

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

УДК 004.8

Отчет об исследовательском проекте на тему:

Кластеризация аудио

(промежуточный, этап 1)

Выполнил студент:

группы #БПМИ213, 3 курса

Бонич Дмитрий Сергеевич

Принял руководитель проекта:

Сендерович Александра Леонидовна

Научный сотрудник

Факультет компьютерных наук НИУ ВШЭ

Москва 2024

Содержание

Аннотация	3
1 Введение	4
1.1 Постановка задачи	4
1.2 Метрики качества	4
1.3 Классические методы решения	5
1.3.1 K-means	5
1.3.2 DBSCAN	5
1.3.3 MeanShift	5
1.3.4 AgglomerativeClustering	5
2 Обзор литературы	5
2.1 DeepCluster	6
2.2 SPICE	6
3 План дальнейшей работы	7
4 Примеры	7
4.1 Ссылки на статьи	7
4.2 Рисунки	7
4.3 Таблицы	7
4.4 Формулы	8
Список литературы	9

Аннотация

Существующие методы глубинного обучения для кластеризации изображений работают довольно хорошо. Однако вопрос качественной кластеризации аудио остается открытым. В этой работе мы планируем адаптировать лучшие методы кластеризации изображений для задачи кластеризации аудио.

Ключевые слова

Глубинное обучение, обучение без учителя, кластеризация

1 Введение

1.1 Постановка задачи

В задаче классификации мы имеем обучающую выборку, где для каждого объекта известен его класс и от нас требуется моделировать распределение классов на пространстве объектов. Задача кластеризации более сложная – необходимо разбить объекты на осмысленные группы не зная ни самих групп, ни их распределения.

1.2 Метрики качества

Чтобы замерить качество кластеризации как правило метод решает задачу на размеченном для кластеризации датасете, разумеется не используя метки классов. Полученные после кластеризации номера кластеров будем называть *псевдометками*. Псевдометки сравниваются с настоящими метками и качество определяется как некоторый вид корреляции между ними.

Одной из самых популярных метрик качества является NMI(Normalized Mutual Information). Пусть у нас есть пара случайных величин X и Y , тогда:

$$\text{NMI}(X, Y) = \frac{\text{KL}(p_{(X,Y)} | p_X \otimes p_Y)}{\sqrt{H(X)H(Y)}}$$

При подсчете NMI мы считаем распределения меток и псевдометок за X и Y и вычисляем *оценку максимального правдоподобия*¹. NMI принимает значения от 0 до 1.

Еще одной используемой метрикой является ARI(Adjusted Rand Index). Это скорректированная версия метрики RI² (Rand Index). Определяется ARI следующим образом:

$$\text{ARI}(X, Y) = \frac{\text{RI}(X, Y) - \text{E}[\text{RI}(X, Y)]}{1 - \text{E}[\text{RI}(X, Y)]}$$

ARI принимает значения от 0 до 1.

Также в качестве метрики качества можно использовать долю правильных ответов как и в задаче классификации. Однако предварительно необходимо решить *задачу о назначениях*³. между псевдометками и метками. Мы переименуем псевдометки так, чтобы достичь максимального качества. Затем на переименованных псевдометках мы посчитаем долю правильных ответов, которую и используем в качестве метрики качества.

¹далее ОМП

²https://en.wikipedia.org/wiki/Rand_index

³https://en.wikipedia.org/wiki/Assignment_problem

1.3 Классические методы решения

С точки зрения классических методов кластеризуемые объекты это точки в многомерном пространстве. Такие методы обычно имеют итерационную природу и решают задачу выделения кластеров точек находя их скопления, области с повышенной плотностью, некоторые структуры. Посмотрим на несколько примеров.

1.3.1 K-means

Одним из самых простых и известных методов является K-means. Он работает на базе ЕМ-алгоритма⁴. K-means итерационно пытается найти два неизвестных набора переменных — номера кластеров для объектов и центры этих кластеров. Количество кластеров фиксировано и задается до начала работы алгоритма в качестве гиперпараметра k .

В начале своей работы классический K-means инициализирует все центры кластеров случайно. Затем чередуются Е и М шаги до сходимости. На Е-шаге мы назначаем каждому объекту номер кластера к центру которого объект ближе всего в качестве псевдометки. На М-шаге мы вычисляем ОМП для каждого центра кластера, то есть берем в качестве центра кластера усредненную точку из всех объектов с псевдометкой этого кластера.

