实验一

1. 实验内容：使用C++语言实现三种频繁项集挖掘算法：(1)Apriori[AS94b]; (2)FP-growth[HPY00]; (3)ECLAT[Zak00]（使用垂直数据格式挖掘）。在各种不同的数据集上比较每种算法的性能。写一个报告，分析在哪些情况下（如最小支持度阈值），一种算法比其他算法好，并陈述理由。
2. 算法实现源码：见附件。
3. 实现平台：VS2015专业版，win10专业版64位系统，Core i5，CPU2.5GHz。
4. 算法性能分析

受编程能力的限制，无法对三个算法的实现采用相对一致的水平，也就是说，对于易于实现的算法如Eclat能够以十分高效的方式进行实现，对于比较难的算法则实现起来就不那么优雅了，很是凌乱。因此，各个算法之间的性能差异很大，以至于很难进行三者的比较，但还是强行比较了一下（图1），由于Apriori算法使用了大量的STL容器以及很多递归以及其自身算法的缺陷，使得其时间性能最坏。虽然通过图1很难直观的看到三个算法的具体区别，但还是能够看出在时间性能上Apriori算法明显比Eclat算法和FP-growth算法要差很多，可能是具体实现的原因，另外两种算法都比Apriori算法快不止一个数量级。但这三种算法都有一个共同点：当最小支持度变小时，三种算法的需要的时间开始增长缓慢，但超过某一点便急剧增长。为了更好的理解每个算法的性能，我们分别对这三种算法进行了分析。

图1 Apriori、Eclat和FP-growth算法性能比较

（1）Apriori算法：通过运行，算法需要时间如表1。

表1 Apriori算法

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| min\_sup | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 |
| time(ms) | 12644308 | 2547410 | 793640 | 267516 | 120234 | 57527 | 38143 | 7431 |

通过表1，我们绘制了该算法时间性能的散点图（图2），从图中我们可以看出，当最小支持度变小时，三种算法的需要的时间开始增长缓慢，但超过某一点便急剧增长。

图2 Apriori算法时间性能分析

（2）Eclat算法：通过运行，算法需要时间如表2。

表2 Eclat算法

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| min\_sup | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 |
| time(ms) | 2879 | 1120 | 635 | 419 | 328 | 261 | 230 | 179 |

通过表2，我们绘制了该算法时间性能的散点图（图3），从图中我们可以看出，当最小支持度变小时，三种算法的需要的时间开始增长缓慢，便增长较快，但没有Apriori那么明显。

图3 Eclat算法时间性能分析

（3）FP-growth算法：通过运行FP-growth算法，需要时间如表3。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| min\_sup | 300 | 400 | 500 | 600 | 700 | 800 | 900 |
| time(ms) | 11720 | 4314 | 2565 | 1402 | 788 | 689 | 361 |

通过表3，我们绘制了该算法时间性能的散点图（图4），从图中我们可以看出，当最小支持度变小时，三种算法的需要的时间开始增长缓慢，但超过某一点便增长较快。但没有Apriori那么明显。

图4 FP-growth算法时间性能分析

总得来说，三种算法从时间性能上来讲Eclat > FP-growth > Apriori，即Eclat算法最优，FP-growth次之，Apriori算法最差。

1. 实验结果：通过对老师给的数据集进行挖掘，我们得出最小支持度阈值从200-900的挖掘结果（结果保存在map<vector<int>, vector<int>>中，为了方便只输出每个最小支持度阈值对应的最长的频繁项集）。如下：

