

Классификация букв в изображениях на основе медиального представления

А.Л. Липкина¹, Л.М. Местецкий¹
lipkina96@mail.ru|mestlm@mail.ru

¹Московский государственный университет имени М.В. Ломоносова, Москва, Россия

В статье предлагается математическая модель понятия графемы для символов кириллических шрифтов и метод построения графем на основе непрерывного медиального представления букв в цифровых изображениях. Также предлагается метод распознавания изображения печатного текста на базе математической модели графемы, используемой при генерации признаков и для построения классификатора. Приводятся результаты экспериментов, подтверждающих работоспособность графемного подхода, высокое качество распознавания текста в различных вариантах шрифта и в различных качествах изображения текста.

Ключевые слова: цифровое изображение текста, цифровой шрифт, графема, медиальное представление, агрегированный скелетный граф

Letters Classification in Images Based on Medial Representation

A.L. Lipkina¹, L.M. Mestetskiy¹
lipkina96@mail.ru|mestlm@mail.ru

¹Lomonosov Moscow State University, Moscow, Russia

In the article we propose a mathematical model of grapheme for Cyrillic characters and a method of constructing graphemes based on the continuous medial representation of letters in digital images. We also suggest the recognition method of the printed text image on the basis of mathematical model of the grapheme used at generation of features and for classifier construction. The results of experiments confirming the efficiency of the grapheme approach, high quality of text recognition in different font variants and in different qualities of the text image are presented.

Keywords: digital text image, digital font, grapheme, medial representation, aggregated skeleton graph

1. Введение

Когда грамотный человек читает текст, он по виду символа может сразу определить, какую букву этот символ изображает. Причём может это сделать вне зависимости от различных вариантов художественного начертания символа (с засечками, наклонного, прямого, декоративного и т. д. [7]). То есть существует некий «образ» буквы, который легко узнаваем человеком и легко отличим от «образов» других букв. Такой «образ» называется *графемой* [11].

Определение 1.1. Буква — отдельный символ алфавита.

В процессе развития письменности и скорописи [8, 12] у букв появилось несколько начертаний: строчное и прописное, а позднее — различные варианты написания одной и той же буквы. Зачастую эти написания могут довольно сильно различаться, хотя и обозначают они произношение одного и того же звука, например: **А** и **а**. Для описания таких различий вводится понятие графемы:

Определение 1.2. Графема — единица письменности, некоторый графический примитив, имеющий вид геометрического графа и изображающий канонические начертания буквы.

Графему можно представлять себе как изображение буквы в тонком шрифте, например, как на рис. 1.



Рис. 1. Изображения букв в шрифте Lato и выделенные вершины геометрических графов

Графемы должны обладать следующими свойствами:

1. Любые две графемы хорошо отличимы между собой.
2. Пусть изображения I_1 и I_2 представляют одну и ту же графему. Тогда различие между I_1 и I_2 несущественное. Таким образом, похожесть определяется некоторой мерой сходства между I_1 и I_2 .



Рис. 2. Различные варианты графемы прописной буквы «А»

Понятие графемы вводится дизайнерами и шрифтовиками в словесной форме, через «общую» конструкцию начертаний графем. В настоящей статье предлагается математическое описание этой «общей» конструкции и проверяется гипотеза о том, что такого опи-

сания достаточно для распознавания букв в большинстве шрифтов.

2. Структура алгоритма классификации букв

Алгоритм фактически состоит из двух частей:

1. Построение математической модели графемы.
2. Разработка алгоритма на основе построенной модели.

Определим несколько базовых понятий.

Определение 2.1. *Фигура* — множество точек на плоскости.

Определение 2.2. *Пустой круг фигуры* — круг, целиком лежащий в фигуре.

Определение 2.3. *Вписанный пустой круг фигуры* — пустой круг фигуры, не содержащийся ни в каком другом пустом круге фигуры.

Определение 2.4. *Скелетное представление фигуры* — множество центров всех вписанных пустых кругов фигуры (см. рис. 3).

