Project A

Project Plan for Topic Modeling Implementation

Chen Xi, Yiran Li, Chengcheng Cai, Jingyi Liu, Jiayi Wang

**Objective:**

To identify meaningful topics in a collection of news articles and track their evolution over time using LDA, BERTopic, and Prompt+LLM models.

**Project Goal:**

To have a robust, adaptable topic modeling system that provides insightful analysis of news article topics over time.

| Phase | Task | Timeline |
| --- | --- | --- |
| Data Preparation | * Collect a static corpus of news articles. * Clean and preprocess the text (tokenization, stopword removal, stemming/lemmatization). | Weeks 1-4 |
| Model Selection | * LDA for baseline topic modeling. * BERTopic for advanced topic modeling using transformers. * Prompt+LLM for leveraging large language models in identifying and generating topic insights. | Weeks 4-6 |
| Model Training and Tuning | * Train the LDA model on the corpus to establish baseline topics. * Train BERTopic to extract topics using deep contextual embeddings. * Design prompts for LLM to generate topics and insights. | Weeks 7-8 Midterm Presentation |
| Integration and Iteration | * Integrate the outputs of all models for a comprehensive topic analysis. * Iterate over the models' parameters and training data to improve results. | Weeks 9-10 |
| Evaluation and Adaptation | * Evaluate the coherence and relevance of the topics generated. * Adjust models and refine prompts based on feedback. | Week 11 |
| Visualization and Tracking | * Develop a timeline visualization of topic progression. * Implement tracking mechanisms to monitor topic evolution. | Week 12 |
| Deployment | * Set up a Streamlined interface for model interaction and visualization. * Ensure the system can update with new data, reflecting the latest trends. | Weeks 13-14 Final Presentation |