2024-2 Unstructured Data Analysis
2024.11.18

Sentiment Analysis

Week 12 – Code Practice

Professor: Misuk Kim
Teaching Assistant: Minjoo Son
minjoo77@hanyang.ac.kr

HANYANG UNIVERSITY

Sentiment Analysis
Week 12 – Code Practice

Contents

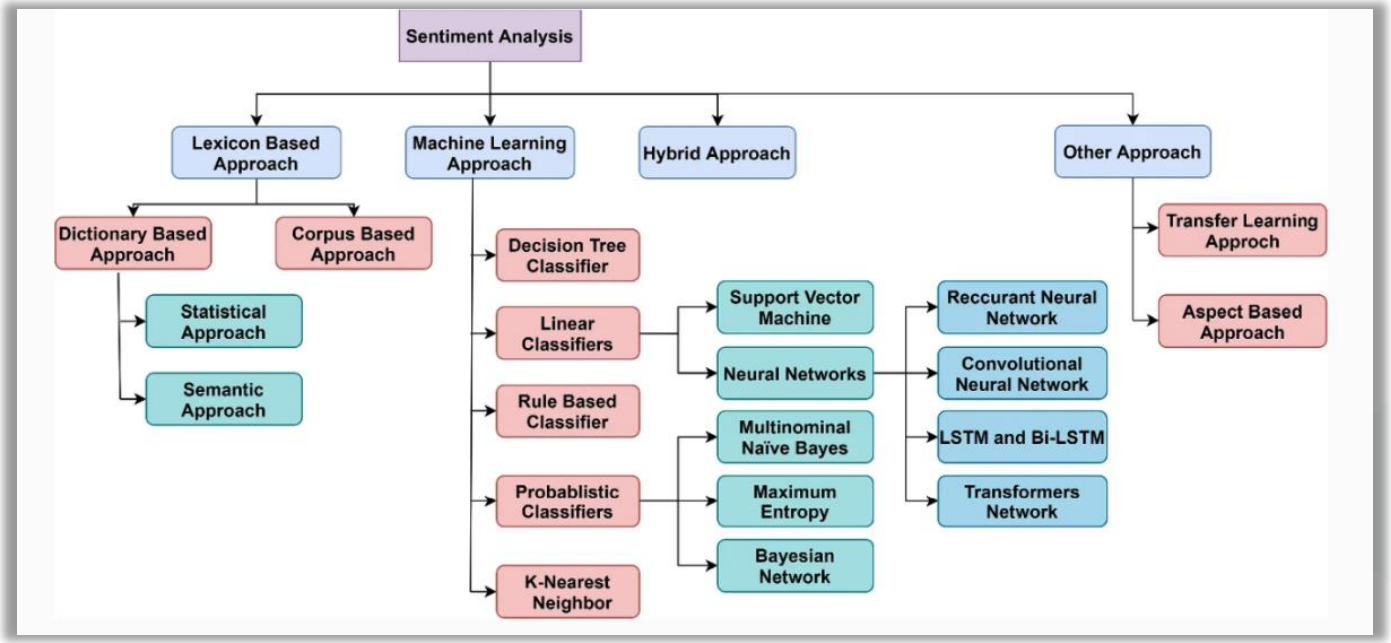
1. Introduction

❖ Week 12 Objective

- Sentiment Analysis
  - Lexicon-based
  - Machine learning-based through training

