

**NANYANG  
TECHNOLOGICAL  
UNIVERSITY**  

---

**SINGAPORE**

Report  
On  
EE4483/IM4483  
Artificial Intelligence and Data Mining  
Continuous Assessment  
BY FP growth

Xiong ChenYu  
U1521516C  
EEE

Oct. 20, 2017

## Contents

1	Abstract	2
2	Question 1 (How many frequent itemsets have the minimum support of 20%, 10%, 5%, and 3% respectively?)	3
3	Question 2 (What are the respective percentages of frequent 3-itemsets, and 2-itemsets, with respect to all possible itemsets, which have a minimum support of 3%?)	4
4	Question 3 (How many association rules have a minimum confidence of 50% and a minimum support of 5% and 10%, respectively? Briefly explain how the minimum support affects the strong rules generated. )	5
5	Question 4 (List three association rules that have the highest support with 100% confidence?)	5
6	Question 5 (Do you find any “interesting” rules? What are they? Briefly explain why.)	5
	References	5
7	APPENDIX A	6

## List of Tables

1	System Specification . . . . .	2
2	Frequent itemsets counts . . . . .	3
3	3% minsupport itemsets distribution . . . . .	4
4	respective percentages . . . . .	4
5	Rules counts on 50% minimum confidence . . . . .	5
6	Rules(X itemsets => Y itemsets) . . . . .	5

## List of Figures

# 1 Abstract

This report answer questions of the research and study on the continues assessment.  
The algorithms used to do this project is the FP growth.

Table 1: System Specification

Usage	Application
System	Mac Sierra
Algorithm	FP growth
Language	Scala
Package	Spark

**2 Question 1 (How many frequent itemsets have the minimum support of 20%, 10%, 5%, and 3% respectively?)**

Table 2: Frequent itemsets counts

Minimum support (Unit : %)	Itemsets size counts
20	20
10	68
5	268
3	659

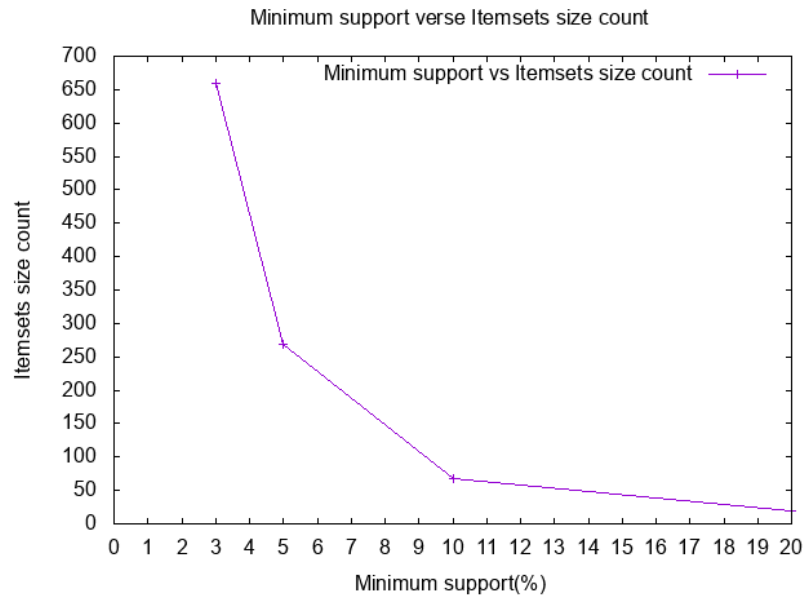


Table 3: 3% minsupport itemsets distribution	
Itemsets size	Itemsets size count (Total: 659)
1	20
2	190
3	424
4	25
5	0

**3 Question 2 (What are the respective percentages of frequent 3-itemsets, and 2-itemsets, with respect to all possible itemsets, which have a minimum support of 3%?)**

