Embeddings and Time Series

GitHub Link: https://github.com/Pritam0705/MACS-40123/blob/main/ITR4/DTM.ipynb

Google Colab Link (for all outputs of the cells):
https://colab.research.google.com/drive/1YBy1VNF1MnoWa_TAHplQ8o0kX7d0784z?usp=sharing

Part 1: Analysis of Peer-Reviewed Papers

1. **Topic evolution analysis of radar research using a dynamic topic model based on latent Dirichlet allocation**
   Author: Xiaoguang Huang and Hui Fang
   Published: CISAT 2021
   Summary: The paper presents a comprehensive study on the evolution of research themes within radar technology. The authors employ a Dynamic Topic Model (DTM), an advanced extension of Latent Dirichlet Allocation (LDA), to explore how radar research topics have changed over time. The authors applied DTM to a dataset of radar research publications, organizing them into chronological slices to understand the temporal evolution of topics. This enabled identifying emerging research areas gaining prominence, declining topics that have lost focus, and persistent themes that have remained consistently relevant. Their analysis provides insights into key trends within the radar research community, mapping how the field has developed and adapted to new challenges and technologies.

2. **Fake News Detection Using Time Series and User Features Classification**
   Author: Previti, M., Rodriguez-Fernandez, V., Camacho, D., Carchiolo, V., Malgeri, M.
   Published: Applications of Evolutionary Computation, 2020
   Summary: The paper introduces a methodology for identifying fake news on online social networks (OSNs) by analyzing the temporal dynamics of information dissemination and user-related features. The authors categorize existing fake news detection approaches into content-based and context-based methods. Content-based approaches focus on linguistic cues within the text, such as pronoun usage, conjunctions, and emotionally charged words, to differentiate between true and false news. In contrast, context-based approaches examine external factors like user interactions and the propagation patterns of news across networks. In their methodology, they classify features extracted from time series data that represent

the temporal evolution of rumors—defined as collections of tweets about the same topic—and user information associated with each rumor. This involves analyzing how rumors spread over time and the characteristics of users who engage with them.

3. **Dynamic Topic Modeling to Mine Themes and Evolution during the Initial COVID-19 Vaccine Rollout**
   Author: Agarwal, Ankita; Patel, Dixit B; Burwell, Emily; Romine, William L.; Banerjee, Tanvi
   Published: Health Behavior and Policy Review, 2023
   Summary: examines public discourse on COVID-19 vaccines by analyzing Twitter data from December 2020 to March 2021. Utilizing dynamic topic modeling, the authors identified prevalent themes and their evolution over time, revealing key topics such as vaccine distribution, side effects, public sentiment, and policy discussions. The analysis highlighted shifts in public focus, with initial concerns about vaccine availability transitioning to discussions on efficacy and adverse effects as the rollout progressed. The study underscores the utility of dynamic topic modeling in understanding public concerns and information dissemination during health crises, providing valuable insights for policymakers and health communicators to address public sentiment and misinformation effectively.

Part 2: Interpretation

1. **BERTopic**
   BERTopic provides a framework for extracting topics from textual data by leveraging contextual embeddings (e.g., BERT) and clustering techniques. For the misinformation dataset, the topics extracted provide insights into **thematic clusters of claims** based on shared semantic content.
   - BERTopic identified distinct topics like "party_minister_congress_modi" and "woman_hindu_muslim_communal," reflecting political narratives, communal issues, and gender-based misinformation.
   - The clustering of semantically similar claims allows for a better understanding of how misinformation is framed and targeted toward specific societal issues.
   - Topics span multiple domains, including politics, communal tensions, health crises (e.g., COVID-19), and protests. This shows the diverse landscape of misinformation targeting **emotional triggers** and **polarized events**.

