

Модификация нейронной сети неокогнитрон для увеличения качества распознавания текста

А. Кугаевских

Институт математики и компьютерных наук

Тюменский государственный университет, Тюмень, Россия

a-kugaevskikh@yandex.ru

Аннотация

В работе представлен новый подход к распознаванию иероглифических текстов со слабо изученной лингвистикой. Классическая нейронная сеть была модифицирована для улучшения качества распознавания. Применение фильтра Габора позволило улучшить выявление границ символов, а совмещение сегментации и распознавания в одной нейросетевой архитектуре улучшить восстановление структуры текста.

Ключевые слова: распознавание текстов, нейронная сеть, фильтр Габора.

1. ВВЕДЕНИЕ

Изучение письменности исчезнувших культур порождает дополнительные трудности, связанные со слабой изученностью лингвистики языка и существованием в древних языках синтаксических элементов, связанных с группировкой иероглифов, а также наличием поврежденных иероглифов и других синтаксических элементов.

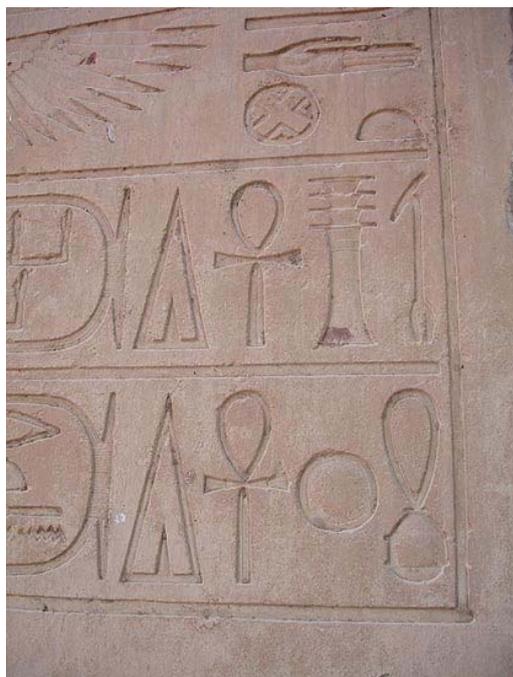


Рис 1. Исходное изображение

Так как лингвистика слабо изучена, мы не можем полагаться на лингвистический анализ при исправлении ошибок распознавания для увеличения качества работы.

В [1] приведены результаты поиска оптимального метода сегментации иероглифического текста и показано, что наиболее подходящий принцип – выявление границ на перепадах яркости. Там же приведена архитектура нейронной сети распознавания выявленных изображений символов.

К настоящему времени наиболее качественным методом для этого является фильтр Габора [3].

Существующие методы сегментации плохо приспособлены для подобного типа текстов ввиду высокой степени повреждения изображения, также возможно неправильное разделение на строки из-за разной высоты символов.

Нет гарантии верной сегментации из-за невозможности достоверно классифицировать компоненты изображения на предмет «шум / элемент изображения символа». А это возможно только после распознавания символа. Поэтому необходимо совместить процесс сегментации и распознавания в одной системе.

2. ОБЩИЙ АЛГОРИТМ

Распознавание иероглифического текста осуществляется в два этапа:

- 1) идентификация структурных элементов текста;
- 2) определение синтаксических параметров текста и его элементов (группы, высота текста и т.д.).

За первый этап отвечает модифицированная нейронная сеть неокогнитрон (рис. 2). Второй этап основывается на результатах первого, поэтому нейронная сеть не только должна указать класс, к которому относится изображение символа, но и определить некоторые параметры (угол поворота, степень повреждения, направление письма, высоту символа).