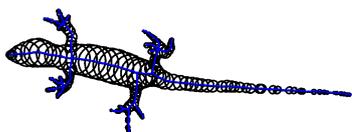


Рис. 3. Скелетное представление фигуры

В действительности скелетное представление фигуры представляет из себя граф \mathcal{S} , называемый *скелетом (скелетным графом)* фигуры. Вершинами графа являются центры вписанных пустых кругов, имеющих либо одну, либо три общих точки с границей фигуры, а ребрами — линии из центров вписанных пустых кругов, касающихся границы ровно в 2 точках. Более подробно скелетное представление фигуры рассматривается в [10].

Определение 2.5. *Силуэт скелетного графа* — фигура, состоящая из объединения всех вписанных пустых кругов, центры которых лежат в скелетном графе \mathcal{S} . Обозначение: $\mathcal{V}_{\mathcal{S}}$.

Определение 2.6. *Стрижка скелетного графа (с параметром α)* — процесс регуляризации скелетного графа \mathcal{S} , основанный на удалении несущественных ребер из скелетного графа (см. рис. 4). В процессе такого удаления возникает минимальный подграф \mathcal{S}' исходного скелетного графа, для которого выполняется $H(\mathcal{V}_{\mathcal{S}}, \mathcal{V}_{\mathcal{S}'}) \leq \alpha$, где $H(\mathcal{V}_{\mathcal{S}}, \mathcal{V}_{\mathcal{S}'})$ — хаусдорфово расстояние [2] между силуэтом скелетного графа \mathcal{S} и силуэтом скелетного графа \mathcal{S}' .



Рис. 4. Пример скелета без стрижки (слева) и со стрижкой (справа)

3. Построение математической модели графемы

Для построения математической модели графемы предлагается сделать два шага:

1. Сегментация изображения текста на изображения отдельных символов (графем).
2. Выделение структурного описания (математической модели) изображения каждой графемы.

Второй шаг разделяется на следующие действия:

1. Получение скелетного графа из изображения буквы.
2. Агрегирование скелетного графа.
3. Обработка скелетного графа, а именно удаление шумовых ребер.

Получение скелетного графа

Построение скелетного графа фигуры подробно описывается в [10]. После его построения производится последующая стрижка с параметром α .

Агрегирование скелетного графа

Полученный скелетный граф содержит только следующие типы вершин: вершины степени 1 (листья), вершины степени 2, вершины степени 3 (развилки).

Основную информацию о скелетном графе несут листья и развилки, а также виды связей между ними. Чтобы выделить эти связи, производится операция *агрегирования скелетного графа*: «склеивание» в одну цепь всех таких последовательных ребер, инцидентные вершины которых имеют степень либо 1, либо 2. После такой «склейки» в качестве вершин остаются только листья и развилки (см. рис. 5).

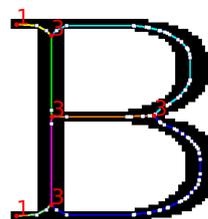


Рис. 5. Пример агрегированного скелетного графа. Белым цветом отмечены вершины степени 2 исходного графа.

После агрегирования скелетный граф представляет собой гиперграф $\mathcal{S}_{agg,1}$, вершинами которого являются листья и развилки, а ребрами — выделенные в процессе цепочки.

Обозначения и понятия

1. Рассматривается входное бинарное изображение символа. \mathcal{B} — минимальная по площади прямоугольная рамка с горизонтальными и вертикальными сторонами, ограничивающая данный символ. \mathcal{B}_H и \mathcal{B}_W — высота и ширина рамки \mathcal{B} соответственно.
2. Пусть e — ребро неагрегированного скелетного графа \mathcal{S} . $v_1(e), v_2(e)$ — концевые вершины этого ребра без учёта какого-либо порядка.
3. $l(e)$ — длина ребра e . Она вычисляется через евклидово расстояние между двумя точками $v_1(e)$ и $v_2(e)$:

