

Результаты проекта

"Разработка системы предсказания успешного завершения учебной дисциплины"

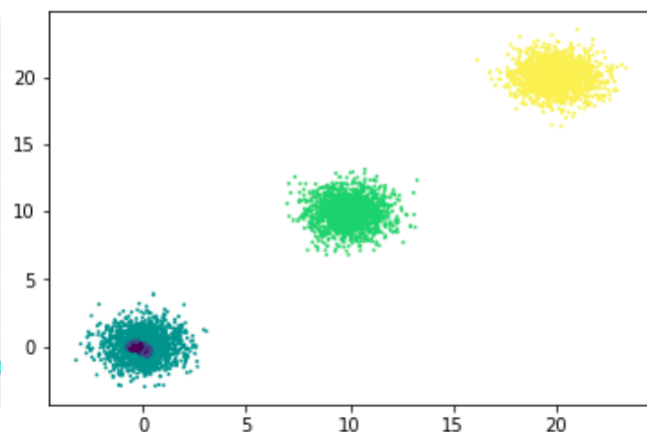
Наумова Анастасия 185 - 1

Результаты работы алгоритмов

Оба алгоритма BIRCH и OPTICS были имплементированы на языке Python3 и протестированы на сгенерированных данных.



Пример работы алгоритма BIRCH



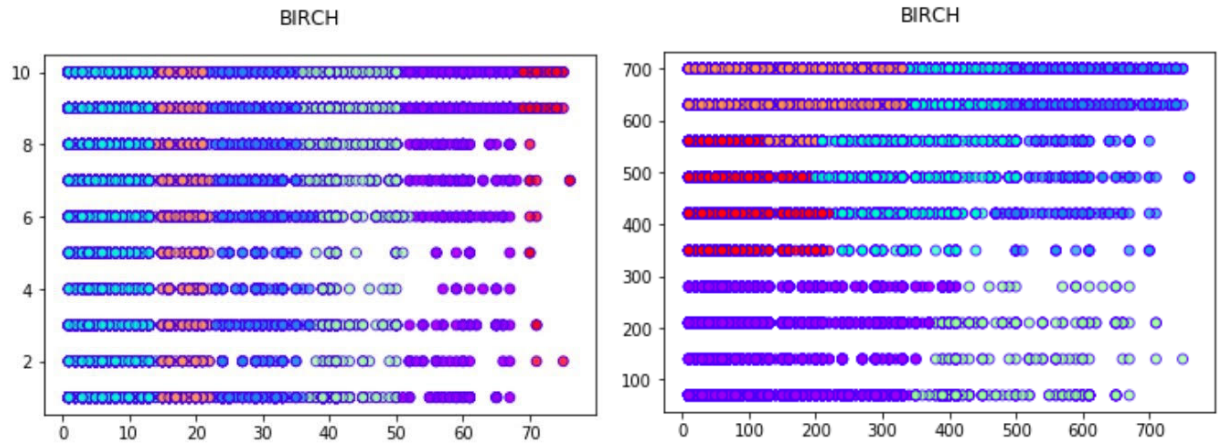
Пример работы алгоритма OPTICS

Можно заметить, что алгоритм OPTICS разделил нижнюю область на кластеры по различной плотности количества точек. Это происходит из-за свойств алгоритма.

Далее оба алгоритма были запущены на датасете отзывов о репетиторах объёмом около 300 000 элементов (данные собраны с сайта <https://repetitors.info/>). Рассмотрим подробнее работу каждого алгоритма в отдельности.

BIRCH

Каждой учебной дисциплине сопоставили число. Можно видеть, что в датасете их больше 70. В основном в каждой учебной дисциплине подавляющее большинство отзывов получала только одна оценка из всех возможных, кроме того, точки, сопоставленные оценкам одного предмета, находятся близко (что происходит из-за их огромного количества по сравнению с числом предметов), в связи с чем в подавляющем большинстве случаев все оценки полученные за один предмет были сгруппированы в один



Работа алгоритма на датафрейме

Работа на отмасштабированном датафрейме

кластер и объединены с ближайшими по номеру предметами, у которых близки оценки с наибольшим числом отзывов.

Для получения лучших результатов кластеризации параметры датасета можно отмасштабировать (существует более 70 различных дисциплин и всего 10 различных оценок).

