Readme

Dataset:

The data size is 32736 and there are 15 features for each data. We are going to take four features, **"name"**, **"number"**, **"rating"**, **"add_time"** to implement multi-scale event detection.

How to use:

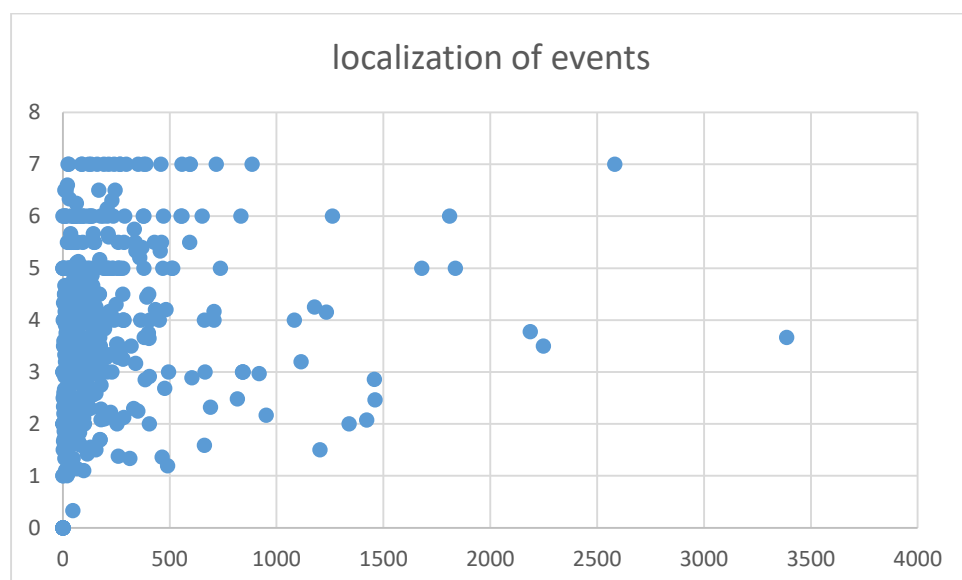1. Run data_seg.py
2. Run run.mat

First, We run the data_seg.py to separate the Chinese words with Jieba, so we generate the segmentation words dictionary ("data.xlsx") .

Later, we just need to run the "run.mat" whenever the input data file is named as "origin.xlsx". The result is the clusters that represent different events.
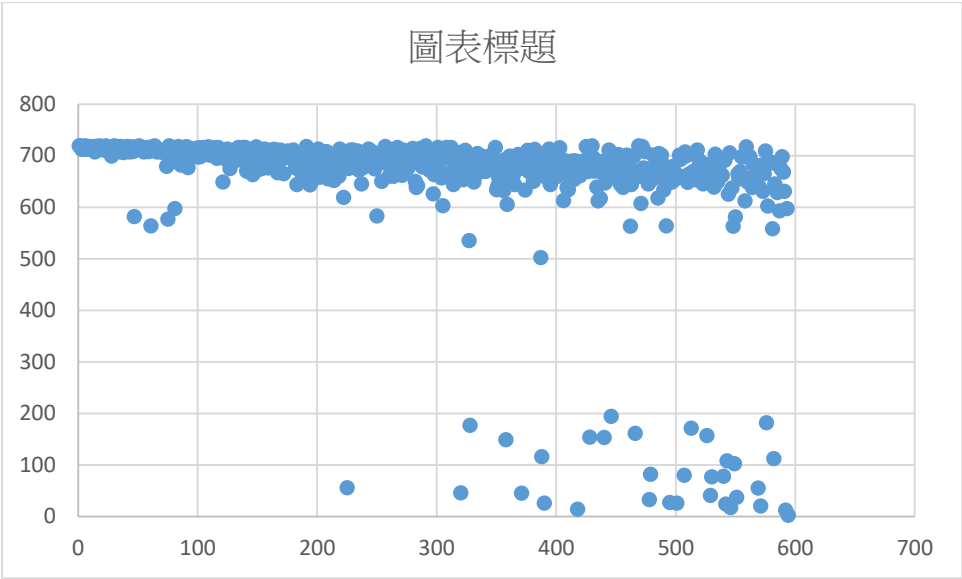
Figure:

There are 594 clusters that contain more or equal to 10 tweets.

The figure is the localization(x-axis: number, y-axis: rating) of the clusters

The figure is the time length that the cluster occur.

圖表標題

This sheet is the analysis of part of clusters. The last two columns represents the most frequent tweet and word occur in that cluster respectively.

numberlen	ratinglen	timelen	number of tweets	tweets	word
240	7	719	419	'玫瑰'	'玫瑰'
2583	7	719	417	'深層清潔面膜火山泥'	'深層'
1262	6	712	304	'玻尿酸面膜(單片裝)'	'玻'
557	7	717	284	'高效美顏組'	'高效'
388	7	718	245	'海洋礦物潔膚泥 Total Detox Body Mask'	'海洋'
377	6	719	230	'完美淨膚保濕乳液 Intacte'	'完美'
87	7	712	223	'極緻精華 Super Restorative Serum'	'極'
268	7	713	213	'全效眼唇卸妝液 Universal Eye Make-Up Remover'	'全效'
1680	5	718	191	'青春面皰組'	'青春'
553	6	714	191	'美白洗顏乳 Whitening Cleanser'	'美白'
47	4	716	189	'緊緻立顯面膜'	'緊'
455	5.333333	717	188	'保濕修護唇膏 Lip Care'	'保濕'
259	5.5	718	187	'毛孔管理清痘筆 Blemish Control'	'毛孔'
352	7	707	168	'白金級活顏再生乳霜/乳液 20+1 Nutri-Restructuring Rich Cream/Fluid'	'白金'
652	6	718	164	'淨白雙效卸妝乳'	'淨白'
379	6	718	163	'維他眼部保養組合'	'維他'
198	5	717	162	'膠原彈力精華乳液 Milk Concentrate'	'膠原'
268	5	719	162	'活膚精華 Nourishing Serum Capsules'	'活膚'
594	5.5	717	160	'活力爽膚水'	'活力'
236	5	715	144	'Q10緊緻煥膚霜 Q10 Cream'	'Q10'
1809	6	711	143	'蘆薈面膜撕除式'	'蘆薈'
127	4.25	712	139	'紅酒水療 Beautifying Gelee'	'紅'
718	7	719	131	'茶樹'	'茶樹'
429	5.5	716	128	'清爽控油脂日間保濕乳 Hydra-Matte Day Lotion'	'清爽'
74	5	712	123	'有機可循純淨潔顏露'	'有機'
379	7	707	120	'水漾保濕面膜'	'水漾'
459	7	715	116	'天然植物美容油(油性) Aromessence? Ylang Ylang'	'天然'
209	6	699	113	'黃玉霜 Topaz Cream'	'黃'
290	6	708	113	'亮采化粧水 Ultra-Brightening Whitening Lotion'	'亮'
470	6	719	112	'淨白日間乳液SPF10.PA+'	'淨'
509	5	710	112	'頂級保濕青春精華霜 Premier Moisture Complex'	'頂級'
3387	3.666667	715	111	'超完美修護乳液'	'超'