$$l(e) = \sqrt{(v_1(e)_x - v_2(e)_x)^2 + (v_1(e)_y - v_2(e)_y)^2}.$$
4. Для ребра (цепочки) e_{agg} гиперграфа \mathcal{S}_{agg} через $v_1(e), v_2(e)$ обозначаются концевые вершины этой цепочки.
5. Ребро e_{agg} обозначенного гиперграфа \mathcal{S}_{agg} состоит из n последовательных рёбер исходного графа \mathcal{S} , попавших в данную цепочку e_{agg} : $\{e_{agg}^1, e_{agg}^2, \dots, e_{agg}^n\}$.
6. $l(e_{agg})$ — длина цепочки e_{agg} . Она считается как сумма длин всех входящих в данную цепочку рёбер:

$$l(e_{agg}) = \sum_{i=1}^n l(e_{agg}^i).$$

7. $\deg v$ — степень вершины v .

Определение 3.1. Пусть дан отрезок $d = [v_1(e_{agg}), v_2(e_{agg})]$. Среди всех вершин цепочки e_{agg} находится вершина v_h , наиболее удалённая от отрезка d . По трём точкам $v_1(e_{agg}), v_2(e_{agg}), v_h$ строится окружность. Тогда *аппроксимирующей дугой* называется дуга наименьшей длины, ограниченная точками $v_1(e_{agg}), v_2(e_{agg})$ (см. рис. 6).

Определение 3.2. Угол кривизны цепочки — центральный угол её аппроксимирующей дуги.

Замечание. В случае, когда три точки $v_1(e_{agg}), v_2(e_{agg}), v_h$ лежат на одной прямой или когда в цепочке e_{agg} нет вершин, угол кривизны цепочки полагается равным 0.

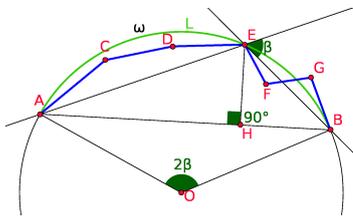


Рис. 6. Пример аппроксимирующей цепочку $[A, C, D, E, F, G, B]$ дуги L и центрального угла BOA

Удаление шумовых рёбер

После регуляризации и агрегации скелетного графа и в $\mathcal{S}_{agg,1}$ все ещё могут содержаться шумовые рёбра. Это проявляется на буквах, изображённых в шрифтах с засечками [7].

Засечки это некое украшение для буквы, и их наличие или отсутствие не мешает человеку распознать, какая именно буква изображена. Таким образом в модель графемы буквы засечки не должны входить. Поэтому следующий этап построения математической модели графемы — удаление из $\mathcal{S}_{agg,1}$ рёбер, являющихся засечками (см. рис. 7). Множество рёбер гиперграфа $\mathcal{S}_{agg,1}$, являющихся засечками, обозначим через \mathcal{E}_S .

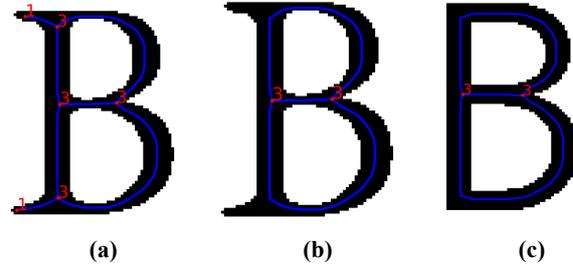


Рис. 7. 7а: скелет буквы в шрифте с засечками; 7б: тот же скелет с удалёнными рёбрами из \mathcal{E}_S ; 7с: скелет буквы в шрифте без засечек.

Для множества \mathcal{E}_S можно выделить следующие признаки:

1. $|\mathcal{E}_S| \geq 2$, то есть если засечки в скелетном графе присутствуют, то их не меньше двух.
2. $\forall e_{agg} \in \mathcal{E}_S$ характерны следующие признаки:
 - ровно одна из вершин $\{v_1(e_{agg}), v_2(e_{agg})\}$ является листовой, и ровно одна из них является развилкой;
 - длина ребра $l(e_{agg})$ не превышает некоторого порога $\mathcal{L}(\mathcal{B})$;
 - центральный угол 2β аппроксимирующей e_{agg} дуги не меньше некоторого порога \mathcal{A} .

