

README

Name : Tushar Kumar

NetId : tusharku

Collaborators : None

October 29, 2018

1 Introduction

This project, which was implemented by myself, and **did not have any collaboration from anybody else**, implements Apriori and FPGrowth algorithm. I also implement an improvement on top of Apriori. For extra credit I generate association rules and run prediction on test data for generating association rules. I also have implemented Naive Bayes algorithm and test it on test data. This project was undertaken as part of the graduate course(CSC 440) at University of Rochester.

2 Technology Used

- Java as the programming language (Java 10.0.2) with no external libraries
- Eclipse used as IDE
- [Overleaf](#) for generating reports in L^AT_EX

3 Building the project

- Download the **CSC440MinorProject-tusharku.zip** file(which you would have if you are reading this file.
- Unzip the file to get the CSC440MinorProject-tusharku folder
- Run the following commands in sequence once you are in the location where you downloaded the zip file: Please mind the line break created because the command being of greater length than width of the page. Actual command would be
"javac -d executable -sourcepath src
-cp . src/com/uofr/csc440/project/project1/FrequentItemSetPatternMiner.java"

```
cd CSC440MinorProject-tusharku
```

```
javac -d executable -sourcepath src -cp .
```

```
↪ src/com/uofr/csc440/project/project1/FrequentItemSetPatternMiner.java
```

In case , you were not able to compile this, not to worry, I have provided the already compiled binaries of classes in the bin folder. So you can use that straight away to run the application

4 Running the application

- Run the following command to execute the program. This will just run Apriori. You would need to add arguments to run and test different algorithms

```
cd CSC440MinorProject-tusharku
```

```
java -cp executable
```

```
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
```

This, by default would mine the dataset with Apriori algorithm with a support of 0.23. These are the commandline arguments with their description and possible values that can be provided in order to run the program.

- **strategy** : This represents strategy that you want to use to mine the adult dataset.
Possible Values : apriori(for Apriori), apriori-improved(for running Apriori-Improved), fpgrowth for FPGrowth
Default Value : apriori
- **support** : This represents the minimum support you want to mine the dataset with.
Possible Values : Double
Default Value : 0.23
- **confidence** : This represents the minimum confidence you want to use while generating association rules..
Possible Values : Double
Default Value : 0.0 (Will not generate association rules)
- **runnaivebayes** : Whether you want to run Naive Bayes Algorithm or not.
Possible Values : true or false
Default Value : false
- **usetest** : If you want to run generates itemsets of test data.
Possible Values : true or false
Default Value : false

4.1 Commands to test fulfillment of project requirements

Here are some sample scenarios with their commands provided:

- **Running Apriori with support of 0.23**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
```

- **Running Apriori with a different support say 0.5**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↪ -support 0.5
```

- **Running Apriori-Improved with default support of 0.23**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↪ -strategy apriori-improved
```

- **Running Apriori-Improved with a different support say 0.5**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↪ -strategy apriori-improved -support 0.5
```

- **Running FPGrowth with default support of 0.23**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↪ -strategy fpgrowth
```

- **Running FPGrowth with a different support of say 0.5**

```
java -cp executable
↪ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↪ -strategy fpgrowth -support 0.5
```

- **Extra Credit : Generating Association Rules and Running Prediction**

Do note that this will also run the pattern mining algorithm as that will always run.

```
java -cp executable
↳ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↳ -strategy fpgrowth -support 0.02 -confidence 0.99
```

- **Extra Credit : Running Naive Bayes Algorithm**

Do note that this will also run the pattern mining algorithm as that will always run.

```
java -cp executable
↳ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↳ -runnaivebayes true
```

- **Generate item sets only of test data**

By Default I would ONLY run the itemset generation on adult.data. This is incase one wants to test the itemset generation of adult.test

```
java -cp executable
↳ com.uofr.csc440.project.project1.FrequentItemSetPatternMiner
↳ -usetest true
```