экономика	более-менее	3	6
	вполне хорошо	3	4
	лучше не бывает!	3	201
	отлично	3	143
	плохо	3	2
	почти отлично	3	6
	так себе...	3	12
	ужасно	3	654
	хорошо	3	4
электротехника	более-менее	0	2
	вполне хорошо	0	3
	лучше не бывает!	0	40
	отлично	0	31
	плохо	0	1
	почти отлично	0	4
	почти хорошо	0	1
	так себе...	0	1
	ужасно	0	6
	хорошо	0	3
японский язык	более-менее	1	2
	вполне хорошо	1	9
	лучше не бывает!	1	145
	отлично	1	182
	плохо	1	2
	почти отлично	1	5
	почти хорошо	1	2
	так себе...	1	7
	ужасно	1	5
	хорошо	1	5

Работа алгоритма на датафрейме

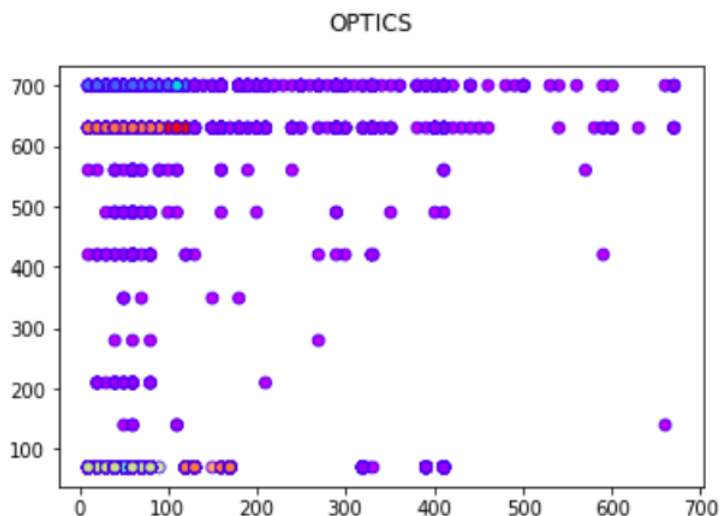
экономика	более-менее	0	6
	вполне хорошо	2	4
	лучше не бывает!	2	201
	отлично	2	143
	плохо	3	2
	почти отлично	2	6
	так себе...	3	12
	ужасно	3	654
	хорошо	2	4
электротехника	более-менее	3	2
	вполне хорошо	1	3
	лучше не бывает!	1	40
	отлично	1	31
	плохо	3	1
	почти отлично	1	4
	почти хорошо	1	1
	так себе...	3	1
	ужасно	3	6
	хорошо	1	3
японский язык	более-менее	0	2
	вполне хорошо	2	9
	лучше не бывает!	2	145
	отлично	2	182
	плохо	0	2
	почти отлично	2	5
	почти хорошо	2	2
	так себе...	0	7
	ужасно	3	5
	хорошо	2	5

Работа на отмасштабированном датафрейме

При выводе результатов в виде таблицы, можно видеть аналогичный результат. В первом случае, все оценки одного предмета отнесены к одному кластеру, в другом случае (при масштабировании) кластеры различны.

OPTICS

Поскольку алгоритм OPTICS имеет квадратичную сложность, локально я не смогла протестировать его на полном датасете в связи с нехваткой памяти. При использовании алгоритма сразу отмечалась его чувствительность к задаваемым параметрам и выделялась его особенность, создание кластеров разной плотности.

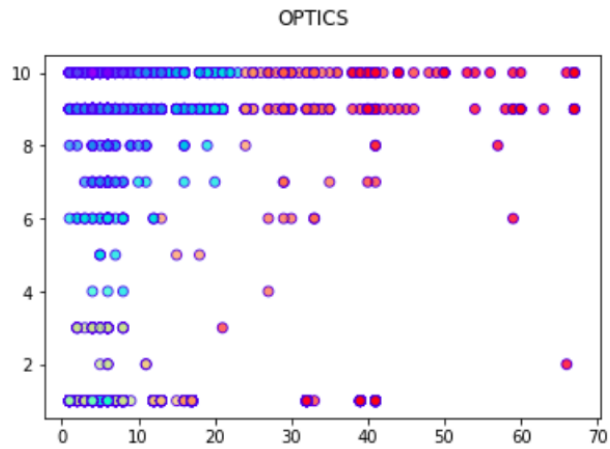


Работа алгоритма на уменьшенном в объёме и отмасштабированном датафрейме

На отмасштабированном датафрейме можно увидеть ярковыраженный разреженный кластер и несколько плотных кластеров. Это же видно и в таблице. Оценки по каждому из предметов, которые редко встречаются в отзывах, собираются в один кластер.