Алгоритм удаления шумовых рёбер из $\mathcal{S}_{agg,1}$:

1. Определение множества \mathcal{E}_S на основе его признаков.
2. Удаление из агрегированного скелетного графа $\mathcal{S}_{agg,1}$ всех рёбер из \mathcal{E}_S .

Гиперграф, полученный после удаления рёбер и повторной агрегации, обозначим через $\mathcal{S}_{agg,2}$. Он и является предлагаемой математической моделью графемы.

4. Распознавание графем

На этом этапе из $\mathcal{S}_{agg,2}$ будут выделяться признаки для последующего построения классификатора графем.

Генерация признаков

В данном методе предлагается выделять 2 типа признаковов описаний: *признаки верхнего уровня* \mathcal{F}_a и *признаки нижнего уровня* \mathcal{F}_d . Они обладают следующими свойствами:

- Если из гиперграфов $\mathcal{S}'_{agg,2}, \mathcal{S}''_{agg,2}$ выделяются одинаковые признаки верхнего уровня $\mathcal{F}'_a = \mathcal{F}''_a$,

то признаки нижнего уровня \mathcal{F}'_d и \mathcal{F}''_d лежат в одном признаковом пространстве.

- Если из гиперграфов $\mathcal{S}'_{agg,2}, \mathcal{S}''_{agg,2}$ выделяются различные признаки верхнего уровня $\mathcal{F}'_a \neq \mathcal{F}''_a$, то признаки нижнего уровня \mathcal{F}'_d и \mathcal{F}''_d лежат в разных признаковых пространствах.

Признаки верхнего уровня

Рамка \mathcal{B} , в которую заключена графема, разбивается на n равных частей горизонтальными прямыми и на m равных частей — вертикальными. В каждом из получившихся $n \cdot m$ прямоугольников считается количество листьев и количество развилок, и эти количества добавляются в признаковое описание верхнего уровня. Помимо этого в качестве признака верхнего уровня рассматривается количество компонент связности графемы внутри рамки \mathcal{B} (см. рис. 8).

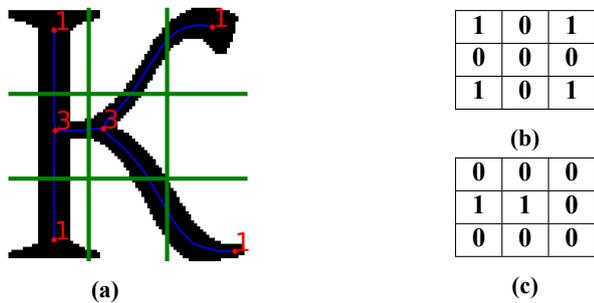


Рис. 8. 8a: скелет $\mathcal{S}_{agg,2}$ буквы «К» и разбиение рамки на 9 прямоугольников ($n = m = 3$); 8b: количества листьев в каждом из прямоугольников; 8c: количества развилок в каждом из прямоугольников.

Признаки нижнего уровня

Считаем, что признаковое описание верхнего уровня \mathcal{F}_a зафиксировано. Это означает, что фактически зафиксировано строение гиперграфа $\mathcal{S}_{agg,2}$: для каждого из $n \cdot m$ прямоугольников разбиения известно количество листьев и развилок, попавших в него, и известно количество рёбер гиперграфа, связанных с каждым прямоугольником разбиения. Таким образом, теперь появляется возможность генерировать фиксированное число признаков для каждого из $n \cdot m$ прямоугольников.

Признаковое описание нижнего уровня \mathcal{F}_d предлагается генерировать из структуры рёбер $\mathcal{S}_{agg,2}$.

