Name: Morgan Gere

Course: IST 736- Text Mining

Date: 4/11/2022

Homework #1 Sentiment Analysis comparison

**Introduction**

At this point in our history, it is safe to assume most people have heard of or even know a little about artificial intelligence (AI). For anyone that doesn’t know or just so we can clarify exactly what we are discussing, Merriam-webster defines artificial intelligence in two ways 1) ”A branch of computer science dealing with the simulation of intelligent behavior in computers” and 2) “the capability of a machine to imitate intelligent human behavior” (Merriam-Webster). Based on my interactions with people and talking about AI I would expect that people are divided on how they feel about AI.

To test and see what the public sentiment about AI is, two sentiment analysis programs will be used to score tweets collected that pertain to AI. The two packages will be VADER and Flair. Both are pretrained sentiment analysis models. VADER uses predetermined lexicon of words to assign specific words a score (positive, negative, or neutral) and then compounds the scores to find an overall score (Shahul ES). Flair is a trained off IMDB dataset and gives a score and a value stating if the score is positive or negative (Shahul ES).

**Method**

Snscrape was used to collect tweets from twitter. 1001 tweets were collected, and the contents of the tweet were placed into a list. Snscrape collected tweets that had “Artificial Intelligence” in the tweet themselves. This is a random sampling of the latest available tweets pertaining to AI. This is a small sample size but should give us some insight into the sentiment analysis packages as well as an idea of the public sentiment of AI. A larger sample size would give us more information and would be the next step after determining which sentiment analysis package worked best.

Text

Description automatically generated

These tweets were manually inspected to see if any data preparation was required. Since Vader is assigning scores to specific words and Flair is giving scores to sentences based on prior training. It was decided no data preprocessing was required. Wanting to see how the packages work with the full content of the tweets. If it is possible to achieve good results with less coding, then it makes this process easier to accomplish in the future with a much larger amount of data.

Text

Description automatically generated

A way to compare these two sentiment analysis packages was required. Since they do not have the same output. Vader gives a negative, neutral, positive, and compound score, Flair gives a score and a value stating if the score is positive or negative. To be able to compare, the compound score was used from Vader, and a function was created to make the score ascertained from Flair positive or negative so it would be more understandable. These scores were graphed, and the extreme values were manually inspected to see accuracy. This will give a reasonable understanding of what the sentiment analysis packages are telling us.

Vadar was imported and a loop was run to assign each tweet a sentiment score. These scores were placed into a list so they can be graphed to have a visual inspection of the data. As stated previously only the compound score was collected.

Graphical user interface, text, application

Description automatically generated

A loop was created to create a list for naming the tweets in the graph. The loop assigned each tweet a name with the number it was in the list along with a T at the start.

Graphical user interface, text, application

Description automatically generated

Matplot was used to create a graph of the Sentiment Scores for visual inspection. Below is the code and a small image this will be shown and examined in the results section.

A picture containing graphical user interface

Description automatically generated

A loop was used to find the index of the highest and lowest scored tweets. These tweets will be manually inspected in the results section.

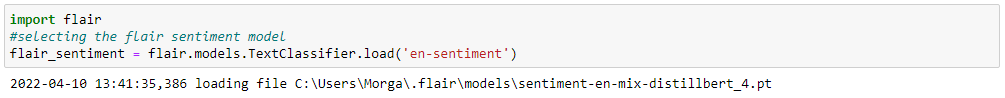
Text

Description automatically generated

Text

Description automatically generated with medium confidence

Flair was imported and the pretrained sentiment analysis model was selected.



As stated above a function is created so the score found by flair is either multiplied by 1 for a positive score value or -1 for a negative score value. This gives us an overall score that is comparable to the compound score of Vader sentiment analysis.

Graphical user interface, text, application, email

Description automatically generated

A loop was used to run each tweet through the function created and the scores were captured in a list.

Graphical user interface, text, application, email

Description automatically generated

Again, Matplot was used to create a graph to visually inspect Flair sentiment analysis. This will be seen in the results section, below is the code and a small image.

A picture containing text

Description automatically generated

A loop was used to find the index of the highest and lowest scored tweets. These tweets will be manually inspected in the results section.

A picture containing graphical user interface

Description automatically generated

Text

Description automatically generated with medium confidence

**Results**

Visually inspecting the graphs below, there is a large difference in the outcomes. Both graphs are attached separately for a much larger view allowing better visual inspection.

**Vader Sentiment Analysis**

Icon

Description automatically generated

**Flair Sentiment Analysis**

**Icon

Description automatically generated with medium confidence**

Even with just this small view it is very clear that the Vader results are almost exclusively positive and have more average values with some extremes. Looking at the Flair results it is much more mixed with about half positive and half negative. The Flair results are also much more extreme, the positive values are very positive, and the negative values are very negative with only a few that are slightly positive or negative. To see which is more accurate the 3 most extreme examples of negative and positive from each were examined to give an understanding of which package is giving the better results.

Selecting the highest positive values from Vader the tweets are below.

Graphical user interface, text, application, email

Description automatically generated

Tweet 1 is not and 2 are about movies. Tweet 1 seems neutral. It is asking questions stating if you are a fan of these things go see this movie. Tweet 2 is unsure of the possible moral dilemmas in fictional AI. It seems negative possibly neutral at best. While tweet 3 is positive they were excited to learn about India’s initiatives in the AI field.

Selecting the highest negative values from Vader the tweets are below.

Text, application

Description automatically generated

Tweets 1, 2, and 3 are negative statements about AI.

Selecting the highest positive values from Flair the tweets are below.

Graphical user interface, text, application, email

Description automatically generated

All three tweets are very positive and are directly about AI.

Selecting the highest negative values from Flair the tweets are below.

Graphical user interface, text, application, email

Description automatically generated

All three tweets are very negative about AI.

**Conclusion**

Overall, Vader did a good job identifying the negative tweets and poor job identifying positive tweets. Since these are the highest scored positive tweets, it would be expected to be less ambiguous in what they are saying. In looking at Flair the most negative and positive tweets are directly talking about AI and are truly positive and truly negative. Flair did the best job and fit better with expected public sentiment of AI. The results are people feel very positive or very negative about AI. The split is fairly even. My recommendation is to begin using flair to get a good image of public sentiment on AI.

For future steps more tweets would be needed to have a larger understanding of the public sentiment of AI based on tweets. The larger sample size over several weeks or months would give a better understanding of the actual public sentiment on AI. While using only tweets from a limited time frame, they are susceptible to outliers. These outliers could be a movie or book just released that was very positive/negative towards AI that people are tweeting about. More tweets would need to be inspected manually to understand how well the package is doing with its analysis. Once enough data had been collected a sentiment analysis specific for AI and tweets, could be trained. Using supervised learning to create a model to identify sentiment in AI based on tweets.

**Citations**

Merriam-Webster. (n.d.). *Artificial Intelligence Definition & meaning*. Merriam-Webster. Retrieved April 11, 2022, from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Shahul ES Data scientist with strong acquaintance in the fields of Machine Learning, ES, S., Data scientist with strong acquaintance in the fields of Machine Learning, & on, F. me. (2021, December 3). *Sentiment Analysis in python: Textblob vs vader sentiment vs Flair vs building it from scratch*. neptune.ai. Retrieved April 11, 2022, from https://neptune.ai/blog/sentiment-analysis-python-textblob-vs-vader-vs-flair