Documentation

Team: Team-RPDD

Team member:Yuran Wang (netid:yuranw3) [captain]
Zhengyu Li (netid:zhengyu6)
Hongru Wang (netid:hongru2)

Project: Reproducing A Paper(Latent aspect rating analysis)

1) An overview of the functions of the code:
    a) Create vocabulary from the dataset.
    b) Use stemmer on the dataset to create a vocabulary corpus of the reviews, which will be used for aspect mining and rating.
    c) Use BootStrapping to mine aspects(stored in\Data\Seeds\hotel_bootstrapping.dat), use regression to calculate weight for each aspect.
    d) Calculate rating per minded aspects.
    e) The results are in Data/Vectors/vector_CHI_4000.dat"
2) Implementation documentation:
    a) Acknowledgment: The authors of the paper released the codes for LARA on his personal website(`http://sifaka.cs.uiuc.edu/~wang296/`) . We modified the original version of the code for our projects.
    b) Review dataset sources: http://sifaka.cs.uiuc.edu/~wang296/Data/index.html
    c) Setup instructions:
        i) Add colt.jar, concurrent.jar, opennlp-maxent-3.0.1-incubating.jar, opennlp-tools-1.5.1-incubating.jar, and JRE System Library[Java-SE-1.7] to the libraries of Java Build Path. Then run the “Analyzer” file.
3) Usage documentations:
    a) In order to run the code, the user needs to import tools from opennlp. You can import the core toolkit directly from Maven. Then run mvn install. The detailed dolumentation can be found here https://github.com/apache/opennlp.
    b) It requires tools from NTLK
    c) In order to run the code, the user also needs Python3 and Java.
4) Work distribution:
    a) We worked together so everyone on the team participated.

5) Final report: The authors of the paper developed the LARAM to effectively solve the problem of LARA, including automatically identifying meaningful topical aspects, inferring interesting differences in aspect ratings within reviews, and modeling users' preferences with the inferred relative emphasis on different aspects. Such detailed analysis of opinions at the level of topical aspects enabled by LARAM can support multiple application tasks, including aspect opinion summarization, ranking of entities based on aspect ratings, and analysis of reviewers rating behavior. Our project borrowed the ideas from the author's work about LARA, and our work is mainly based on aspects-mining.