Генерация признаков из ребра

Пусть $[A, B]$ — ребро гиперграфа $\mathcal{S}_{agg,2}$. Фиксируется маска разбиения этого ребра на k частей:

$$\mathcal{Z}_k = [z_1, z_2, \dots, z_k], \quad z_j \in (0, 1) \quad \forall j = \overline{1, k}.$$

Фиксируется стартовая вершина (не ограничивая общности считаем, что это A). Применяем разбиение \mathcal{Z}_k к ребру $[A, B]$, начиная от вершины A , следующим образом: ребро $[A, B]$ разбивается k точками, считая от точки A , на $k + 1$ отрезков s_i так, что:

$$\sum_{i=1}^j l(s_i) = z_j l([A, B]) \quad \forall j = \overline{1, k}.$$

Пусть концы отрезков s_i имеют координаты C_{i-1}, C_i :

$$s_i = [C_{i-1}, C_i] \quad \forall i = \overline{1, k+1}.$$

Заметим, что $C_0 = A$ и $C_{k+1} = B$. Также отдельно обозначим $\vec{b} = \overrightarrow{AC_1}$.

Пример разбиения можно увидеть на рис. 9.

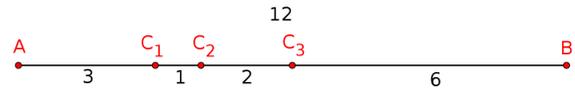


Рис. 9. Пример разбитого маской $\mathcal{Z}_3 = [\frac{1}{4}, \frac{1}{3}, \frac{1}{2}]$ ребра $[A, B]$ длины 12 со стартовой вершиной A

Предлагается выделять следующие признаки нижнего уровня:

1. Рассматриваются вектора $\overrightarrow{AC_i}$, $i = \overline{1, k+1}$. Пусть $m_i = \|\overrightarrow{AC_i}\|_2$, $i = \overline{1, k+1}$. Эти вектора нормируются на свои длины соответственно:

$$\vec{c}_i = \frac{\overrightarrow{AC_i}}{m_i}, \quad i = \overline{1, k+1}.$$

В качестве признаков берутся последовательно (по i) координаты получившихся векторов \vec{c}_i , $i = \overline{1, k+1}$.

2. Пусть $\vec{g} = (1, 0)$ (см. рис. 10). В качестве признаков добавляются последовательно (по i) следующие ориентированные углы:

$$\angle(\vec{g}, \overrightarrow{AC_i}), \quad i = \overline{1, k+1}.$$

3. Следующие ориентированные углы последовательно (по i):

$$\angle(\overrightarrow{C_i C_{i-1}}, \overrightarrow{C_i C_{i+1}}), \quad i = \overline{1, k}.$$

4. Отношения длин соседних векторов:

$$\frac{m_i}{m_{i-1}}, \quad i = \overline{2, k+1}.$$

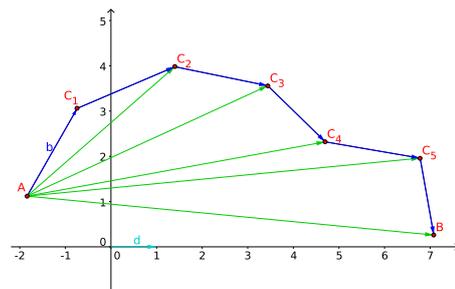


Рис. 10. Пример разбитого ребра $[A, B]$ со стартовой вершиной A

Генерация признаков для одного прямоугольника
 Пусть \mathcal{R} — текущая рассматриваемая прямоугольная область из разбиения рамки \mathcal{B} .

Сначала рассмотрим все листья, затем все развилки. Во всех случаях стартовой вершиной будет являться текущая рассматриваемая вершина v .

1. v — лист. Тогда генерируются признаки f_e для соответствующего ей ребра e , и они добавляются к итоговому признаковому описанию $f_{\mathcal{R}}$.
2. v — развилка. Рассмотрим соответствующие трём выходящим рёбрам вектора $\vec{b}_1, \vec{b}_2, \vec{b}_3$. Выходящие рёбра сортируются в порядке возрастания ориентированных углов $\angle(\vec{b}_i, \vec{g})$, $i = 1, 2, 3$, после чего для них генерируются признаки f_e . Полученные признаки добавляются в порядке сортировки рёбер к итоговому признаковому описанию $f_{\mathcal{R}}$.

