
Классификация дуктов для поиска ключевых слов в рукописном контексте

Феоктистов Дмитрий Дмитриевич
ВМК МГУ
feoktistovdd@my.msu.ru

Местецкий Леонид Моисеевич
ВМК МГУ
mestlm@yandex.ru

2023

Аннотация

В работе решается задача поиска ключевых слов в рукописном контексте. Пусть даны изображения некоторого рукописного файла, в котором требуется находить все вхождения введенного слова. Решение этой задачи может значительно упростить работу с архивными данными. Для решения задачи предлагается работать со словами на уровне их штрихового представления и определить метрику на множестве штрихов, с помощью которой можно решить задачу их классификации. Данный классификатор предлагается использовать как часть ранжирующего алгоритма. Для демонстрации результатов работы используются изображения работ участников "Тотального диктанта".

Ключевые слова: Обнаружение ключевых слов · Преобразование Фурье · Обучение метрик · Компьютерное зрение

1 Введение

Задача поиска ключевых слов в рукописном контексте является актуальной последние десятилетия в силу того, что полноценное распознавание рукописного текста не достигло той точности, при которой можно выполнять полноценное чтение документа и поиск в нем [Gurav et al., 2023, Souibgui et al., 2022]. Одной из главной целей при решении задачи является достижение ее применимости для навигации в архивных документах [Riesen et al., 2014, Giotis et al., 2015].

Изначально задача поиска ключевых слов решалась для напечатанных документов, в которых используется курсивный шрифт [Keaton et al., 1997]. Постепенно задача стала усложняться, и появились алгоритмы, выполняющие поиск в рукописных текстах [Rath and Manmatha, 2003], .

Существует несколько разновидностей рассматриваемой задачи. В первом варианте запрос задается в виде примера искомого слова в данном документе, такой подход позволяет обойти проблему разнообразия почерков [Giotis et al., 2015, Retsinas et al., 2019]. Более общий подход предполагает, что запрос является строкой [Retsinas et al., 2021]. Путь решения задачи также существует несколько. Первый предполагает использование глубоких нейронных сетей напрямую [Retsinas et al., 2023, Sfikas et al., 2022, Cascianelli et al., 2022], второй – использование нейронных сетей для построения векторных представлений слов для осуществления последующего поиска [Retsinas et al., 2021, Krishnan et al., 2023, Jemni et al., 2023]. Третий же стоит отдельно и предполагает использование признаков, полученных из изображения с помощью некоторых алгоритмов, с последующим их применением в алгоритмах машинного обучения. При этом может использоваться как и дискретное представление слов [Sfikas et al., 2017, Yousfi et al., 2021, Kundu et al., 2021], так и непрерывные признаки, полученные из скелетных графов [Giotis et al., 2015, Ameri et al., 2017, Stauffer et al., 2016]. Последний подход является актуальным и по сей день [Yousfi et al., 2021, Kundu et al., 2021, Banerjee et al., 2022], так как есть экспериментально подтвержденная гипотеза о том, что увеличение количества параметров в нейронных сетях не приводит к улучшению результатов поиска [Rusakov et al., 2018]. Наиболее актуальные алгоритмы используют комбинацию

описанных подходов, применяя как и классические признаки, так и полученные с помощью обучения нейронной сети [Jemni et al., 2023, Omayio et al., 2023].

Отдельно стоит выделить алгоритмы, выполняющие сопоставление запроса и слова с помощью различных метрик. Этот подход интересен тем, что метрики являются интерпретируемыми [Ameri et al., 2017, Stauffer et al., 2016]. При этом большинство метрик вычисляются долго, из-за чего требуется разработка специальных фильтров, позволяющих ускорять поиск [Stauffer et al., 2020].

