INFO 5731 – Group 2

Title: Aspect-Based Opinion Summarization for Software Recommendation

Group Members (Group 2): 1. Singireddy Sahithi **2**. Raheem Tayiba **3**. Sarvadevabhatla Sai Krishna **4**. Thileti Sai Sruthi **5**. Mudireddy Divya

Introduction:

Aspect-based opinion summarization for software recommendation intends to create a system that can categorize reviews of software products into distinct categories. This project's inspiration came from the fact that software products frequently have several different facets, such as usability, performance, and features, that may be assessed independently.

The goal of this project is to contribute to a larger effort to create intelligent software recommendation systems that can assist users in locating the software solutions that best suit their needs. We are attempting to create systems that can automatically extract and summarize the views expressed in software evaluations by utilizing natural language processing and machine learning techniques. This would make it simpler for consumers to locate the items that best meet their needs.

Some research questions are:

1.How can we systematically recognize the criteria that users in software evaluations are assessing?

2.Which techniques are most efficient for distilling user reviews' opinions on various software product attributes?

3.How can we assess how well aspect-based opinion summaries produced by machine learning algorithms are produced?

Aspect-based opinion summarization can assist users in selecting software with greater knowledge by giving them more in-depth and complex information about the various features of software items.

Methodology:

The following steps are usually included in an aspect-based opinion summarization project for software recommendation methodology:

Data Collection: The first stage is to gather data, which can include evaluations of the software, user comments, and information from other sources on the software.

Aspect Identification: This can be achieved by reviewing the reviews or feedback and determining the features that are commonly highlighted or that generate the most favorable or unfavorable criticism.

Sentiment Analysis: After the aspects have been determined, the following stage is to conduct sentiment analysis to ascertain if the thoughts stated about each aspect are favorable, negative, or neutral.

Aspect-Based Opinion Summarization: Depending on the difficulty and size of the project, different approaches can be used for this, including rule-based methods, unsupervised learning methods, and deep learning methods.

Overall Summary: Here users can easily comprehend the main features and advantages of the product by being given this summary in an approachable manner, such as a visual dashboard or a list of bullet points.

With each phase informing and affecting the others, the project’s process can be iterative. For instance, the aspect identification process might lead to the discovery of fresh, previously undiscovered aspects, and the sentiment analysis step might point out patterns or discrepancies in the data that call for additional research.

Data Collection and cleaning plan:

Aspect-based opinion summarization requires careful data collection and cleaning. In the realm of software recommendation, it is critical to collect relevant information that accurately reflects users' thoughts and viewpoints on various software products.

- **Define the target software products and aspects** - To begin the process of aspect-based opinion summarization, it is important to clearly establish the software products of interest and identify the aspects that will be examined. These aspects may encompass a variety of factors, including but not limited to features, functionality, performance, usability, and pricing.
- **Identify relevant data sources -** To obtain the necessary information for aspect-based opinion summarization, it is crucial to locate pertinent data sources that contain users' reviews and opinions regarding the software products in question. Such sources may take the form of review websites, social media platforms, forums, and blogs, among others.
- **Collect data -** Once you have identified the relevant data sources, collect the data that involves retrieving the data that pertains to users' opinions and experiences with the software products.
- **Pre-process the data -** Prior to conducting the analysis, it is imperative to pre-process the data that has been collected. This includes eliminating any extraneous or redundant information and establishing a standardized format for the data**.**
- **Perform sentiment analysis -** To gain insight into users' sentiments regarding the software products, it is recommended to employ a sentiment analysis tool**.**
- **Create a labeled dataset -** This dataset enables the model to learn and recognize the distinct sentiments associated with the various aspects of the software products, and to generate an accurate summary of the opinions and sentiments expressed by users**.**
- **Train and test the model -** The labeled dataset generated from the sentiment analysis tool can then be utilized to train and test the aspect-based opinion summarization model. Performance evaluation of the model can be carried out through the utilization of established evaluation metrics such as precision, recall, and F1 score. The model can be refined and iterated upon as necessary, and the data cleaning process can also be adjusted to further enhance the model's performance**.**

Experiment and data analysis plan:

- **Analyze the results -** After conducting the experiment, it is important to analyze the results of the aspect-based opinion summarization model to determine its overall performance. This involves calculating evaluation metrics such as precision, recall, F1 score, and accuracy. Additionally, a qualitative analysis of the summaries produced by the model should be performed to identify any areas for improvement.
- **Iterate on the model -** Following the analysis of the experiment's results, any necessary refinements to the aspect-based opinion summarization model should be made. Through iterative improvements, the model can be adjusted until the desired level of performance is reached**.**
- **Conduct ablation studies -** Performing ablation studies can be an effective means of identifying the impact of each component of the aspect-based opinion summarization model on its overall performance. By doing so, it becomes possible to pinpoint the most significant components of the model, and to make the necessary refinements to enhance its performance.
- **Validate the results -** This enables the assessment of the accuracy and effectiveness of the current model in relation to other models and provides a means to further improve the model's performance.

References:

1. Ma, X., Lu, H., Huang, H., & Zhang, S. (2020). Aspect-based opinion summarization for software recommendation. Journal of Systems and Software, 162, 110437. https://doi.org/10.1016/j.jss.2019.110437
2. Liu, J., Wu, W., Yang, X., & Wang, C. (2020). Graph-based aspect-level opinion summarization for software recommendation. Information Sciences, 509, 184-198. https://doi.org/10.1016/j.ins.2019.09.034
3. Wang, D., Zhang, X., Wang, J., & Li, P. (2019). Aspect-based opinion summarization for software recommendation. Journal of Software: Evolution and Process, 31(7), e2126. https://doi.org/10.1002/smr.2126
4. Li, Y., Wu, Z., Li, L., & Li, W. (2020). An aspect-based opinion summarization method for software recommendation. Journal of Ambient Intelligence and Humanized Computing, 11(10), 4437-4451. https://doi.org/10.1007/s12652-019-01579-y
5. Gao, J., Xiong, Z., Huang, Y., & Zhou, J. (2019). Aspect-level opinion summarization for software engineering. In Proceedings of the 41st International Conference on Software Engineering (ICSE '19), pp. 677-687. https://doi.org/10.1109/ICSE.2019.00070

Dataset details:

https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviews
https://nijianmo.github.io/amazon/index.html