for min\_sup=200 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1 2 3

support count: 315

1 2 4

support count: 279

1 2 5

support count: 287

1 2 6

support count: 265

1 2 7

support count: 268

1 2 8

support count: 263

1 2 9

support count: 263

1 2 10

support count: 237

1 2 11

support count: 211

1 2 12

support count: 217

1 2 13

support count: 208

1 2 14

support count: 204

1 3 4

support count: 276

1 3 5

support count: 291

1 3 6

support count: 286

1 3 7

support count: 276

1 3 8

support count: 266

1 3 9

support count: 258

1 3 10

support count: 236

1 3 11

support count: 232

1 3 12

support count: 237

1 3 14

support count: 207

1 4 5

support count: 256

1 4 6

support count: 251

1 4 7

support count: 259

1 4 8

support count: 234

1 4 9

support count: 222

1 4 10

support count: 221

1 4 11

support count: 205

1 5 6

support count: 251

1 5 7

support count: 252

1 5 8

support count: 256

1 5 9

support count: 232

1 5 10

support count: 234

1 5 11

support count: 201

1 5 12

support count: 214

1 6 7

support count: 257

1 6 8

support count: 255

1 6 9

support count: 237

1 6 10

support count: 217

1 6 11

support count: 215

1 6 12

support count: 211

1 6 13

support count: 201

1 6 14

support count: 203

1 7 8

support count: 231

1 7 9

support count: 232

1 7 10

support count: 214

1 7 11

support count: 218

1 7 12

support count: 211

1 8 9

support count: 230

1 8 12

support count: 203

1 9 10

support count: 200

2 3 4

support count: 291

2 3 5

support count: 297

2 3 6

support count: 271

2 3 7

support count: 279

2 3 8

support count: 285

2 3 9

support count: 269

2 3 10

support count: 257

2 3 11

support count: 239

2 3 12

support count: 216

2 3 13

support count: 205

2 3 14

support count: 212

2 4 5

support count: 254

2 4 6

support count: 245

2 4 7

support count: 262

2 4 8

support count: 248

2 4 9

support count: 233

2 4 10

support count: 226

2 4 11

support count: 212

2 4 13

support count: 200

2 5 6

support count: 228

2 5 7

support count: 249

2 5 8

support count: 259

2 5 9

support count: 232

2 5 10

support count: 233

2 5 11

support count: 202

2 5 13

support count: 200

2 6 7

support count: 239

2 6 8

support count: 241

2 6 9

support count: 229

2 6 10

support count: 207

2 6 11

support count: 202

2 7 8

support count: 237

2 7 9

support count: 227

2 7 10

support count: 217

2 7 11

support count: 202

2 8 9

support count: 241

2 8 10

support count: 233

2 8 13

support count: 204

2 9 10

support count: 215

3 4 5

support count: 254

3 4 6

support count: 242

3 4 7

support count: 247

3 4 8

support count: 235

3 4 9

support count: 204

3 4 10

support count: 218

3 4 11

support count: 209

3 5 6

support count: 244

3 5 7

support count: 253

3 5 8

support count: 251

3 5 9

support count: 228

3 5 10

support count: 239

3 5 11

support count: 215

3 5 12

support count: 205

3 6 7

support count: 249

3 6 8

support count: 241

3 6 9

support count: 213

3 6 10

support count: 202

3 6 11

support count: 214

3 6 12

support count: 209

3 7 8

support count: 229

3 7 9

support count: 214

3 7 10

support count: 210

3 7 11

support count: 214

3 7 12

support count: 205

3 8 9

support count: 220

3 8 10

support count: 210

3 8 11

support count: 208

4 5 6

support count: 220

4 5 7

support count: 220

4 5 8

support count: 226

4 5 10

support count: 204

4 6 7

support count: 228

4 6 8

support count: 219

4 7 8

support count: 209

4 7 9

support count: 202

5 6 7

support count: 219

5 6 8

support count: 227

5 7 8

support count: 218

5 8 9

support count: 207

5 8 10

support count: 202

6 7 8

support count: 216

6 7 9

support count: 212

6 8 9

support count: 201

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=300 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1 2 3

support count: 315

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=400 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1 2

support count: 521

1 3

support count: 515

1 4

support count: 466

1 5

support count: 487

1 6

support count: 480

1 7

support count: 453

1 8

support count: 444

1 9

support count: 433

1 10

support count: 402

2 3

support count: 536

2 4

support count: 485

2 5

support count: 483

2 6

support count: 448

2 7

support count: 454

2 8

support count: 464

2 9

support count: 449

2 10

support count: 412

3 4

support count: 457

3 5

support count: 485

3 6

support count: 452

3 7

support count: 447

3 8

support count: 447

3 9

support count: 417

3 10

support count: 400

4 5

support count: 419

4 6

support count: 422

4 7

support count: 423

4 8

support count: 403

5 6

support count: 410

5 7

support count: 409

5 8

support count: 430

6 7

support count: 411

6 8

support count: 412

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=500 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1 2

support count: 521

1 3

support count: 515

2 3

support count: 536

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=600 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1

support count: 936

2

support count: 952

3

support count: 935

4

support count: 849

5

support count: 873

6

support count: 841

7

support count: 800

8

support count: 812

9

support count: 768

10

support count: 724

11

support count: 691

12

support count: 672

13

support count: 642

14

support count: 620

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=700 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1

support count: 936

2

support count: 952

3

support count: 935

4

support count: 849

5

support count: 873

6

support count: 841

7

support count: 800

8

support count: 812

9

support count: 768

10

support count: 724

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=800 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1

support count: 936

2

support count: 952

3

support count: 935

4

support count: 849

5

support count: 873

6

support count: 841

7

support count: 800

8

support count: 812

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

for min\_sup=900 L:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1

support count: 936

2

support count: 952

3

support count: 935

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*