

"بسمه تعالی"

گزارش پروژه پیاده سازی الگوریتم های FP_Growth و Apriori

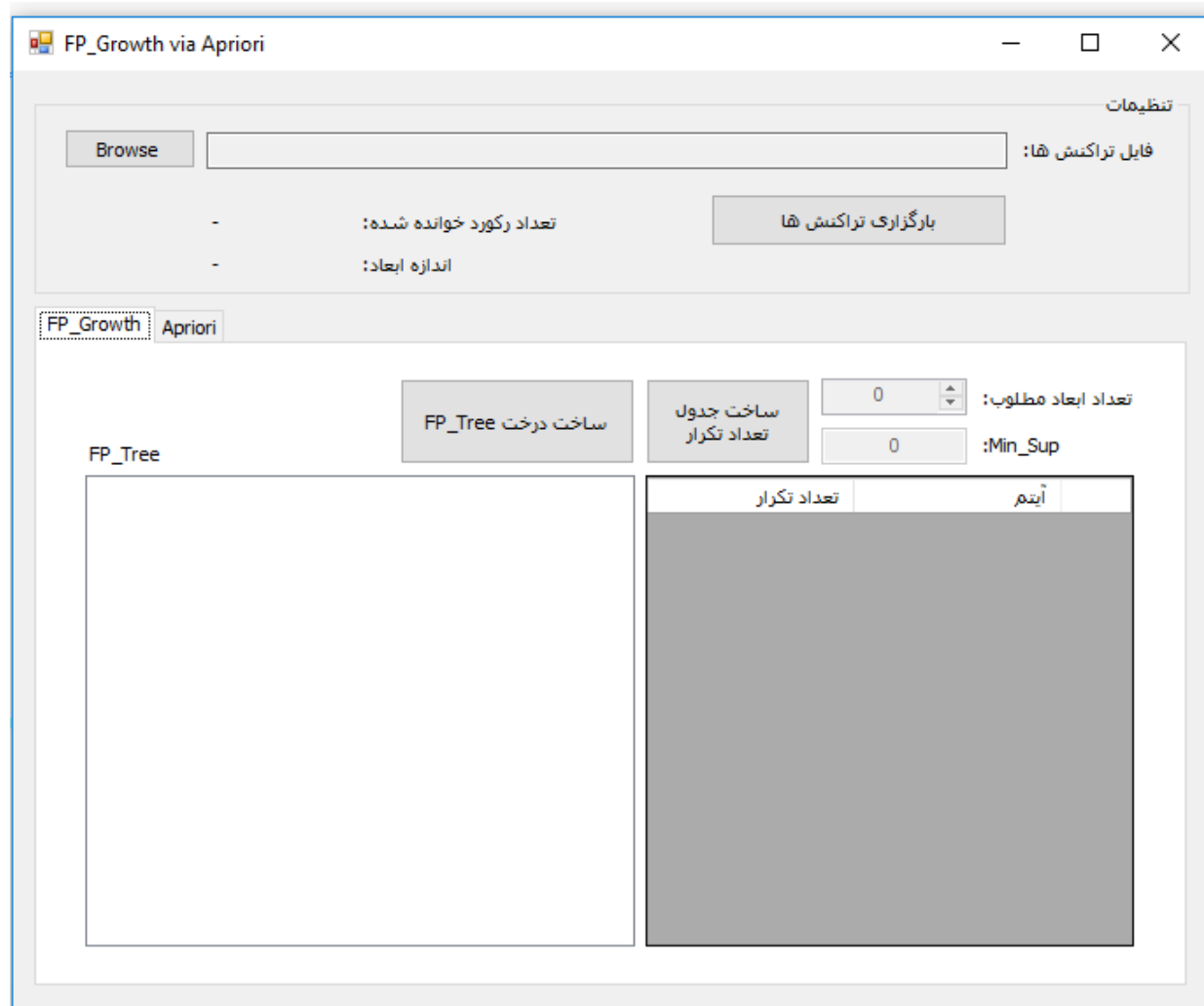
درس : داده کاوی

استاد : جناب آقای دکتر احمدی

تهیه کننده : فرخنده زینالی آق قلعه

شماره دانشجویی : ۹۶۱۱۲۷۴

در این پروژه قصد داریم تا قسمتی از الگوریتم FP_Growth یعنی ساخت درخت FP_Tree و الگوریتم Apriori را پیاده سازی کنیم. برای انجام این کار پروژه ای از نوع Windows Application با زبان C# در محیط Visual Studio 2010 ایجاد کردیم که نمای صفحه اصلی آن را در زیر مشاهده می نمایید. همانطور که ملاحظه می کنید قسمت مربوط به انتخاب فایل داده ها برای هر دو الگوریتم یکی می باشد امام جهت هر یک از الگوریتم های FP_Growth و Apriori یک tab page جداگانه به همراه تنظیمات در نظر گرفته شده است.