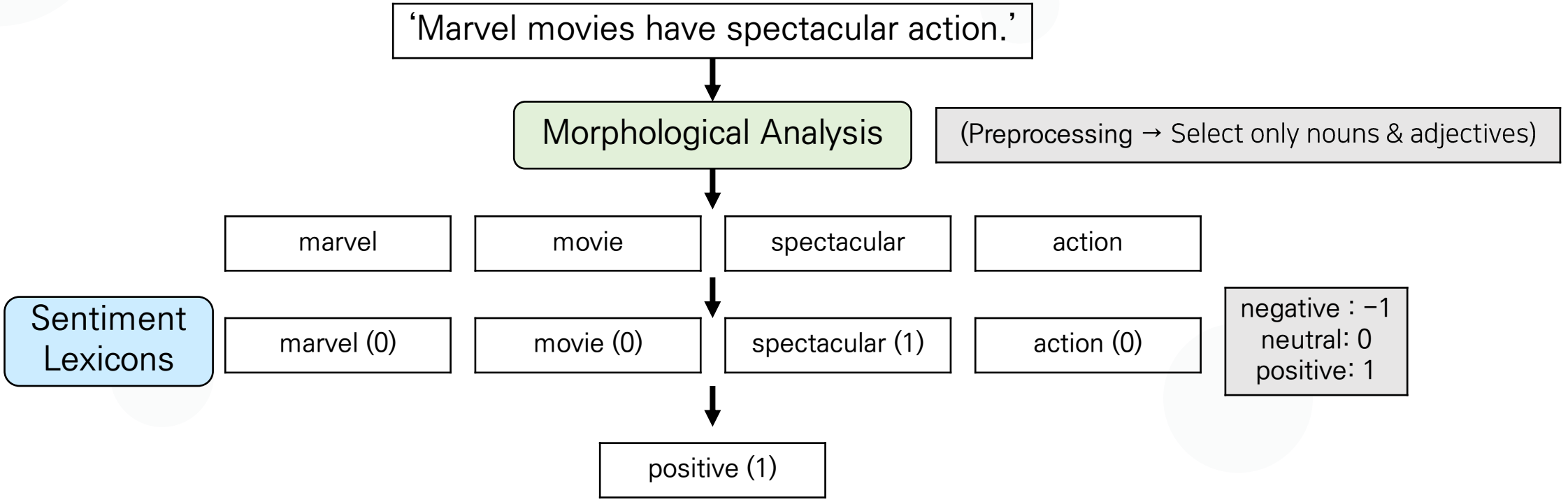
❖ Understanding Sentiment Analysis

- Sentiment analysis refers to the process of analyzing subjective information such as **opinions, evaluations, and attitudes expressed in text**.
- Sentiment is a subjective emotion towards a particular subject, usually classified into three categories: positive, neutral, and negative.
- There are various methodologies for sentiment analysis, which can generally be categorized into lexicon-based analysis and machine learning-based analysis.



❖ Lexicon-based Sentiment Analysis

- This method involves creating a sentiment lexicon by assigning positive or negative sentiments to words like nouns, adjectives, and verbs. Sentiment analysis is then performed using this lexicon.
- While lexicons can be built manually, external sentiment lexicons are often used due to the time and effort required for manual creation.



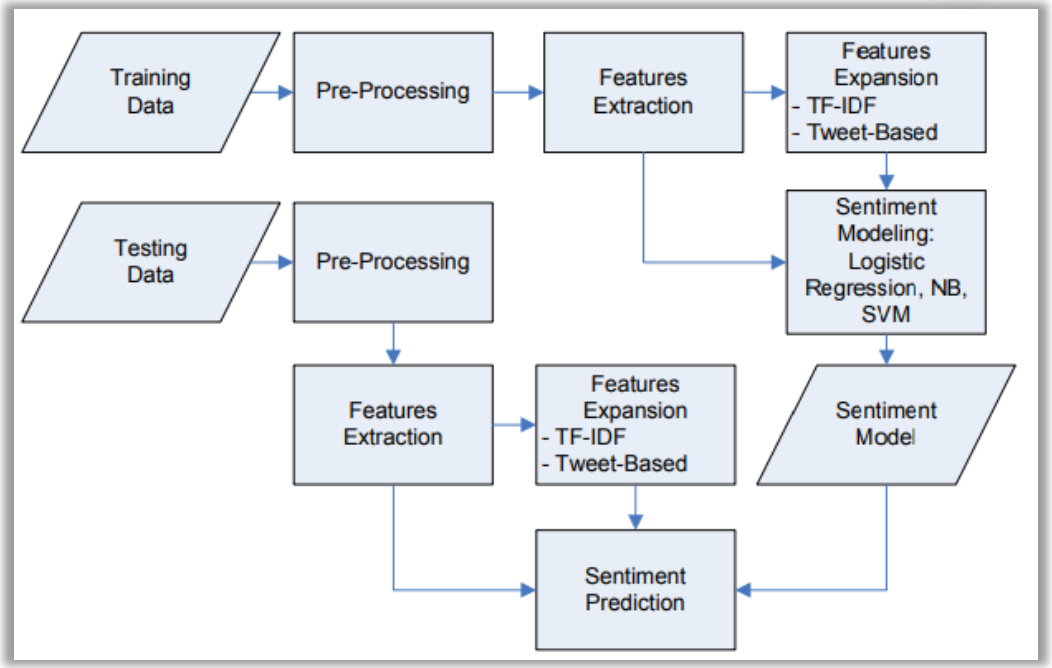
2. Sentiment Analysis

❖ Lexicon-based Sentiment Analysis

- Code Practice
- Codes on Github: https://github.com/ming9oori/Unstructured-Data-Analysis

❖ Sentiment Analysis Based on Machine Learning through Training

- To perform sentiment analysis using machine learning, **a labeled training dataset is needed.** For example, if analyzing movie reviews, the dataset should pair each review with a sentiment label.
- This is a document classification problem with classes like 'positive', 'negative', and 'neutral'.
- With the dataset, various classification algorithms can be used to train a model to predict sentiment.



