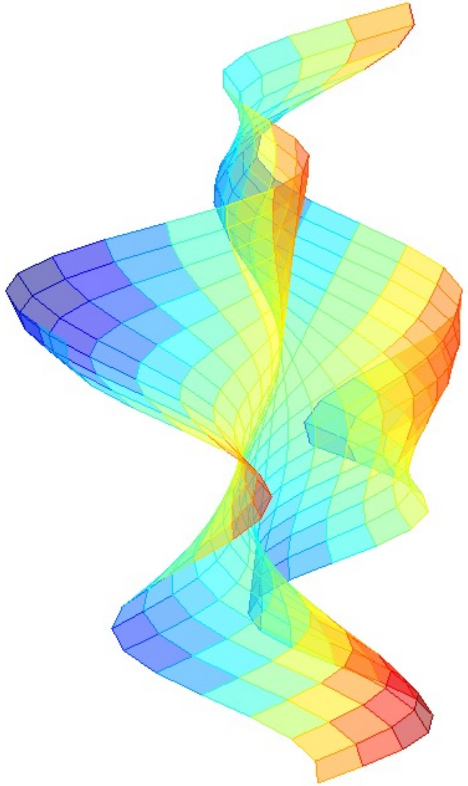
UPPSALA UNIVERSITET

Training Algorithm for AMDN-HAGE

Require: Activity traces (Cs), Account set (U)

Ensure: Generative model (θa, θg and E)

$\theta_a^{(0)}, E^{(0)} \leftarrow \text{argmax}_{\theta_a, E} \log p(C_s|U; \theta_a, E)$
Set i as 1 {Iteration index}.
while not converged **do**
    $\theta_g^{(i)} \leftarrow \text{argmax}_{\theta_g} \log p(U; E^{(i-1)}, \theta_g)$ using EM algorithm
    $\theta_a^{(i)}, E^{(i)} \leftarrow \text{argmax}_{\theta_a, E} \log p(C_s, U; \theta_g^{(i)}, \theta_a, E)$ using SGD or its variants
    $i \leftarrow i + 1$.
end while



A generic concern to such alternating optimization algorithm is its convergence

Jointly Learning

- Issues with directly use stochastic gradient descent (SGD) or its variants like ADAM.

  -> Leading to invalid log-likelihood in training.

- Bi-level optimization.

  -> Use of Expectation–Maximization algorithm

  -> use of SGD (or its variant)

