

Associate Rule

Bochen Wang, Tianyuan Liu

November 29, 2015

Introduction

This is the second part of Homework 4. Implement the example from Harrington Sec. 11.5.

Implementation

Read Data

Read data from votesmart_20bill.txt and votesmart_20bill_meaning.txt.

```
# Read Data
data = []
temp = []
meaning = []

with open('votesmart_20bill.txt') as file1:
    last_tag = -1
    for line in file1:
        line = line.strip()
        index = line.find('I')
        if index != -1:
            tag = int(line[index+1:])
            if tag <= last_tag:
                if len(temp) > 0:
                    data.append(temp)
                temp = []
                temp.append(tag)
            else:
                temp.append(tag)
            last_tag = tag

with open('votesmart_20bill_meaning.txt') as file2:
    for line in file2:
        line = line.strip()
        index = line.find('S')
        if index != -1:
            meaning.append(line[index+2:-1])
```

Apriori and Associate Rule

The code of Apriori algorithm and associate rule is cited from Harrington Sec. 11.5.

createC1, scanD, aprioriGen, apriori, generateRules, calcConf, rulesFromConseq, these functions are cited from Harrington's book.

Frequent Item Sets

The code is as below:

```
def getFrequentItemset():
    minSupport = 0.5
    while minSupport >= 0.3:
        L, support = apriori(data, minSupport)
        ret = str(len(L))
        for s in L:
            ret += ' ' + str(len(s))
        print ret
        minSupport -= 0.05
```

From the code above, we can get the result.

Minimum Support	Itemset Maximum size	Itemset size
0.5	3	5 1 0
0.45	3	10 4 0
0.4	5	14 33 18 3 0
0.35	7	16 51 84 63 27 5 0
0.3	9	21 75 121 114 78 35 9 1 0

Association Rule Mining

When the minumum confidence is set to low, there are so many association rules that we cannot figure out the pattern. The association rules we discover as below:

If	Then	Confidence
Prohibiting Federal Funding of National Public Radio – Yea	Repealing the Health Care Bill – Yea	100.0%
Prohibiting Federal Funding of National Public Radio – Yea	Republican	99.6%
Mine Safety Act – Yea	Democratic	98.1%
Unemployment Benefits Extension – Yea	”Whistleblower Protection” for Offshore Oil Workers – Yea	96.3%
Repealing ”Don’t Ask, Don’t Tell” After Military Review and Certification – Yea	Democratic	96.2%
Prohibiting Use of Federal Funds For Planned Parenthood – Yea	Republican	95.8%
”Whistleblower Protection” for Offshore Oil Workers – Yea	Prohibiting 2010- 2011 Congressional Cost-of-Living Pay Increase – Yea	93.3%