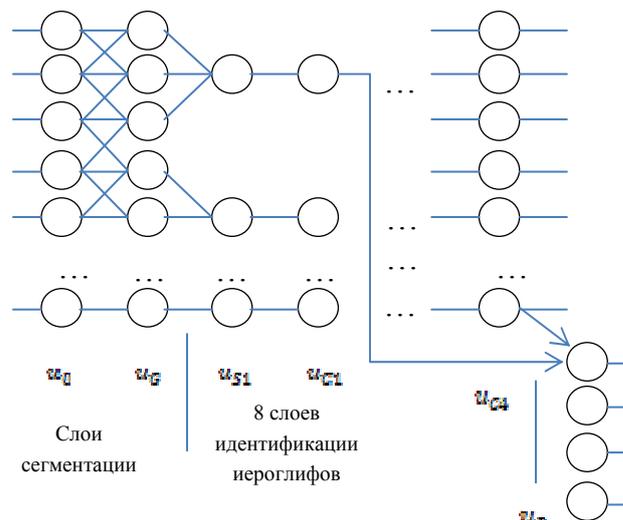


Рис. 2. Модифицированная структура нейросети

Для решения задачи сегментации на вход нейронной сети должно поступать изображение всего текста, а не отдельного символа.

Для улучшения качества выделения границ компонентов изображения в слое U_G стал применяться фильтр Габора, адаптированный под нейросетевую парадигму.

3. МОДИФИКАЦИЯ ИЗВЛЕЧЕНИЯ ГРАНИЦ И СЕГМЕНТАЦИЯ

Ядро фильтра Габора является произведением гауссиана и гармонической функции, фильтрует сигнал, основываясь на параметре предпочтительной пространственной частоты λ^{-1} :

$$g[k, l] = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(\frac{2\pi x'}{\lambda} + \psi\right), \quad (1)$$

Непосредственно же выявление границ сегментов осуществляется при проведении математической операции свертки ядра фильтра Габора и изображения в градациях серого:

$$r = \iint_{\Omega} f(u - k, v - l) g(k, l) dk dl, \quad (2)$$

Размер слоя U_G , отвечающего за выделение контуров сегментов, определяется размером исходного изображения. Значения свертки больше нуля означают наличие границы, поэтому в качестве передаточной функции нейрона используется функция Хевисайда. Преобразуя (2) к взвешенной сумме и подставив формулу (1), получим передаточную функцию нейрона:

$$u_G(m, n) = \varphi \left[\sum_{|k|, |l| < A_G} u_G(m + k, n + l) \exp\left(-\frac{x'^2 + 0.25y'^2}{2A_G^2}\right) \cos\left(\frac{2\pi x'}{\lambda} + \frac{\pi}{2}\right) \right], \quad (3)$$

где $\varphi[x] = \max(x, 0)$ – пороговая функция;

A_G – размер рецептивного поля;

m, n – позиция центра рецептивного поля;

k, l – позиция светового импульса в рецептивной области;

λ – количество параллельных возбуждающих и тормозных зон в рецептивном поле

Согласно исследованиям, опубликованным в [2], оптимальное значение предпочтительной пространственной частоты может быть вычислено, исходя из соотношения:

$$A_G = 0.56\lambda$$

$$x' = k \cos \theta + l \sin \theta$$

$$y' = -k \sin \theta + l \cos \theta$$

При этом для текущего пикселя рассматриваются яркости соседних пикселей в некоторой области A_G , которые влияют на степень граничности текущего пикселя. Таким образом, можно оценить степень граничности всех пикселей на изображении в зависимости от их соседей, и такая локальная обработка будет инвариантна к смене освещенности или цвета изображения.

Выход уровня U_G подается на слой извлечения признаков (U_{S1}). Он содержит ровно столько плоскостей, сколько неповторяющихся признаков встречается на всей обучающей выборке.

На рис. 3 представлен результат работы слоя извлечения границ

Сегментация осуществляется следующим образом. Каждый нейрон слоя извлечения признаков при активации запоминает координаты нейронов слоя извлечения границ, по активированным связям, соединенным с ним. При передаче от слоя к слою координаты отдельных признаков преобразуются в координаты и размер прямоугольника,

описанного около распознанного изображения символа. Таким образом, после распознавания символа, пройдя обратный путь по нейронной сети по активированным нейронам, можно выявить позицию сегмента на изображении. После распознавания всего текста, на основании этой информации, можно формировать строки текста.