Генерация признаков для графемы

Признаки нижнего уровня F_d для графемы получаются путём объединения признаков $f_{\mathcal{R}}$ в порядке упорядочивания прямоугольных областей \mathcal{R} .

Обучение классификатора

Теперь в рамках каждого признака верхнего уровня F_a имеется возможность обучить свой классификатор — каждый на своём признаковом пространстве, соответствующем своему признаковому пространству нижнего уровня F_d .

На этапе обучения берётся размеченная тренировочная выборка $(\mathbb{X}_{tr}, \mathbb{Y}_{tr})$, где $x \in \mathbb{X}_{tr}$ — бинаризованное изображение символа, $y \in \mathbb{Y}_{tr}$ — соответствующий класс изображения.

Алгоритм обучения состоит из следующих шагов:

1. По всей тренировочной выборке $(\mathbb{X}_{tr}, \mathbb{Y}_{tr})$ выделить признаки верхнего уровня и по ним построить словарь классификации \mathcal{D} .
2. Для каждого уникального признака верхнего уровня F_a выделить те объекты, которые имеют данный признак F_a . Для каждого из таких объектов построить признак нижнего уровня F_d . В итоге получается новая подвыборка объектов из признакового пространства F_d , при помощи которой обучается классификатор (см. рис 11).

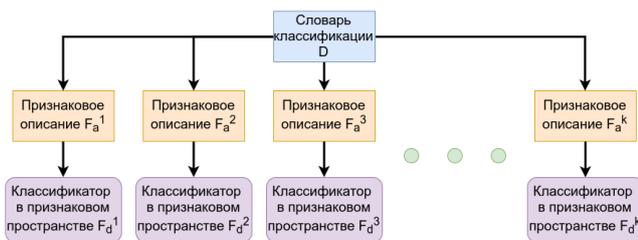


Рис. 11. Устройство словаря классификации \mathcal{D}

Алгоритм классификации

Пусть на вход приходит новый объект x (бинаризованное изображение одного символа), и его необходимо классифицировать. Для его классификации необходимо проделать следующие шаги:

1. Выделение математической модели графемы $\mathcal{S}_{agg,2}$ из x .
2. Построение признаков верхнего уровня F_a из $\mathcal{S}_{agg,2}$.
3. Проверка наличия признака F_a в словаре классификации \mathcal{D} , полученном на этапе обучения. Если признак не присутствует, то производится операция пост-обработки скелетного графа. Если присутствует — переход к следующему шагу.
4. Построение признаков нижнего уровня F_d , применение к нему соответствующего обученного классификатора и получение ответа.

Идея пост-обработки заключается в следующем: продолжение поиска подграфа, который возможно будет классифицировать по обученному словарю классификации \mathcal{D} . Если же такой граф не был найден, то будет возвращен отказ от классификации.

Итоговый алгоритм можно увидеть на рис. 12.

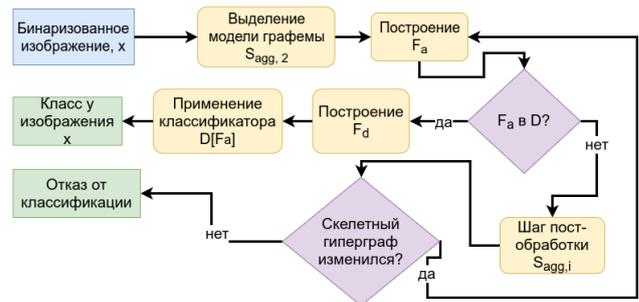


Рис. 12. Алгоритм классификации бинаризованного изображения

Метрика качества

В качестве метрики качества используется *точность классификации (accuracy)*.