логика	лучше не бывает!	0	1
	плохо	0	1
логопеды	вполне хорошо	0	1
	лучше не бывает!	1	5
		2	11
	отлично	5	6
		6	5
	плохо	0	2
макроэкономика	почти отлично	0	2
	отлично	0	1
математика	более-менее	0	1
	вполне хорошо	0	9
	лучше не бывает!	1	106
	отлично	5	99
	плохо	0	2
	почти отлично	0	11
	так себе...	0	7
	ужасно	3	1
		4	687
	хорошо	0	10
математический анализ	лучше не бывает!	0	5
менеджмент	отлично	0	1
микроэкономика	лучше не бывает!	0	1
музыка	лучше не бывает!	0	6
	отлично	0	4
	хорошо	0	1

При запуске алгоритма на разных параметрах получаем сильно отличающиеся кластеризации.



В данном случае выделено большее число разреженных кластеров, а плотные кластеры разделены на более мелкие.

REST API

POST http://127.0.0.1:5000/all_exps Send

Params Authorization Headers (7) Body Pre-request Script Tests Settings

Query Params

KEY	VALUE	DESCRIPTION
Key	Value	Description

Body Cookies Headers (4) Test Results Status: 201 CREATED Time: 31 ms Size: 152 B Save

Pretty Raw Preview Visualize JSON

```
1 {}
```

Создание нового пустого эксперимента.

GET http://127.0.0.1:5000/all_exps Send

Params Authorization Headers (6) Body Pre-request Script Tests Settings

Query Params

KEY	VALUE	DESCRIPTION
Key	Value	Description

Body Cookies Headers (4) Test Results Status: 200 OK Time: 30 ms Size: 1.79 KB Save

Pretty Raw Preview Visualize JSON

```
1 {
2   "1": [ ...
153 ],
154   "2": {}
155 }
```

Множество всех экспериментов, один из которых пуст.

POST http://127.0.0.1:5000/all_exps/2?min_pts=30&eps=5 Send

Params Authorization Headers (7) Body Pre-request Script Tests Settings

Query Params

	KEY	VALUE	DESCRIPTION
<input checked="" type="checkbox"/>	min_pts	30	
<input checked="" type="checkbox"/>	eps	5	
	Key	Value	Description

Body Cookies Headers (4) Test Results Status: 201 CREATED Time: 284 ms Size: 1.18 KB Save

Pretty Raw Preview Visualize JSON

```

1 [
2   1,
3   1,
4   1,
5   1,
6   1,
7   1,

```

Запуск нового эксперимента с введёнными параметрами.

GET http://127.0.0.1:5000/all_exps Send

Params Authorization Headers (6) Body Pre-request Script Tests Settings

Query Params

	KEY	VALUE	DESCRIPTION
	Key	Value	Description

Body Cookies Headers (4) Test Results Status: 200 OK Time: 14 ms Size: 3.4 KB Save

Pretty Raw Preview Visualize JSON

```

1 {
2   > "1": [ ...
153 ],
154 > "2": [ ...
305 ]
306 }

```

Результаты нового эксперименты добавлены к остальным.


GET ▼ http://127.0.0.1:5000/all_exps Send ▼

Params Authorization Headers (6) Body Pre-request Script Tests Settings

Query Params

KEY	VALUE	DESCRIPTION	..
Key	Value	Description	

Body Cookies Headers (4) Test Results Status: 200 OK Time: 14 ms Size: 3.4 KB Save

Pretty Raw Preview Visualize JSON ▼ 

```

1 {
2   > "1": [ ...
153 ],
154   "2": [
155     1,
156     1,
157     1,
158     1,
159     1,
160     1,
161     1,

```

Просмотр результатов среди всех экспериментов.


GET ▼ http://127.0.0.1:5000/all_exps/2?min_pts=30&eps=5 Send ▼

Params ● Authorization Headers (6) Body Pre-request Script Tests Settings

Query Params

KEY	VALUE	DESCRIPTION	..
<input checked="" type="checkbox"/> min_pts	30		
<input checked="" type="checkbox"/> eps	5		
Key	Value	Description	

Body Cookies Headers (4) Test Results Status: 200 OK Time: 9 ms Size: 1.17 KB Save

Pretty Raw Preview Visualize JSON ▼ 

```

1 [
2   1,
3   1,
4   1,
5   1,
6   1,
7   1,

```

Просмотр результатов отдельного эксперимента.


POST http://127.0.0.1:5000/all_exps/2?min_pts=40&eps=6 Send

Params Authorization Headers (7) Body Pre-request Script Tests Settings

Query Params

	KEY	VALUE	DESCRIPTION	••
<input checked="" type="checkbox"/>	min_pts	40		
<input checked="" type="checkbox"/>	eps	6		
	Key	Value	Description	

Body Cookies Headers (4) Test Results Status: 409 CONFLICT Time: 38 ms Size: 174 B Save

Pretty Raw Preview Visualize JSON 

```
1 "begin new experiment"
```

При повторном запуске эксперимента (со старыми или с новыми значениями параметров) будет выведена ошибка с просьбой начать новый эксперимент.