Course Project Overview

Instructor: Yizhou Sun
yzsun@cs.ucla.edu

October 8, 2017
General Goal

- Apply data mining algorithms to real-world problems
  - Choose topic
  - Collect data
  - Apply algorithms to the data
  - Evaluate and compare algorithms
  - Submit a report, together with data and code
Detailed Stages: 1. Form Groups

- Sign-up team: 4-5 members per team
  - Group ID, name, members, topics
- Point: 1
Detailed Stages: 2. Midterm Report

• Submit a 5-page report, indicating
  • Which problem you want to solve
  • How to break the problem into subtasks and formalize them into data mining problems
  • What’s your strategy in crawling Twitter data and describe what you plan to get
  • Schedule of your remaining work
  • Discussion of problems you have met
  • References

• Points: 5
Detailed Stages: 3. Final Report

• Submit a 10-page final report
  • Enrich the major part of midterm report
  • Demo system (if any) or final results
  • Workload distribution
• Submit code and data
• Points: 19
Collaborating Rule

- Every member in a team gets the same score (encourage teamwork)
  - Exception: the team has the right to claim someone as a free rider, and we will lower his/her score

- Final report should include a table describing each member’s duty
- We also collect Peer evaluation form
# Sample of Workload Distribution Table

<table>
<thead>
<tr>
<th>Task</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collecting and preprocessing data</td>
<td>Student A</td>
</tr>
<tr>
<td>2. Implementing Algorithm 1</td>
<td>Student B</td>
</tr>
<tr>
<td>3. Implementing Algorithm 2</td>
<td>Student C and D</td>
</tr>
<tr>
<td>4. Evaluating and comparing algorithms</td>
<td>Student A</td>
</tr>
<tr>
<td>5. Writing report</td>
<td>Student B and C</td>
</tr>
<tr>
<td>6. Slides, demo, and Presentation</td>
<td>Student A, B</td>
</tr>
</tbody>
</table>
Twitter Projects

• Three topics to choose
  • Stock price prediction
  • Mood detection and prediction
  • Trending Event detection
Stock Price Prediction

• Goal
  • Predict stock price for several certain stocks or overall index

• Possible subtasks
  • Decide prediction tasks: short term or long term?
  • Focused crawling: collect tweets that are related to a company or an industry
  • What data mining problem it can be formalized into?
  • Which data mining algorithms can be applied to solve this problem?
  • How to evaluate the performance of different algorithms?
References

• Johan Bollen et al., Twitter mood predicts the stock market, Arxiv, 2010
• Anshul Mittal et al., Stock Prediction Using Twitter Sentiment Analysis
  • http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.375.4517&rep=rep1&type=pdf
Mood detection and prediction

• Goal
  • Detect and predict happiness index for twitter users according to their tweets

• Possible subtasks
  • Decide which mood classification scheme to use
  • Decide the scope of tweets to crawl
  • What features will affect people’s mood, e.g., # of friends, # of tweets?
  • What data mining problem it can be formalized into?
  • Which data mining algorithms can be applied to solve this problem?
  • How to evaluate the performance of different algorithms?
References

• Kirk Roberts et al., EmpaTweet: Annotating and Detecting Emotions on Twitter.
  • http://www.hlt.utdallas.edu/~kirk/publications/robertsLREC2012_2.pdf
  • https://mislove.org/twittermood/

• Johan Bollen et al., Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena, ICWSM’11
Trending Event Detection in LA

• Goal
  • Detect and rank the trending events in a specified location, e.g., LA

• Possible subtasks
  • How to model an event?
  • How to crawl tweets within a specified location?
  • How to detect and track an event?
  • How to summarize an event?
  • How to categorize them into different event types?
  • How to evaluate the performance of different algorithms?
References

• Rui Li et al., TEDAS: A Twitter-based Event Detection and Analysis System, ICDE’12
• Charu C. Aggarwal et al., Event Detection in Social Streams, SDM’12