Lab3

Inference of User Demographics from Multiple Social Networks

Due date: 9th May

Objectives:

Modeling user interests will benefit many services and applications, such as adaptive E-learning, target advertisement and personalized services. On the other hand, user' online behaviors on the prominent social platforms, such as Facebook, Twitter and LinkedIn, are mostly interest-driven, which essentially reflect users' personal preferences and interests. Moreover, users are connected and organized by multiple social platforms, which characterize users from different angles. For example, Facebook, Twitter and LinkedIn emphasize social connection, information exchange and professional archive, respectively. Therefore, aggregating and exploring a user's footprint from multiple social platforms is a promising way to harvest a more comprehensive view of users.

This assignment aims to design an effective multi-label multi-source classifier and then infer demographics of users based on a given dataset collected from multiple social networks. There are several challenges, such aggregating suitable informative features from multiple sources, and designing effective classifier.

What You Need to Do

You need to implement a multi-label multi-source classifier based on the user-generated-content –features extracted from multiple social networks for Demographics inference. It should incorporate the following functions:

1. **Basic Requirements:**
    *   You need to implement any suitable effective machine learning techniques for your multi-label multi-source classifier.
    *   You need to implement the required evaluation metrics (F1 Measure and Accuracy) and tabulate the overall results to demonstrate the effectiveness of each source and the resulting classifier.
    *   It is noted that your classifier should be able to classify a user's interests into given Groups (as it is set up in the Ground Truth File) and perform results testing based on the test set (provided by Qi Yang).
    *   You can appropriately combine features from different sources via early fusion.
    *   You can implement advanced multi-source multi-label models, such as multi-label SVM and even multi-source classification.
    *   One key aspect of testing is to systematically demonstrate the effectiveness of different combinations of multiple sources in the classifier compared to single sources.

What You are Given:

1. **Dataset**
   - Multi-Source Demographics Profiling Dataset
2. **Baseline Approaches described in lecture notes**

Report

You need to submit before the deadline:
   - Source codes of your implementation.
   - Jupyter notebook for representation or in PowerPoint format.

Remarks:

(a) All members are required to present some aspects of the system.

(b) Extra marks will be given for excellent assignments.

Consultation:

For questions regarding this assignment, please consult:
   - Qi Yang

**Late Submission Policy

We impose the following penalties for late submissions. (a) Late but within 24 hours: 25% reduction in grades. (b) Late but within 3 days: 50% reduction in grades. (c) After 3 days: zero marks.