2. **Dynamic Topic Modeling (DTM)**
   Dynamic topic modeling tracks the evolution of topics over time, revealing how misinformation themes gain or lose traction in response to real-world events.
   - Topics like "farmers protest" and "election" exhibit spikes corresponding to significant events, such as state elections, national protests, or public health emergencies. This indicates the event-driven nature of misinformation, where narratives are crafted to exploit current socio-political contexts.
   - Event-Specific Misinformation: COVID-19-related misinformation peaked sharply in early 2020, reflecting the global uncertainty at the start of the pandemic. Similarly, farmers' protests saw a peak in early 2021, corresponding to widespread mobilization and debate around farm laws.
   - Recurring Narratives: Some topics, such as "woman_hindu_muslim_communal," show periodic spikes, indicating the reuse of narratives around communal tensions and gender-based issues during politically sensitive periods (e.g., elections).

Summary

BERTopic excels at uncovering the thematic **diversity of misinformation**, while dynamic topic modeling provides a timeline for **when and why misinformation peaks**. Together, these approaches offer a comprehensive view of how misinformation spreads and evolves, enabling targeted interventions and a deeper understanding of its societal impact.

1. **What worked:**
   - BERTopic effectively identified distinct, interpretable themes such as political narratives, communal tensions, and gender-based misinformation.
   - The semantic embeddings captured nuanced relationships between claims, providing high coherence within topics.
   - DTM successfully revealed **temporal trends** in misinformation, showing how topics evolved in response to real-world events like elections, protests, and the pandemic.
   - DTM provides insights into **persistent vs. transient topics**, showing which narratives have long-lasting impact.
   - Using BERTopic for thematic clustering and Dynamic Topic Modeling for temporal analysis provided a comprehensive view of misinformation's content and timing.
   - Clear identification of recurring and event-specific narratives.

2. **What didn’t work:**

- Some clusters were too broad or lacked granularity, combining unrelated subtopics (e.g., merging political and communal themes in some cases).
- Limited ability to handle overlapping topics, where certain claims were relevant to multiple themes (e.g., communal violence during elections).

3. **Future revisions:**
    - Explore advanced preprocessing techniques, such as domain-specific stopword lists and contextual lemmatization, to enhance the quality of embeddings and clusters.
    - Combine BERTopic and dynamic topic modeling with **fine-tuning** to improve coherence and temporal adaptability
    - Implement **hierarchical clustering** or **sub-topic detection** in BERTopic to break down broad clusters into more interpretable subtopics.
    - Use **fine-grained time bins** for more granular temporal analysis.

Part 3: Social, Cultural, and Behavioral Implications of Findings

Social Implications

Erosion of Trust in Institutions: Misinformation targeting elections, political parties, and government policies undermines trust in democratic institutions and governance. Persistent narratives about electoral fraud or leadership controversies create skepticism about the integrity of democratic processes.

Polarization of Communities: Topics like communal narratives (e.g., Hindu-Muslim relations) and gender-based misinformation deepen societal divisions and reinforce stereotypes. Misinformation amplifies emotional and ideological divides, making reconciliation and dialogue between polarized groups increasingly challenging.

Cultural Implications

Exploitation of Cultural Symbols: Narratives around religious icons (e.g., Ayodhya, Ram Temple) and cultural events are weaponized to provoke emotional responses and inflame cultural tensions. This demonstrates how misinformation exploits culturally significant symbols to manipulate public sentiment and shape political discourse.

Reinforcement of Stereotypes: Misinformation about gender-based violence or communal tensions perpetuates harmful stereotypes about certain groups (e.g., associating specific communities with violence or backwardness). This normalization of biased narratives impacts the perception of minority groups and marginalized communities.

Behavioral Implications

Emotional and Cognitive Manipulation: Misinformation leverages emotionally charged themes (e.g., religion, gender, political allegiance) to drive engagement and influence behavior. This manipulates public decision-making, from voting patterns to participation in protests or communal violence.

Resistance to Fact-Checking: While fact-checking initiatives exist, the persistence of certain misinformation topics reflects **cognitive dissonance**—people resist changing their beliefs even when presented with verified information. This behavioral resistance complicates efforts to combat misinformation effectively.