1.3.2 DBSCAN

1.3.3 MeanShift

1.3.4 AgglomerativeClustering

2 Обзор литературы

При кластеризации сложных объектов таких как изображения или аудио недостаточно вытянуть данные объекта в вектор и применить классический алгоритм кластеризации для полученных точек. Чтобы алгоритмы кластеризации хорошо работали входное пространство точек должно обладать некоторыми свойствами. В идеале близкие в этом пространстве точки должны принадлежать к одной группе.

Таким образом задачу кластеризации сложных объектов можно разделить на две части — получение представлений объектов образующих пригодное для кластеризации пространство точек и сама кластеризация этих представлений. Метод переводящий объекты в признаки будем называть *энкодером*.

⁴https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm

2.1 DeepCluster

Одним из первых методов использующих глубинное обучение для кластеризации изображений был DeepCluster [1]. DeepCluster использует сверточную нейросеть f_θ в качестве энкодера. К полученным признакам применяется K-means и мы получаем псевдометку y_n для каждого изображения x_n . Затем к сверточной сети прикрепляется классификационная голова g_W . g_W предсказывает вероятности кластеров для объекта по его признакам полученным из f_θ . Псевдометки используются как настоящие для подсчета логистической функции потерь. Таким образом мы решаем следующую задачу оптимизации:

$$\min_{\theta, W} \frac{1}{N} \sum_{n=1}^N l(g_W(f_\theta(x_n)), y_n)$$

Здесь N – это размер батча, l логистическая мультиномиальная функция потерь

По усредненному значению функции потерь для батча делается проход назад и веса θ и W обновляются стохастическим градиентным спуском. Затем шаги кластеризации и оптимизации параметров сети повторяются заданное количество эпох.

Однако в такой постановке метод выражается в тривиальные решения. Их существует 2 типа - пустые кластеры и тривиальная параметризация

Чтобы не допустить возникновения пустых кластеров сделаем следующую модификацию K-means. Пусть на некоторой итерации у нас появился пустой класс А. Тогда возьмем случайный непустой класс В. Добавим к центру В шум получив новый центр для кластера А. Переназначим классы точек из В так чтобы каждая точка имела класс ближайшего центра кластера.

Тривиальная параметризация возникает когда распределение классов становится сильно неравномерным. Тогда классификационной голове становится выгодно выдавать только несколько наиболее часто встречающихся классов игнорируя остальные. Чтобы этого избежать необходимо сэмплировать объекты для батча из равномерного распределения по псевдометкам с предыдущего шага.

2.2 SPICE

SPICE предложенный в [8], является

3 План дальнейшей работы

4 Примеры

4.1 Ссылки на статьи

Ссылки на статьи оформляются с помощью пакета `biblatex`, например [2]. В описании статье в `bib` файле нужно обязательно указывать место публикации работы (журнал или конференцию) и год. Обратите внимание, что для описания статей из разных источников в списке литературы используются разные команды в `bib` файле: статья из журнала [4], статья с конференции [2], книга [7], глава книги [6]. Если статья еще не опубликована нигде, а только выложена на `arXiv`, то на нее тоже можно сослаться [3], но предпочтительно ссылаться на опубликованную версию, если она уже существует. Если вы хотите сослаться на сайт, то можно либо так же внести его в список литературы [5] (рекомендуется, если таких ссылок у вас много из-за особенностей темы вашего проекта), либо использовать ссылку внизу страницы⁵. При работе с онлайн ресурсами не забывайте указывать дату обращения к этому ресурсу, так как в отличие от опубликованных статей, эти ресурсы могут измениться в любой момент.

4.2 Рисунки

Все рисунки в тексте должны иметь подписи и вы на них должны ссылаться в тексте. Например, на Рисунке 4.1 изображен пример графика. Не забывайте подписывать все оси на графиках, добавлять легенду и пояснять все обозначения, а также используйте адекватного размера шрифты и толщину линий на графиках (все должно быть видно и понятно без многократного увеличения). На рисунке из примера явно не хватает обозначения синей линии в легенде.

4.3 Таблицы

Все таблицы в тексте тоже должны иметь подписи и вы на них должны ссылаться в тексте. Например, в Таблице 4.1 показаны результаты примерного эксперимента.