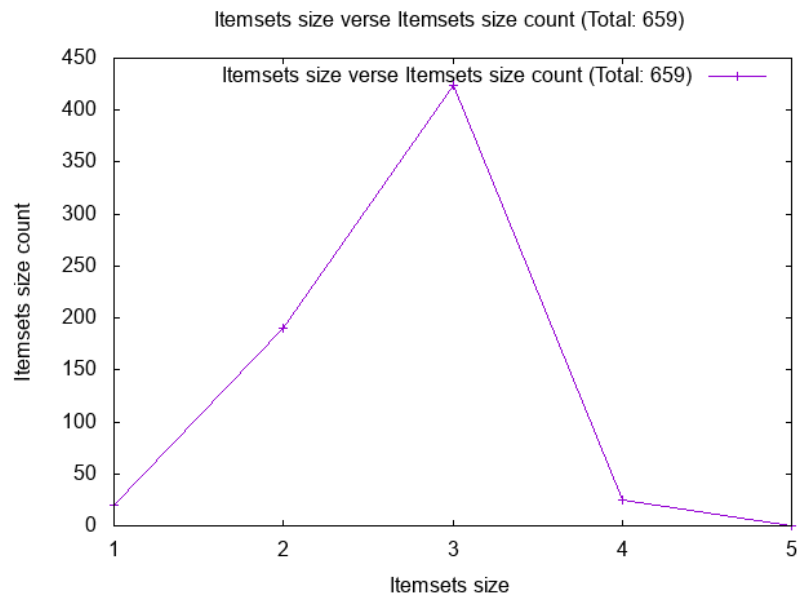


Table 4: respective percentages	
Itemsets size	Itemsets size count (Total: 659)
1	$\frac{20}{659} = 0.03$
2	$\frac{190}{659} = 0.288$
3	$\frac{424}{659} = 0.643$
4	$\frac{25}{659} = 0.038$
5	$\frac{0}{659} = 0$

- 4 Question 3 (How many association rules have a minimum confidence of 50% and a minimum support of 5% and 10%, respectively? Briefly explain how the minimum support affects the strong rules generated.)**

Table 5: Rules counts on 50% minimum confidence  
Minimum Support(Unit: %) Rules counts

10	0
5	117

- 5 Question 4 (List three association rules that have the highest support with 100% confidence?)**

Table 6: Rules(X itemsets => Y itemsets)

X itemsets	Y itemsets	support
Salad,Ham,Banana	Apple	0.33
IceCream,Olive,Tea	Banana	0.33
Ham,Coffee,Diaper	IceCream	0.27

- 6 Question 5 (Do you find any “interesting” rules? What are they? Briefly explain why.)**

## References

- [1] EN.WIKIPEDIA.ORG (2017). Data Mining Algorithms In R/Frequent Pattern Mining/The FP-Growth Algorithm - Wikibooks, open books for an open world. [online] Available at: [https://en.wikibooks.org/wiki/Data\\_Mining\\_Algorithms\\_In\\_R/Frequentpattern\\_Mining/The\\_FP-Growth\\_Algorithm](https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Frequentpattern_Mining/The_FP-Growth_Algorithm)[Accessed7Nov.2017].
- [2] EN.WIKIPEDIA.ORG (2017). Data Mining Algorithms In R/Frequent Pattern Mining/The FP-Growth Algorithm - Wikibooks, open books for an open world. [online] Available at: [https://en.wikibooks.org/wiki/Data\\_Mining\\_Algorithms\\_In\\_R/Frequentpattern\\_Mining/The\\_FP-Growth\\_Algorithm](https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Frequentpattern_Mining/The_FP-Growth_Algorithm)[Accessed7Nov.2017].

## 7 APPENDIX A

```
package example
import org.apache.spark.{SparkContext, SparkConf}
import org.apache.spark.rdd.RDD
import org.apache.spark.mllib.fpm.FPGrowth

object Hello extends Greeting with App {

  println(greeting)
  val conf = new SparkConf().setAppName("Ex1_SimpleRDD").setMaster("local[4]")
  val sc = new SparkContext(conf)

  sc.setLogLevel("ERROR")

  val data = sc.textFile("data.csv")
  val transactions: RDD[Array[String]] = data.map(s => s.trim.split(", "))

  val fpg = new FPGrowth().setMinSupport(0.0276).setNumPartitions(10)
  val model = fpg.run(transactions)

  model.freqItemsets.collect().foreach { itemset =>
    println(itemset.items.mkString("[", ",", "]")
      + ", "
      + itemset.freq) }

  val minConfidence = 1
  model.generateAssociationRules(minConfidence).collect().foreach { rule =>
    println(rule.antecedent.mkString("[", ",", "]")
      + " => "
      + rule.consequent.mkString("[", ",", "]")
      + ", "
      + rule.confidence) }
  println(model.generateAssociationRules(minConfidence).collect().size)
  sc.stop()
}

trait Greeting {
  lazy val greeting: String = "hello"
}
```