پس از خوانده شدن فایل تراکنش ها می توان یکی از عملیات FP_Growth یا Apriori را از قسمت tab page پایین صفحه انتخاب و به شرح مراحل زیر اجرا نمود:

FP_Growth

برای اجرای عملیات FP_Growth ابتدا می بایست جدول تعداد تکرار آیتم ها را ایجاد نمود لذا تعداد ابعاد مطلوب یا همان تعداد آیتم های در نظر گرفته شده و همچنین مقدار Min_Sup را وارد نموده و بر روی دکمه ساخت جدول تعداد تکرار کلیک نمایید. همانطور که دو تصویر زیر را مشاهده می کنید که یکی تعداد آیتم ها را ۲۰ و Min_Sup ۰ انتخاب شده است و دیگری همانطور که صورت مسئله خواسته است تعداد آیتم های ۱۰ و مقدار Min_Sup ۴۲۰۰۰ را انتخاب نموده ایم.

آیتم	تعداد تکرار
57398	0
54216	7
53514	9
49706	3
49402	1
47530	2
47229	8
43592	4

آیتم	تعداد تکرار
62020	15
58042	10
57712	17
57398	0
55984	13
55227	18
54563	16
54216	7
53514	9
49706	3
49689	11
49402	1
47530	2
47229	8
46839	12
45988	14
43592	4
43120	19
40229	5
36228	6

کار را با تنظیمات تصویر دوم یعنی تعداد ابعاد ۱۰ و Min_Sup ۴۲۰۰۰ ادامه می دهیم. برای ساخت درخت FP_Tree بر روی دکمه ساخت درخت FP_Tree کلیک می کنیم. پس از اتمام عملیات درخت مذکور به شکل زیر ساخته و نمایش داده می شود:

تنظیمات

فایل تراکنش ها: E:\Projects\Farkhondeh Projects\DataMining\Dataset.txt

تعداد رکورد خوانده شده: 100000

بارگزاری تراکنش ها

اندازه ابعاد: 20

تعداد ابعاد مطلوب: 10

ساخت جدول

تعداد تکرار

Min_Sup: 42000

ساخت درخت FP_Tree

FP_Tree

0:57398

7:25588

2:2592

4:946

8:1704

3:7571

1:3512

9:12835

7:28628

4:732

9:13017

3:5837

8:1478

1:3800

2:1841

9:15837

3:7119

1:3806

4:681

8:1116

2:2070

2:3256

8:1109

4:762

3:9935

8:1340

4:847

1:4240

2:1451

8:813

4:260

4:608

2:1934

8:662

4:207

4:431

1:5789

2:2080

4:429

8:746

8:1174

4:356

4:819

4:1129

8:2737

4:828

تعداد تکرار	آیتم
57398	0
54216	7
53514	9
49706	3
49402	1
47530	2
47229	8
43592	4

همانطور که مشاهده می کنید به علت کمبود جای نمایش نمی توان تمامی گره های درخت را در حالت باز نشان داد اما در همان نگاه اول می توان تعداد تکرار آیتم ۰ را که بیشترین تکرار را دارد با فرکانس مربوطه در جدول تعداد تکرار و پس از آن آیتم بعدی یعنی آیتم شماره ۷ نیز مجموع دو گره بالایی فرکانس آن را نشان می دهد و به همین صورت تا آخرین آیتم.

Apriori

همانند بخش قبل فایل داده ها را انتخاب کرده و پس از بارگزاری تراکنش ها تعداد کل تراکنش ها و تعداد آیتم ها قابل مشاهده می شود سپس وارد tab page مربوط به Apriori می شویم و مقادیر تعداد ابعاد مطلوب و Min_Sup را وارد می کنیم و پس از آن بر روی دکمه انجام عملیات کلیک می کنیم. همانطور که در صورت مسئله خواسته شده تعداد آیتم های مطلوب را برابر ۱۰ و مثلاً Min_Sup را برابر ۱۰۰۰ قرار می دهیم و نتیجه تعداد تراکنش ها برای ترکیبات ۱ تا ۱۰ تایی که بیشتر از Min_Sup هستند را در جدول می بینیم.