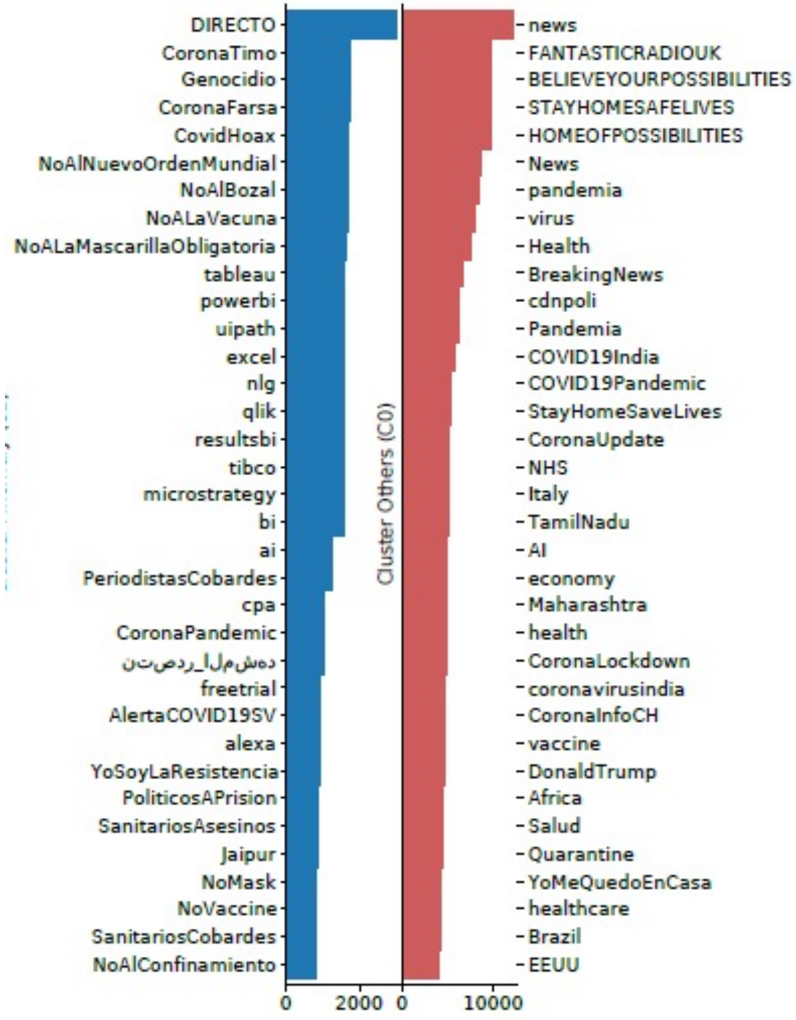
EXPERIMENT RESULTS

- Data Collection - Social media posts from with keywords related to COVID-19.

- Uncovering coordinated groups in COVID-19 data.

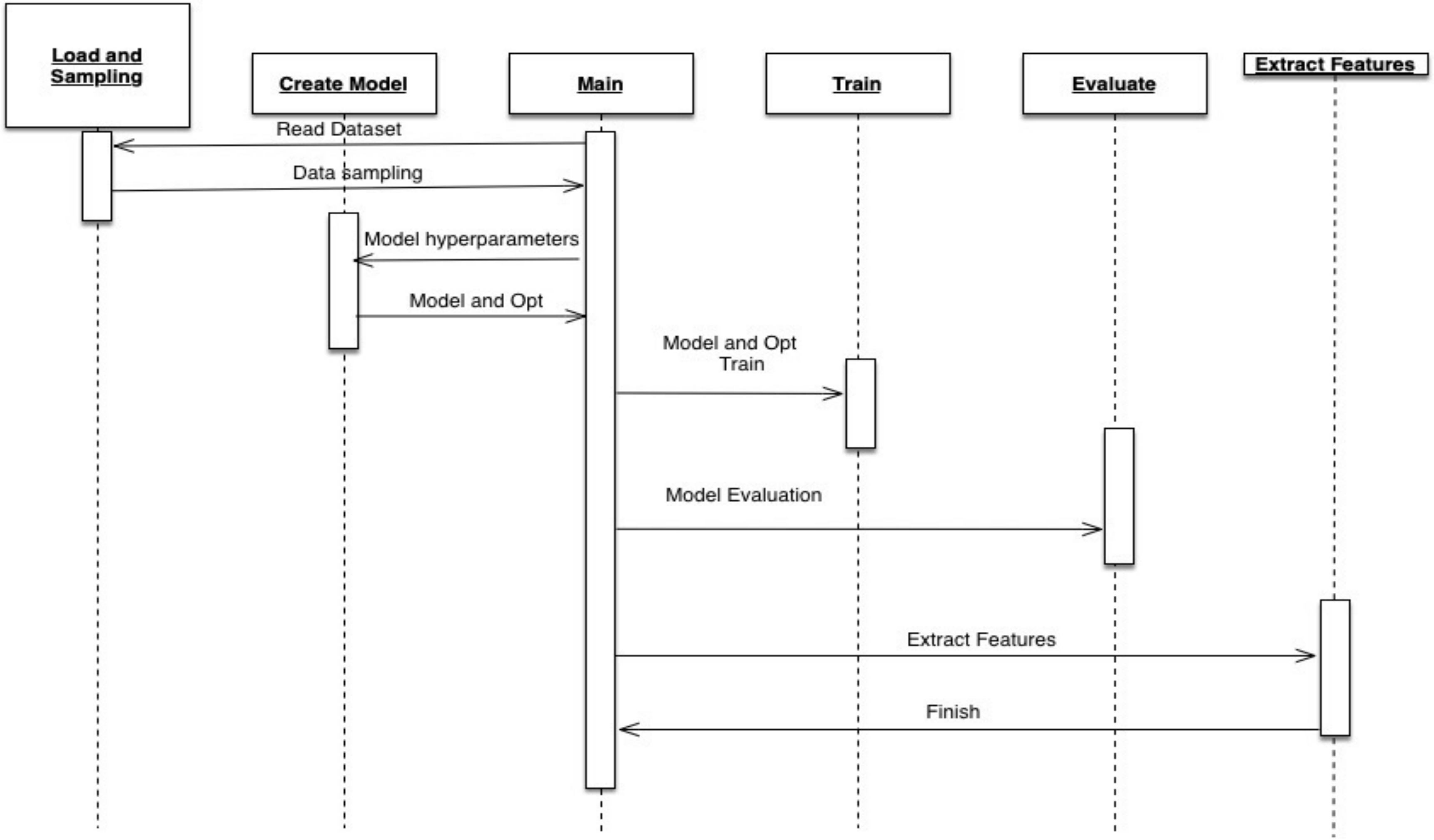
  -> AMDN-HAGE method identifies two clusters.