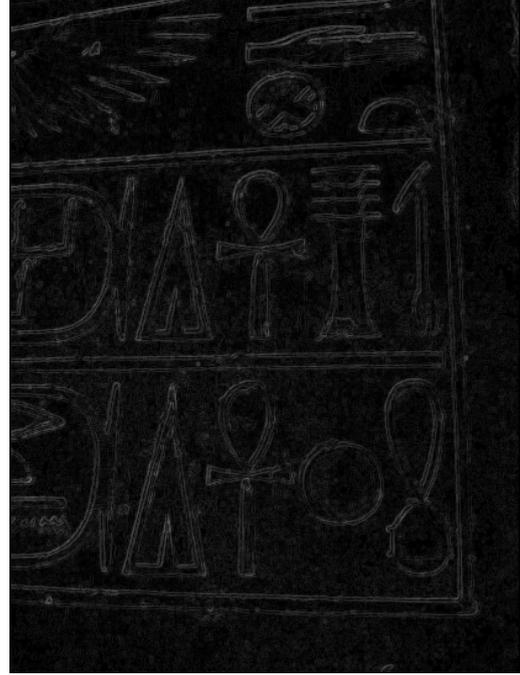


Рис. 3. Применение фильтра Габора

4. МОДИФИКАЦИЯ ДЛЯ ОПРЕДЕЛЕНИЯ ПАРАМЕТРОВ СИМВОЛА

Для придания сети инвариантности к поворотам символов был добавлен параметр α ($\alpha \in [0, \pi]$ с шагом 15°). В дальнейшем параметр угла наклона признака символа понадобится только при определении общего угла наклона символа при восстановлении структуры текста.

Шаг изменения параметра α выбран не случайно. Оптимально настроенный и обученный неоконгитрон перестает отвечать на признаки, отклоненные более чем на $15 - 20$ градусов.

Подход, предложенный Фукушимой, позволяет классифицировать изображение по признакам, но не дает ответа на вопрос о степени повреждения изображения элемента текста. Для выявления степени повреждения сегмента нейронная сеть дополняется слоем U_D , состоящим из 4 нейронов, отвечающих за соответствующую четверть сегмента.

Каждый нейрон слоя U_D получает связи от нейронов слоя U_{C1} . Ответ такого нейрона должен обладать обратной пропорциональной зависимостью от степени активации признаков символа (т.е. слоя U_{C1}). Чем меньше активация соответствующих нейронов слоя U_{C1} , тем выше вероятность, что символ поврежден в соответствующей четверти. Поэтому передаточной функцией является функция Хевисайда, при которой нейрон будет активироваться только при положительном сигнале. Ответом каждого нейрона U_D будет:

$$u_D(m) = \varphi \left[1 - \frac{\sum_{k=1}^{K_{C1}} \sum_{|l| < A_D} a_D(v) u_{C1}(n + v, k, \alpha)}{\theta_D} \right] \quad (4)$$

где A_D – размер области активации нейрона;

$\alpha_D(v)$ – сила постоянных возбуждающих связей;

θ_D – порог чувствительности.

Модификация нейросетевой модели К. Фукушимы в части дополнения слоев выявления степени повреждения требует обучения нейронов этого слоя.

Обучение нейрона этого слоя производится с применением правила Хебба, ставящего изменение весовых коэффициентов в зависимость от входных и выходных значений нейрона. Алгоритм обучения слоя выявления повреждения заключается в выполнении следующих шагов:

Шаг 1. Обучение слоев распознавания.

Шаг 2. Начальным значениям $\alpha_D(v)$ присваиваются соответствующие значения $\alpha_{S2}(v)$.

Шаг 3. На вход нейросети подается поврежденный образ элемента

Шаг 4. Расчет изменения веса

$$\Delta \alpha_D(v) = q_D u_D(m) u_{C1}(n + v, k, \alpha) \quad (5)$$

Шаг 5. Переход к следующему нейрону и возврат на шаг 4.