5. Вычислительные эксперименты

Обучающая выборка

В качестве обучающей выборки было выбрано 88 различных шрифтов, из каждого сгенерировано по 33 буквы русского алфавита в строчном и прописном вариантах (то есть всего 66 графем) в трёх размерах шрифта: 30, 50, 100 пикселей. Генерация изображений производилась без сглаживания, то есть сразу в бинарном формате. Размер тренировочной выборки $(\mathbb{X}_{tr}, \mathbb{Y}_{tr})$ составляет $n_{tr} = 17424$ бинаризованных изображений букв. В качестве истинного класса \mathbb{Y}_{tr}^i для объекта \mathbb{X}_{tr}^i обучающей выборки бралась изображенная буква в нижнем регистре.

Параметры предложенного алгоритма

1. Параметр стрижки $\alpha = 0.06 \cdot \mathcal{B}_H$.

2. Порог для отсекаения по длине:

$$\mathcal{L}(\mathcal{B}) = \frac{2}{7} \max(\mathcal{B}_H, \mathcal{B}_W).$$

3. Порог для отсекаения по длине на стадии постобработки увеличивается в 1.8 раза:

$$\mathcal{L}'(\mathcal{B}) = 1.8 \cdot \mathcal{L}(\mathcal{B}) \quad (\kappa = 1.8).$$

4. Порог для отсекаения по центральному углу:

$$A = \frac{\pi}{5}.$$

5. На стадии извлечения признаков верхнего уровня полагается $n = m = 3$.

6. Фиксируемая сетка полагается равной:

$$Z_8 = \left[\frac{1}{50}, \frac{1}{5}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{4}{5} \right].$$

7. В качестве классификаторов нижних уровней рассматривается случайный лес (Random forest) [3].

Базовый алгоритм

В качестве базового алгоритма была выбрана сверточная нейронная сеть [6][1], архитектура которой представлена на рис. 13:

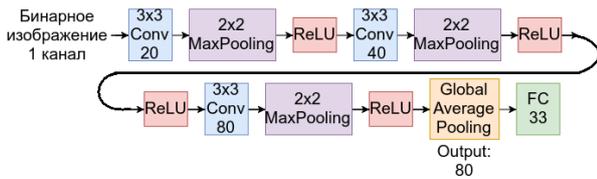


Рис. 13. Архитектура нейронной сети

Из соображений решения задачи классификации над выходным слоем (x^1, x^2, \dots, x^{33}) из 33 нейронов производится softmax-активация.

В качестве оптимизируемой функции потерь берётся кросс-энтропия.

Эксперимент 1

В качестве тестовой выборки берутся те же самые 88 шрифтов, которые использовались при обучении, но другого размера шрифта, который равен 80 пикселям. Таким образом, $n_{te} = 5800$. Изображения букв генерируются при помощи программы, то есть изображения качественные, без шумов и бинаризованные. Результаты работы двух методов (структурный анализ — описываемый в статье метод распознавания) представлены в таблице 1:

	Структурный анализ	Нейронная сеть
Качество, Q	0.99689	0.99862
Доля отказов от классификации	0.00086	0

Таблица 1. Результаты работы двух методов

Эксперимент 2

В качестве тестовой выборки берутся 50 шрифтов, которые не использовались при обучении [5]. Размер шрифта равен 80 пикселям, $n_{te} = 3300$. Изображения букв генерируются при помощи программы. Результаты работы двух методов представлены в таблице 2:

	Структурный анализ	Нейронная сеть
Качество, Q	0.97	0.96515
Доля отказов от классификации	0.01364	0

Таблица 2. Результаты работы двух методов

Эксперимент 3

В качестве тестовой выборки берутся те же самые 50 шрифтов, что и в предыдущем эксперименте 5, и того же размера. Сначала генерируется документ (.doc) со всеми буквами из тестовой выборки, затем этот документ конвертируется в изображение в формате png с разрешением 300 dpi (см.рис. 14). Сначала изображения из цветного представления RGB переводились в серые тона Y по формуле:

$$Y = 0.299R + 0.587G + 0.114B.$$

Затем изображения были бинаризованы при помощи метода Оцу [4].