- In Fig, we find most frequent hash-tags in tweets posted by accounts in the groups, and plot the top hash-tags unique to each group



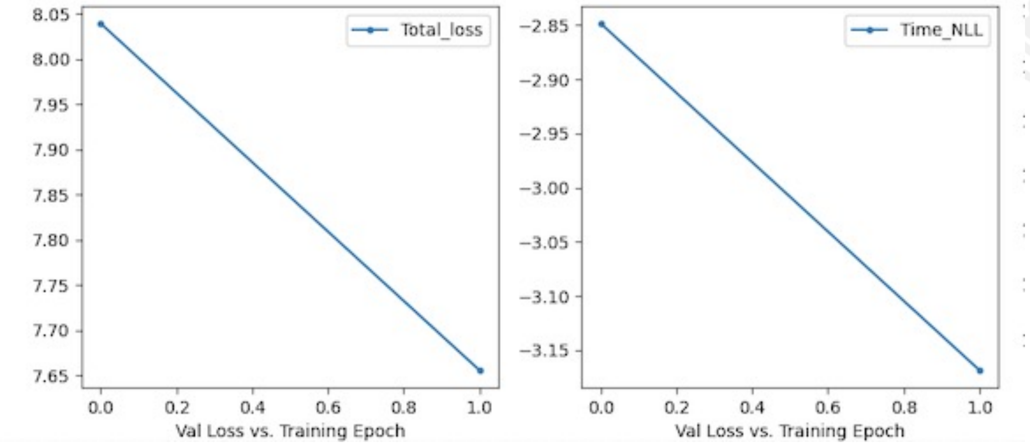
UPPSALA UNIVERSITET

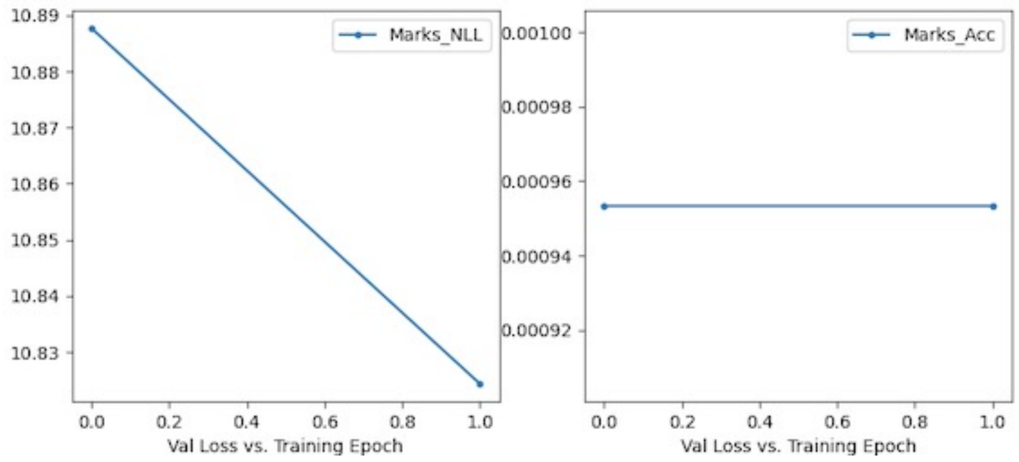
Code Explanation



UPPSALA UNIVERSITET

Training Result





UPPSALA
UNIVERSITET