Шаг 6. Если не конец обучающей выборки – возврат на шаг 3, иначе – выход.

В результате обучения нейроны этого слоя становятся отзывчивыми к отсутствию наиболее важных признаков, характеризующих степень повреждения элемента текста. При этом признаки обладают разной степенью важности.

5. РЕЗУЛЬТАТЫ

На неповрежденных образцах изображения нейросеть показала среднее качество распознавания на уровне 82%. Этот показатель зависит от материала-носителя текста, степени повреждения.

Тестовая выборка для проведения экспериментов по качеству распознавания текстов составила 358 изображений текстов.

Тестовую выборку текстов можно условно разделить на четыре группы по материалу:

1. папирус, холст, кожа;
2. камень;
3. керамика;
4. дерево.

Материал из каждой группы обладает разной фактурой и степенью повреждения, что по-разному влияет на качество распознавания. Результаты проверки на текстах из каждой группы и сравнение показателей качества распознавания без применения и с применением фильтра Габора приведены в таблице.

Результаты проверки качества распознавания

Материал	Кол-во текстов	Качество распознавания, без применения фильтра Габора	Качество распознавания, с применением фильтра Габора
Папирус, холст, кожа	32	79,9%	82,1%

Камень	294	79,1%	84,6%
Керамика	7	78,4%	80,3%
Дерево	25	78%	81%
Итого	358	78,85%	82%

На рис. 4 приведена зависимость качества распознавания от степени повреждения образа символа. При увеличении степени повреждения до 50% и более наблюдается резкое снижение качества распознавания. Это объясняется большим количеством символов алфавита и их похожестью, так как при повреждении изображений символов может быть потеряна информация о важных признаках, отличающих символ от других.

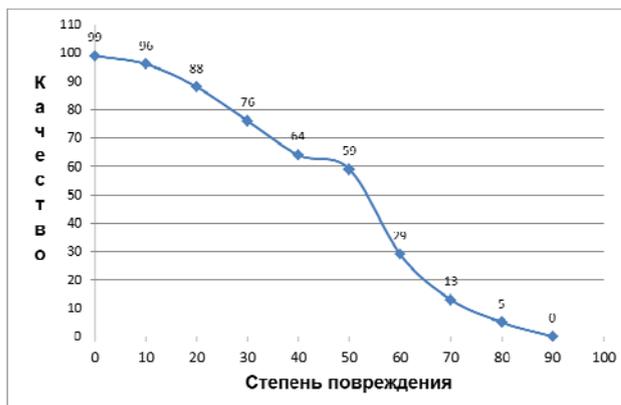


Рис. 4. Зависимость качества распознавания от степени повреждения

6. ЗАКЛЮЧЕНИЕ

Результаты проведенных экспериментов подтверждают, что применение фильтра Габора улучшает распознавание символов. А совмещение сегментации и распознавания в одной системе позволяет восстановить структуру текста. Среднее качество распознавания синтаксических элементов составило 82%, что подтверждает адекватность разработанных методов и моделей распознавания иероглифических текстов.

7. ССЫЛКИ

- [1] Кугаевских А.В., Ивашко А.Г. Возможность применения искусственных нейронных сетей "неокогнитрон" для распознавания древнеегипетских иероглифов // Вестник Тюменского государственного университета. – Тюмень: ТюмГУ, 2009 – № 6. – С. 209 - 214.
- [2] DeValois R.LAlbrecht., D.G. and Thorell L.G. Spatial frequency selectivity of cells in macaque visual cortex, Vis. Res., vol. 22, 1982. – P. 545 - 559.
- [3] Grigorescu S.E., Petkov N. and Kruizing P. Comparison of Texture Features Based on Gabor Filters, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 11, NO. 10, OCTOBER 2002. – 1160 - 1167

Об авторах

Кугаевских Александр – к.т.н., доцент кафедры ИСТюмГУ.
E-mail: a-kugaevskikh@yandex.ru