Рис. 14. Пример буквы из входного изображения

Результаты работы двух методов представлены в таблице 3:

	Структурный анализ	Нейронная сеть
Качество, Q	0.94818	0.94454
Доля отказов от классификации	0.01485	0

Таблица 3. Результаты работы двух методов

Эксперимент 4

В этом эксперименте в качестве тестовой выборки берутся 18 выборочных шрифтов из 50 шрифтов эксперимента 5. Размер шрифтов полагается равным 80 пикселям, $n_{te} = 1188$. Генерируется документ (.doc) со всеми буквами из тестовой выборки, затем этот документ распечатывается на принтере. Далее полученные образцы сканируются с разрешением 300 dpi. То есть изображения получаются менее качественные, нежели чем в предыдущем случае (см.рис. 15).



Рис. 15. Пример буквы из входного изображения

Результаты работы двух методов представлены в таблице 4:

	Структурный анализ	Нейронная сеть
Качество, Q	0.95538	0.94696
Доля отказов от классификации	0.01263	0

Таблица 4. Результаты работы двух методов

Анализ экспериментов

Проведённые эксперименты показывают, что: по качеству предлагаемый метод не уступает выбранному базовому алгоритму и он имеет небольшую долю отказов от классификации, которая растёт с ухудшением качества изображения.

6. Выводы

В данной статье предложена формализация понятия «графема», а именно предложена математическая модель графемы. На основе этой модели предложен метод генерации признаков, использующихся для последующего построения алгоритма классификации изображений букв (то есть определена мера сходства между математическими моделями графем). По этому подходу к извлечению признаков был сделан доклад в [9]. Также в данной статье предложен алгоритм распознавания текста на изображениях.

Достоинства предлагаемого метода распознавания букв: независимость от размера, типа шрифта и типа начертания буквы; выделение общей структуры (математической модели графемы) букв, которой достаточно для распознавания букв в новых шрифтах; интерпретируемость признаков.

Недостатки метода: наличие отказов от классификации и зависимость качества распознавания от качества бинаризации изображения.

Проведённые эксперименты подтверждают, что предлагаемая математическая модель графемы показала свою работоспособность.

7. Благодарности

Работа выполнена при поддержке РФФИ, грант 17-01-00917.

8. Литература

[1] Bishop C. M. Pattern recognition and machine learning. — Springer. — 2006.

- [2] Felix Hausdorff. Grundzüge der mengenlehre. — 1914.
- [3] Kam, H. T. Random decision forests / Ho Tin Kam // Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC. — 1995.
- [4] Otsu, N. A threshold selection method from gray-level histograms / N. Otsu // IEEE Trans. Sys., Man., Cyber. — 1979.
- [5] Fonts database. — <https://www.fontsquirrel.com/>.
- [6] Gradient-based learning applied to document recognition / LeCun Yann, Léon Bottou, Yoshua Bengio, Patrick Haffner // Proceedings of the IEEE. — 1998.
- [7] ParaType. Цифровые шрифты. — ParaType. — 2008.
- [8] А. А. Зализняк. Русское именное словоизменение приложением избранных работ по современному русскому языку и общему языкознанию. — М.: Языки славянской культуры. — 2002.
- [9] А. Л. Липкина. Генерация структурных признаков для классификации букв в цифровых изображениях / А. Л. Липкина // Сборник тезисов XXV Международной научной конференции студентов, аспирантов и молодых ученых «ЛОМОНОСОВ-2018». — Макс-Пресс, 2018.
- [10] Л. М. Местецкий. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры. — М.: ФИЗМАТЛИТ. — 2009.
- [11] О. В. Осетрова. СЕМИОТИКА ШРИФТА. — Вестник Воронежского государственного университета. Серия: Филология. Журналистика. — 2006.
- [12] А. Соломоник. О языке и языках. — М.: Издательство «Спутник+». — 2017.

Об авторах

Липкина Анна Львовна, бакалавр кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Московского государственного университета. lipkina96@mail.ru.

Местецкий Леонид Моисеевич, д. т. н., профессор кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Московского государственного университета. mestlm@mail.ru.