



تمرین چهارم



داده کاوی - پاییز ۱۳۹۹

گزارش کار

دانشکده علوم مهندسی

نام و نام خانوادگی:
مینو احمدی

استاد:
دکتر علی فهیم

مقدمه:

یکی از الگوریتم های مورد استفاده در datamining الگوریتم charm است که در این تمرین ما آن را پیاده سازی کرده ایم .



در این تمرین ایتm های close frequent را پیدا کردیم همین طور فراوانی

برخی ایتm هارا پیدا کردیم .

نحوه اجرای برنامه :

Python HW۰۴.py mushroom.txt MINSUP

خروجی برنامه :

فایل خروجی در کنار کد برنامه در فایل اپلود شده آورده شده .

قسمتی از خروجی با مین ساپ ۳۰۰۰:

۳۱۰۴ - {۳۴, ۱۱۶, ۹۰, ۸۶}

۳۱۰۴ - {۹۰, ۱۱۶, ۸۵}

$\{34, 116, 85, 86\} - 3130$
 $\{116, 85, 86\} - 3148$
 $\{90, 85, 6\} - 3020$
 $\{34, 85, 86, 6\} - 3096$
 $\{34, 85, 6\} - 3098$
 $\{85, 86, 6\} - 3102$
 $\{85, 6\} - 3152$
 $\{34, 36, 10, 86, 90\} - 3016$
 $\{90, 10, 36, 85\} - 3016$

..

$\{85\} - 8124$

۲۴۱

قسمتی از خروجی با مین ساپ ۵۰۰۰:

$\{59, 85\} - 5176$
 $\{34, 85, 86, 39\} - 5402$
 $\{85, 86, 39\} - 5420$
 $\{85, 39\} - 5612$
 $\{34, 36, 90, 86\} - 6272$
 $\{90, 36, 85\} - 6464$
 $\{34, 36, 85, 86\} - 6602$
 $\{36, 85, 86\} - 6620$
 $\{36, 85\} - 6812$
 $\{34, 90, 85, 86\} - 7288$
 $\{90, 34, 85\} - 7296$
 $\{90, 85\} - 7488$
 $\{34, 85, 86\} - 7906$
 $\{34, 85\} - 7914$

{۸۵, ۸۶} - ۷۹۲۴

{۸۵} - ۸۱۲۴

۱۶

توضیح :

اگر کد را بدون دادن دیتا و مین ساپ اجرا کنید ارور bad call را میگیریم .
ابتدا دیتا را در یک لیست ذخیره کرده . (نمیتوان از نامپای استفاده کرد زیرا در هر ردیف تعداد ایتم ها متفاوت است).

کلوز فریکوینت ایتم ست ها را در یک دیکشنری ذخیره میکنیم و مرتب میکنیم به طوری که ایتم ست ها را key و تعداد تکرار را value در نظر میگیریم .

تابع چارم را با مقادیر زیر فراخوانی میکنیم :

`charm(Dataset.list(), MINSUP:int, C:dict())`

خروجی ایتم ست ها و تعداد تکرار ان ها خواهد بود .

تابع چارم :

ورودی ها :

P : Dataset

Minsup: minsup

C : All closed frequent itemsets

ابتدا p را sort کرده (بر اساس افزایش ساپورت ها)

سپس روی p و iteration میزنیم تا زمانی که اگر ساپورت ایتمی از مین ساپ بیشتر بود ان دیتا را انتخاب میکنیم.

اگر دیتا ست جدید خالی نبود به صورت بازگشتی تابع چارم را با دیتاست جدید فراخوانی میکنیم.
بعد از ان ایتم ست های اولین حلقه را بررسی میکنیم اگر شرایط کلوز را داشت ان را به مجموعه کلوز اضافه میکنیم .

تعریف فریکوینت کلوز ها : مجموعه ای از اقلام فریکوینت باشند ولی سوپرستی از ان ها نباید موجود باشد که عدد ساپورت ان با ساپورت خودشان برابر باشد

$$\text{Sup}(x) \neq \text{sup}(\text{superset}(x))$$