**Spring School: Data Science Methods within Geo-Informatics**

**Final Project: Evaluating the Performance of a Location Extractor Model using the f-beta Score.**

**Introduction**

These days, the availability of large data streams –structured and unstructured – gives us the opportunity to extract information for different use cases that didn’t exists until now. This information extraction process is useful in multiple domains and fosters innovations, support policy and decision making and serves as the starting point of many research interests. Within the geoinformatics domain, central to these information extraction techniques, is the extraction of location information in form of latitude, longitude, street names, states, county and other information entities that relates to location. An important source of location data these days is social media and search engine platforms. For instance, Google processes over 3.5 billion search queries everyday and at least 500 million tweets are made every day.

In most cases, through accessible social media APIs and text databases, thousands of text information can be access and analysed. At the heart of this information extraction process is Natural Language Processing, which serves multiple research interests. In this case, we utilized the Spacy library developed by Explosion to extract the location information from a set of social media posts.

**Methods**

After applying the Named Entity Recognition algorithm available in SpaCy, the newly generated dataset with location information was annotated using a binary class of 1 (contains location entity) and 0 (no location entity). This was based on the guideline provided by the Spring School annotation guideline. To fast track the analysis, the first 300 sets of the social media posts were annotated and coded. The annotated data was store in a JSON file (*ss\_california\_tweets\_annotatedRufai\_Vitoria\_FS*) accessible in the files folder.

In order to develop a machine learning model that could be used to predict the location of other text information, we designed a simple Decision tree classifier. This was done after the location entities fields have been vectorized in a binary form as well to match the label field, see the figure below.

Table

Description automatically generated

Table

Description automatically generated

The dataset was split into training (80%) and test sets (20%) and the model was trained on the annotated data, and a prediction was carried out on the test sets. The random state was set to 42, to minimize variation in model output at each run. The code cell below shows how these algorithm was called from the SK-Learn API and applied to our use case.

Graphical user interface, text, application, email

Description automatically generated

Afterwards, the evaluation metrics (f1 scores and f-beta score) were computed. We wanted to understand our well the model performed in identifying true negatives and true positives. To do this, we need an evaluation metric that can suitably identify the different weightings of recall and precision. In this case, f-beta score (also available in SK-learn API) is the most robust metric to use. The harmonic mean of recall and precision -- f1-Score – provides a weighted balance between these precision and recall.

To best evaluate the model performance, the extracted locations were coded as 1 and 0, in the spacy extraction of locations entities [FAC, GPE, ORG and LOC] to fit with the annotated sets. The spacy library has four location entities [FAC, GPE, ORG and LOC] including any mentions of any geographical objects such as countries, cities, schools, organizations, buildings, bridges, water bodies, mountains, monuments etc. See the computed value below:

Graphical user interface, text, application, email

Description automatically generated

|  |  |  |
| --- | --- | --- |
| f-beta-Recall | f-beta-Precision | f1 Score |
| 0.912709 | 0.89011 | 0.89011 |

The table shows a summary of the output of evaluation metric with the computation of the differing focus on recall, precision and just the weighted mean (f1 score).

What does the value imply?

Q: It is important to note that, in this analysis, the sets of social media posts used were not focused on a particular topic but encompasses a range of topic interests. By large, the cost of deciding which evaluation is considered robust depends on the application context. Given an application context of locating endangered persons during an humanitarian crises, it will be better to have more precision (more TPs amongst the real TPs) than recall. As such, we can easily tweak the f-beta score to evaluate for.

**In cases, where a correct location is extracted together with noise (text that are not part of the location), we can better deal with it by …**

**Model Evaluation**

**Discussion and Considerations**

**Conclusion**

**References**

What is F-Score? Thomas Wood (n,d) retrieved from https://deepai.org/machine-learning-glossary-and-terms/f-score