В данной работе предлагается новый метод решения задачи поиска ключевых слов в рукописном контексте, основанный на следующей гипотезе: все почерки являются вариацией некоторого эталонного. Действительно, уже несколько веков обучение письму производится с помощью прописей, в которых не меняются правила написания штрихов (дуктов), из которых строятся буквы. Соответственно, если задать метрику близости на штрихах и произвести их классификацию, то получится обоснованное представление слова в виде частично упорядоченного множества (порядок возникает из-за порядка появления штрихов). Имея описанное выше представление, можно производить поиск ключевых слов с помощью различных мер схожести множеств штрихов. Также описанная выше гипотеза позволяет решать задачу поиска ключевых слов в формулировке, в которой запрос передается в виде строки, которая преобразуется в изображение, написанное с помощью эталонного почерка (в данной работе с помощью шрифта Propisi). Насколько авторам известно, в подобной формулировке задача решалась в [Пронина, 2023, Пацация, 2023], где для сравнения штрихов использовалось расстояние Фреше и DTW соответственно, но в этих работах рассматривался подход основанный на кластеризации, в то время как классификация способна дать больше информации. Таким образом в данной статье описаны:

1. Алгоритм выделения штрихов из изображения рукописного слова, основанный на построении скелета изображения.
2. Обучение метрики для задачи классификации на пространстве дуктов, основанной на преобразовании Фурье ломанных, описывающих штрихи.

Для демонстрации результатов работы полученного алгоритма используются изображения работ участников "Тотального диктанта". Эти данные позволяют показать работоспособность подхода для различных почерков.

2 Задача поиска ключевых слов в рукописном контексте

Цель поиска ключевых слов состоит в том, чтобы извлечь изображения слов из заданной коллекции изображений документов и ранжировать их по релевантности определенному запросу. Запрос может быть либо уже найденным пользователем изображением слова (т.е. обрезанное изображение документа) (QbE), либо строкой, отвечающей слову, которое необходимо найти (QbS). Главным преимуществом второй постановки является отсутствие необходимости искать пользователю запрашиваемое слово самостоятельно, что сильно экономит время при работе с редкими словами, поэтому в данной работе рассматривается вариант QbS. Формализуем данную постановку. Пусть у нас есть множество строк S и множество изображений слов W . Пусть $c(w)$ - строка, соответствующая изображению w . Тогда нам надо найти такое отображение $f : S \times W \rightarrow \mathbb{R}$, что:

$$\operatorname{argmin}_w f(c(w_1), w) = w_1 \quad \forall w_1 \in W$$

В задаче поиска ключевых слов общепринятой метрикой является *mean average precision*:

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AP(q)$$

Где $AP(q)$ – площадь под кривой *precision – recall* кривой для запроса q .

3 Задача метрической классификации штрихов

Математической моделью штриха является ломаная, которая может быть как замкнута, так и являться цепью. Существенным ограничением является отсутствие самопересечений у ломанных. Для сравнения штрихов предлагается ввести метрику, которую можно использовать в алгоритме классификации с помощью k ближайших соседей. В качестве метрики задачи используется *accuracy*.

Примеры синтетически сгенерированных данных

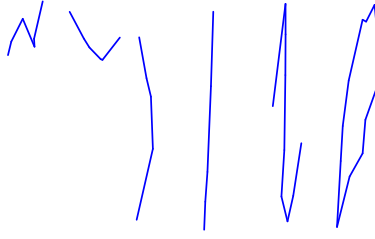


Рис. 1: Примеры сгенерированных данных, каждый объект является представителем своего класса.

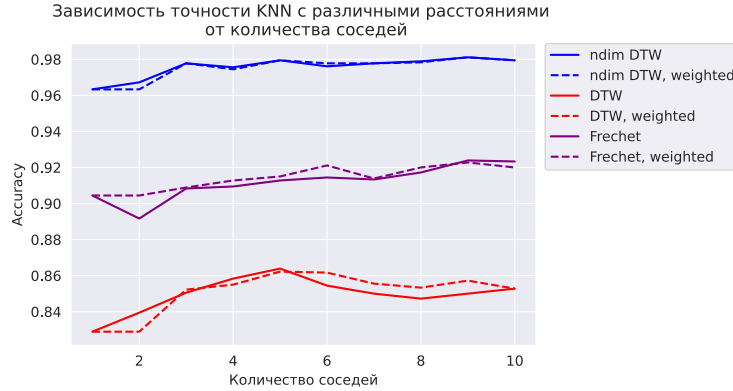


Рис. 2: Зависимость *accuracy* на кросс-валидации от количества соседей в алгоритме *KNN*, синим отмечены значения для двумерного DTW, фиолетовым для расстояния Фреше, красным – для DTW на коэффициентах сплайнов. Пунктиром отмечены запуски, при которых предсказания зависели от расстояний до соседей, а не только от их классов.