⁵Книги доступны по ссылке: <http://www-cs-faculty.stanford.edu/~uno/abcde.html>, дата обр. 16.05.2013

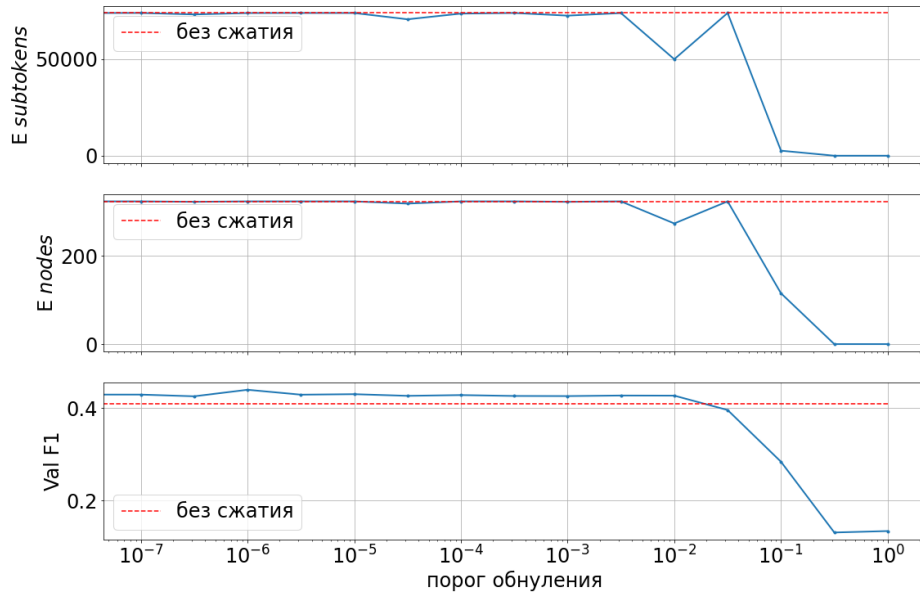


Рис. 4.1: Пример графика. Тут должна быть подпись, поясняющая что происходит на рисунке (краткая, но достаточная для понимания основной идеи графика).

Таблица 4.1: Пример таблички. Тут должна быть подпись, поясняющая что происходит в таблице (краткая, но по делу).

	Val			Test			nodes	subtokens
	Prec	Rec	F1	Prec	Rec	F1		
запуск 1	0.4894	0.3775	0.4263	0.4824	0.3683	0.4177	10029	179
запуск 2	0.4887	0.3739	0.4237	0.4891	0.3724	0.4228	10039	177
запуск 3	0.4820	0.3751	0.4219	0.4838	0.3677	0.4178	10037	180
среднее	0.4867	0.3755	0.4239	0.4851	0.3695	0.4195		
дисперсия	0.0041	0.0019	0.0022	0.0036	0.0025	0.0029		

4.4 Формулы

Формулы стоит центрировать, а также нумеровать, если вы ссылаете на них в тексте. Также не забывайте пояснять все обозначения в формулах. Например, запишем следующую задачу оптимизации:

$$\theta^* = \min_{\theta} F(\theta), \quad (1)$$

где F – квадратичная функция от параметра θ . При необходимости, далее в тексте можно сослаться на формулу (1). При этом, в зависимости от конкретных формул, можно использовать разные слова: формула, уравнение, задача оптимизации и т.п.

Список литературы

- [1] Mathilde Caron, Piotr Bojanowski, Armand Joulin и Matthijs Douze. “Deep Clustering for Unsupervised Learning of Visual Features”. в: *Proceedings of the European Conference on Computer Vision (ECCV)*. сент. 2018.
- [2] Nadezhda Chirkova, Ekaterina Lobacheva и Dmitry Vetrov. “Bayesian Compression for Natural Language Processing”. в: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018.
- [3] Nadezhda Chirkova, Ekaterina Lobacheva и Dmitry Vetrov. “Bayesian Compression for Natural Language Processing”. в: *arXiv preprint, arXiv:1810.10927, version 2* (2018).
- [4] George D. Greenwade. “The Comprehensive Tex Archive Network (CTAN)”. в: *TUGBoat* 14.3 (1993), с. 342—351.
- [5] Donald Knuth. *Knuth: Computers and Typesetting*. URL: <http://www-cs-faculty.stanford.edu/~uno/abcde.html> (дата обр. 16.05.2013).
- [6] Donald E. Knuth. “Fundamental Algorithms”. в: Addison-Wesley, 1973. гл. 1.2.
- [7] Donald E. Knuth. *The Art of Computer Programming*. Four volumes. Seven volumes planned. Addison-Wesley, 1968.
- [8] Chuang Niu и Ge Wang. *SPICE: Semantic Pseudo-labeling for Image Clustering*. 2021. arXiv: [2103.09382 \[cs.CV\]](https://arxiv.org/abs/2103.09382).