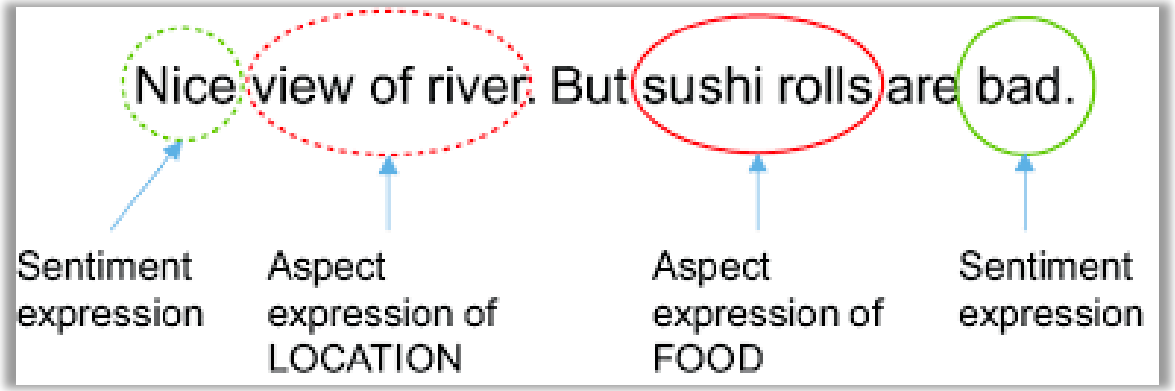
2. Sentiment Analysis

❖ Sentiment Analysis Based on Machine Learning through Training

- Code Practice
- Github repository: https://github.com/ming9oori/Unstructured-Data-Analysis

❖ Advanced: Aspect-Based Sentiment Analysis (ABSA)

- **Definition**: Classifying sentiment towards specific aspects (topics) in a text.
- **Importance**: A single text may contain multiple aspects, each requiring independent sentiment analysis.
- **Challenge**: Accurately aligning sentiment expressions with their corresponding aspects for effective opinion mining.



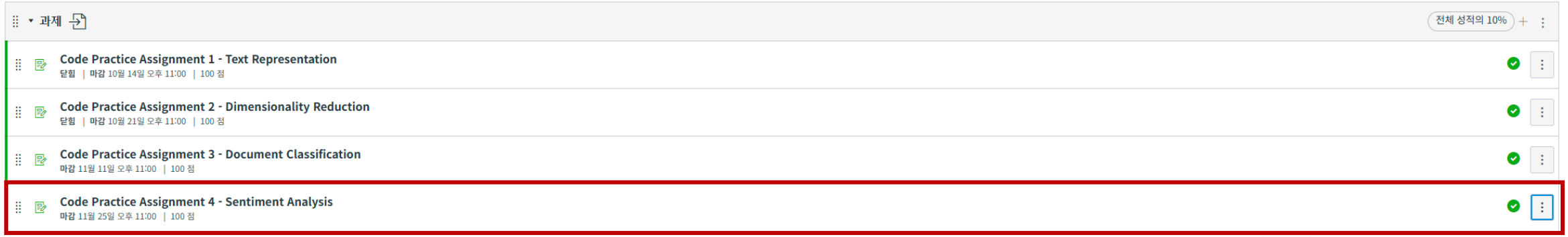
2. Sentiment Analysis

❖ Advanced: Sentiment Analysis using Pre-trained Language Models

- Reference link: https://medium.com/@manjindersingh_10145/sentiment-analysis-with-bert-using-huggingface-88e99deeec9a

❖ Assignment 4 – Sentiment Analysis

- Choose only one of the two tasks below, complete it, and save the files as 'Assignment_YourName_YourStudentID.ipynb' and submit it to the 'Code Practice Assignment 4 – Sentiment Analysis' section under Assignments.
  1) Download the **'Week 12 – Sentiment Analysis – Assignment Skeleton Code.ipynb'** from the GitHub page and perform machine learning-based sentiment analysis on the IMDB review dataset.
     - Github repository: https://github.com/ming9oori/Unstructured-Data-Analysis
  2) Follow the instructions in the link below to perform **Sentiment Analysis with BERT using Huggingface**.
     - Github repository: https://medium.com/@manjindersingh_10145/sentiment-analysis-with-bert-using-huggingface-88e99deeec9a



Sentiment Analysis
Week 12 – Code Practice

2024-2 Unstructured Data Analysis
2024.11.18

Q & A

Thank you for your attention. Any questions are welcome!

Minjoo Son

HANYANG UNIVERSITY