The image shows three screenshots of the 'FP_Growth via Apriori' software interface. Each screenshot displays a table of results for the Apriori algorithm, with columns for 'Item' (آیتم), 'Count' (تعداد تکرار), and 'Support' (نسبت). The 'Support' column is calculated as 'Count / Total Transactions' (تعداد رکورد خوانده شده / اندازه ابعاد). The 'Min_Sup' threshold is set to 1000 in all three screenshots.

Screenshot 1 (Left): Shows the results for Min_Sup = 1000. The table lists items and their counts, with the support column showing values like 0.0001, 0.0002, etc.

Screenshot 2 (Middle): Shows the results for Min_Sup = 1000. The table lists items and their counts, with the support column showing values like 0.0001, 0.0002, etc.

Screenshot 3 (Right): Shows the results for Min_Sup = 1000. The table lists items and their counts, with the support column showing values like 0.0001, 0.0002, etc.

FP_Growth via Apriori

تنظیمات

فایل ترانسش ها:

Browse

E:\Projects\Farhaden Projects\Dataseting\Dataset.txt

100000

تعداد رکورد خوانده شده:

20

اندازه ابعاد:

بارگذاری ترانسش ها

FP_Growth

Apriori

تعداد ابعاد مطلوب:

10

تعداد ابعاد:

1000

Min_Sup

انجام عملیات

تعداد شکار	آیتم
1818	(4,5,6,7)
1883	(4,5,6,8)
1996	(4,5,6,2)
2041	(4,5,6,9)
2145	(4,5,7,8)
2227	(4,5,7,2)
2780	(4,5,7,9)
2812	(4,5,8,2)
2484	(4,5,8,9)
2302	(4,5,2,9)
2029	(4,6,7,8)
2353	(4,6,7,2)
2316	(4,6,7,9)
1931	(4,6,8,2)
1901	(4,6,8,9)
2068	(4,6,2,9)
2415	(4,7,8,2)
2948	(4,7,8,9)
2802	(4,7,2,9)
2449	(4,8,2,9)
1780	(5,6,7,8)
2012	(5,6,7,2)
2071	(5,6,7,9)
1952	(5,6,8,2)
1765	(5,6,8,9)
2014	(5,6,2,9)
2411	(5,7,8,2)
2947	(5,7,8,9)
2818	(5,7,2,9)
2376	(5,8,2,9)
2345	(6,7,8,2)
2467	(6,7,8,9)
2982	(6,7,2,9)
2332	(6,8,2,9)
2144	(7,8,2,9)
1028	(8,1,3,4,5)

FP_Growth via Apriori

تنظیمات

فایل ترانسش ها:

Browse

E:\Projects\Farhaden Projects\Dataseting\Dataset.txt

100000

تعداد رکورد خوانده شده:

20

اندازه ابعاد:

بارگذاری ترانسش ها

FP_Growth

Apriori

تعداد ابعاد مطلوب:

10

تعداد ابعاد:

1000

Min_Sup

انجام عملیات

تعداد شکار	آیتم
1887	(3,4,6,9)
2380	(3,4,7,8)
2343	(3,4,7,2)
2636	(3,4,7,9)
1940	(3,4,8,2)
2324	(3,4,8,9)
2271	(3,4,2,9)
1713	(3,5,6,7)
1796	(3,5,6,8)
2110	(3,5,6,2)
1829	(3,5,6,9)
2382	(3,5,7,8)
2430	(3,5,7,2)
2573	(3,5,7,9)
2250	(3,5,8,2)
2265	(3,5,8,9)
2482	(3,5,2,9)
2112	(3,6,7,8)
2897	(3,6,7,2)
2279	(3,6,7,9)
2185	(3,6,8,2)
2136	(3,6,8,9)
2376	(3,6,2,9)
2897	(3,7,8,2)
2838	(3,7,8,9)
2939	(3,7,2,9)
2512	(3,8,2,9)
1818	(4,5,6,7)
1883	(4,5,6,8)
1996	(4,5,6,2)
2041	(4,5,6,9)
2145	(4,5,7,8)
2227	(4,5,7,2)
2780	(4,5,7,9)
2812	(4,5,8,2)
2484	(4,5,8,9)

توجه: همانطور که مشاهده شد تعداد ابعاد مطلوب و همینطور Min_Sup در پروژه قابل تنظیم می باشد.