In The Name of God

Faculty of Electronics and Computer Engineering

Statistical Pattern Recognition – Dr. AzimiFar

Proposal

From Mohammad Ali Khodayari and Bahareh Gholami for statistical pattern recognition's final project.

Problem

Sentiment analysis is a form of Natural Language Processing (NLP) to determine the aspect of given data. These aspects could be positive, negative, or neutral. Sentiment analysis is usually performed on textual data and helps organizations to achieve a level of understanding of customers' opinions toward their company and products. This knowledge can be used to make better decisions to benefit the company.

Most of the similar works have been based on English language. However, there is not much investigation in Persian compared to English. For this project, we want to use Machine Learning techniques to perform sentiment analysis on Persian text.

Related Works

1. Using bag of words for feature selection, PCA for dimension reduction, and finally Stacked Bidirectional LSTM Neural Network for classification, they were able to reach 81% accuracy for a dataset from movie reviews. [1]
2. The proposed model in this article contains a combination of classifiers and deep neural network. Both parts use information from local disclosure and external knowledge base with 95.61% accuracy. The dataset for this article was from reviews of hotels. [2]
3. In this article LSTM, Word2Vec, and CharEmbed are used to classify comments from DigiKala online shop into three classes (positive, negative, and neutral). Their final F1-Score was 78.3. [3]

Our Approach

We want to use a dataset containing 63000 comments from DigiKala that we have labeled manually. Despite other methods that use two classes (positive and negative), we are using three (the third one is neutral). For our project, we want to use Logistic Regression, Softmax, Naïve Bayes, and SVM to classify our data and compare their results with each other and our manual labels. If the size of feature space gets out of hand, using dimensionality reduction methods is an option.

---

[1] Dashtipour, Kia, Mandar Gogate, Ahsan Adeel, Hadi Larijani, and Amir Hussain. "Sentiment analysis of Persian movie reviews using deep learning." *Entropy* 23, no. 5 (2021): 596.

[2] Dehkharghani, Rahim, and Hojjat Emami. "A novel approach to sentiment analysis in Persian using discourse and external semantic information." *arXiv preprint arXiv:2007.09495* (2020).

[3] Heydari, Mohammad, Mohsen Khazeni, and Mohammad Ali Soltanshahi. "Deep Learning-based Sentiment Analysis in Persian Language." In *2021 7th International Conference on Web Research (ICWR)*, pp. 287-291. IEEE, 2021.