4 Вычислительный эксперимент

В качестве базового эксперимента предлагается решить задачу шестиклассовой классификации 1800 сгенерированных штрихов. Каждый экземпляр датасета представляет собой искажение одного из типов ломанной, которая потенциально может встречаться в виде штриха. Подробно описывать процесс генерации не буду, так как 27.10.2023 в 11:00 планируется утвердить классы и начать разметку данных, поэтому к middle talk здесь будет описание реальных данных, а выполнять одну и ту же работу дважды не хочется. Пример данных изображен на Рис. 1.

Будем сравнивать метрики, предложенные в [Пронина, 2023, Пацация, 2023]: расстояние Фреше между ломанными, DTW расстояние на коэффициентах сплайнов, аппроксимирующих ломанную, двумерное DTW расстояние между ломанными. Во всех трех случаях возьмем предобработку штрихов, которую использовали авторы. Расстояния будем сравнивать с помощью алгоритма *KNN*, как с равномерными весами, так и с весами, зависящими от расстояния. В качестве метрики воспользуемся средним значением *accuracy* на кросс-валидации на пяти фолдах. Результаты эксперимента представлены на Рис. 2. Исходя из графика, можно сделать несколько выводов:

1. Многомерное расстояние DTW является лучшим для классификации штрихов.
2. Стоит брать количество соседей около 5, так как с увеличением их количества точность не увеличивается, но может уменьшиться устойчивость алгоритма.
3. Разницы между взвешенным и невзвешенным вариантами *KNN* в рассматриваемой задаче нет.

Список литературы

- Aniket Gurav, Joakim Jensen, Narayanan C. Krishnan, and Sukalpa Chanda. Respho(sc)net: A zero-shot learning framework for norwegian handwritten word image recognition. In Antonio Pertusa, Antonio Javier Gallego, Joan Andreu Sánchez, and Inês Domingues, editors, *Pattern Recognition and Image Analysis*, pages 182–196, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-36616-1.
- Mohamed Ali Souibgui, Alicia Fornés, Yousri Kessentini, and Beáta Megyesi. Few shots are all you need: A progressive learning approach for low resource handwritten text recognition. *Pattern Recognition Letters*, 160:43–49, 2022. ISSN 0167-8655. doi:<https://doi.org/10.1016/j.patrec.2022.06.003>. URL <https://www.sciencedirect.com/science/article/pii/S016786552200191X>.
- Kaspar Riesen, Darko Brodić, Zoran N. Milivojević, and Čedomir A. Maluckov. Graph based keyword spotting in medieval slavic documents – a project outline. In Marinos Ioannides, Nadia Magnenat-Thalmann, Eleanor Fink, Roko Žarnić, Alex-Yianing Yen, and Ewald Quak, editors, *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, pages 724–731, Cham, 2014. Springer International Publishing. ISBN 978-3-319-13695-0.
- Angelos P. Giotis, Giorgos Sfikas, Christophoros Nikou, and Basilis Gatos. Shape-based word spotting in handwritten document images. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 561–565, 2015. doi:10.1109/ICDAR.2015.7333824.
- P. Keaton, H. Greenspan, and R. Goodman. Keyword spotting for cursive document retrieval. In *Proceedings Workshop on Document Image Analysis (DIA’97)*, pages 74–81, 1997. doi:10.1109/DIA.1997.627095.
- T.M. Rath and R. Manmatha. Word image matching using dynamic time warping. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–II, 2003. doi:10.1109/CVPR.2003.1211511.
- George Retsinas, Georgios Louloudis, Nikolaos Stamatopoulos, and Basilis Gatos. Efficient learning-free keyword spotting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1587–1600, 2019. doi:10.1109/TPAMI.2018.2845880.
- George Retsinas, Giorgos Sfikas, Christophoros Nikou, and Petros Maragos. From seq2seq to handwritten word embeddings. In *British Machine Vision Conference (BMVC)*, 2021.
- George Retsinas, Giorgos Sfikas, and Christophoros Nikou. Keyword spotting simplified: A segmentation-free approach using character counting and ctc re-scoring. In Gernot A. Fink, Rajiv Jain, Koichi Kise, and Richard Zanibbi, editors, *Document Analysis and Recognition - ICDAR 2023*, pages 446–464, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-41676-7.
- Giorgos Sfikas, George Retsinas, Angelos P. Giotis, Basilis Gatos, and Christophoros Nikou. Keyword spotting with quaternionic resnet: Application to spotting in greek manuscripts. In Seiichi Uchida, Elisa Barney, and Véronique Eglin, editors, *Document Analysis Systems*, pages 382–396, Cham, 2022. Springer International Publishing. ISBN 978-3-031-06555-2.
- Silvia Cascianelli, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Boosting modern and historical handwritten text recognition with deformable convolutions. *International Journal on Document Analysis and Recognition (IJDAR)*, 25(3):207–217, Sep 2022. ISSN 1433-2825. doi:10.1007/s10032-022-00401-y. URL <https://doi.org/10.1007/s10032-022-00401-y>.
- Praveen Krishnan, Kartik Dutta, and C. V. Jawahar. Hwnet v3: a joint embedding framework for recognition and retrieval of handwritten text. *International Journal on Document Analysis and Recognition (IJDAR)*, Jan 2023. ISSN 1433-2825. doi:10.1007/s10032-022-00423-6. URL <https://doi.org/10.1007/s10032-022-00423-6>.
- Sana Khamekhem Jemni, Sourour Ammar, Mohamed Ali Souibgui, Yousri Kessentini, and Abbas Cheddad. St-keys: Self-supervised transformer for keyword spotting in historical handwritten documents, 2023.
- Giorgos Sfikas, George Retsinas, and Basilis Gatos. A phoc decoder for lexicon-free handwritten word recognition. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 513–518, 2017. doi:10.1109/ICDAR.2017.90.
- Douaa Yousfi, Abdeljalil Gattal, Chawki Djeddi, Imran Siddiqi, and Ameer Bensefia. Keyword spotting in modern handwritten documents using obifs. In *Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, pages 240–250. Springer, 2021.
- Subhranil Kundu, Samir Malakar, Zong Woo Geem, Yoon Young Moon, Pawan Kumar Singh, and Ram Sarkar. Hough transform-based angular features for learning-free handwritten keyword spotting. *Sensors*, 21(14):4648, 2021.

- M Ameri, Michael Stauffer, Kaspar Riesen, T Bui, and Andreas Fischer. Keyword spotting in historical documents based on handwriting graphs and hausdorff edit distance. In International graphonomics society conference, pages 105–108, 2017.
- Michael Stauffer, Andreas Fischer, and Kaspar Riesen. Graph-based keyword spotting in historical handwritten documents. In Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, S+ SSPR 2016, Mérida, Mexico, November 29–December 2, 2016, Proceedings, pages 564–573. Springer, 2016.
- Debanshu Banerjee, Pratik Bhowal, Samir Malakar, Erik Cuevas, Marco Pérez-Cisneros, and Ram Sarkar. Z-transform-based profile matching to develop a learning-free keyword spotting method for handwritten document images. International Journal of Computational Intelligence Systems, 15(1):93, 2022.
- Eugen Rusakov, Sebastian Sudholt, Fabian Wolf, and Gernot A Fink. Expolring architectures for cnn-based word spotting. arXiv preprint arXiv:1806.10866, 2018.
- Enock Osoro Omayio, Sreedevi Indu, and Jeebananda Panda. Word spotting and character recognition of handwritten hindi scripts by integral histogram of oriented displacement (ihod) descriptor. Multimedia Tools and Applications, pages 1–28, 2023.
- Michael Stauffer, Andreas Fischer, and Kaspar Riesen. Filters for graph-based keyword spotting in historical handwritten documents. Pattern Recognition Letters, 134:125–134, 2020.
- Местецкий Пронина. Штриховая сегментация рукописного текста по медиальному представлению. 2023.
- Местецкий Пацадия. Распознавание рукописных слов на основе штриховой сегментации. 2023.