Метод определения надежности распознавания в задаче распознавания тисненых символов

В. В. Арлазаров, К. Б. Булатов, С. М. Карпенко

Аннотация. В данной статье описывается метод построения функции надежности распознавания изображений тисненых символов по полученным значениям вектора альтернатив классификатора (к примеру, нейронной сети). На основе этой функции строится правило отбраковки, эффективно классифицирующее результаты распознавания на надежные и не надежные. Оценки эффективности правила отбраковки проводится путем оценки суммарной стоимости ошибок отбраковки. В работе показано, что правило отбраковки, построенное описанным методом, является более эффективным, чем традиционные методы, основанные на оценке надежности распознавания по значениям максимальных альтернатив.
Ключевые слова: *распознавание символов, надежность распознавания, машинное обучение.*

Введение

Важной задачей, возникающей при автоматизации процессов жизнедеятельности, является оптическое распознавание образов, в частности — распознавание символов.

В практических задачах, включающих в себя распознавания символов, важную роль играет то, каким образом обрабатываются ошибки распознавания. В таких системах, как, например, ввод рукописных форм, автоматическая обработка банковских документов, пропущенные ошибки, занесенные в базу данных, обходятся очень дорого. Как следствие, при разработке таких систем возникает задача определения, насколько надежно распозналось то или иное информационное поле, или насколько надежно распознан тот или иной символ.

Целью данной работы является исследование методов решения задачи определения надежности результатов распознавания символов для систем, в которых классификатор основан на нейронных сетях.

Задача определения надежности результатов распознавания в рамках единой модели взаимодействия системы автоматической обработки документов была рассмотрена в работе [1]. В указанной работе автор описал схему построения функции оценки эффективности правила отбраковки, исходя из вычисленных апостериорных вероятностей ошибок классификатора и стоимостей этих ошибок, задаваемых извне. Также были рассмотрены простейшие правила отбраковки результата распознавания (правило первой альтернативы, правило двух альтернатив) и сформулированы общие принципы построения комплексных правил отбраковки.

В настоящей работе была предпринята попытка реализовать общий метод построения функции надежности и соответствующего ей решающего правила, основываясь на заранее определенных предикторах (признаках надежности), вычисляемых по вектору альтернатив классификатора [1], минимизировав при этом функционал стоимости.

Работа была проведена применительно к задаче определения надежности результатов распознавания изображений эмбоссированных номеров банковских или дисконтных пластиковых карт. Существует множество эффективных алгоритмов и подходов к решению задачи распознавания печатных символов на документах, но, к сожалению, эти алгоритмы нельзя непосредственно применять к таким конкретным задачам, как распознавание эмбоссированных символов. В качестве классификатора был использован трехслойный персептрон [2], а в качестве вектора признаков — структурный тензор второго порядка, построенный по изображению символа [3]. Отличительной особенностью данной системы является относительно невысокое качество распознавания (приблизительно 98,5 %).

1. Надежность распознавания

В таких системах, как, например, системы распознавания печатного или рукописного текста, полей документов или форм, и других составных объектов, как правило, существует единый модуль, отвечающий за распознавание одиночных символов. Если какое-либо подмножество символов, принадлежащих составному объекту, было распознано с ошибкой, то в рамках конкретной задачи существуют методы верификации распознанного поля (контекстные, словарные и т. п. методы). Но и сам модуль распознавания одиночных символов располагает информацией о том, насколько надежен (достоверен) полученный им результат распознавания.

Рассмотрим задачу распознавания одиночных символов. Имеется некоторое множество классов (алфавит распознавания):

$$A = \{C_i\}_{i=1}^{M}, \tag{1}$$

где M — размер алфавита распознавания.

Классификатор (к примеру, нейронная сеть) реализует классифицирующую функцию, которая ставит образу x в соответствие вектор альтернатив. Каждая альтернатива состоит из индекса класса и оценки принадлежности объекта распознавания к этому классу:

$$\begin{gathered} C(x) = \vec{a}(a_1, ..., a_M),\ a_k = (i_k, g_k), \\ i_k \in \{1, ..., M\},\ g_k \in [0, 1], \end{gathered} \tag{2}$$

где $x \in X$- образ символа, $\vec{a}$ — вектор альтернатив, i_k — индекс класса, g_k — оценка принадлежности символа к классу с индексом i_k. Альтернативы в векторе $\vec{a}$ упорядочены по убыванию оценки принадлежности g_k, т. е. $g_1 = \max_{k\in\{1,...,M\}} g_k$ и $g_1 \geqslant g_2 \geqslant ... \geqslant g_M$. Результатом распознавания является индекс класса с максимальной оценкой принадлежности, т. е. i_1.

Задача определения надежности результатов распознавания состоит в том, чтобы построить правило, позволяющее эффективно отбраковывать ошибочные результаты (т. е. классифицировать результаты на надежные и ненадежные). Приотбраковкерезультатамогутвозникнутьдватипаошибок:

1. **Ошибка первого рода —** объект распознан верно, но результат признан недостоверным.
2. **Ошибка второго рода —** объект распознан с ошибкой, но результат при этом признан достоверным.

Предположим, что имеется некоторый оператор, или пользователь, взаимодействующий с системой распознавания. Введем стоимость контроля оператором правильно распознанного объекта, но признанного недостоверным, как стоимость ошибки первого рода: $W_{rc} \in \mathrm{R}_+$. Стоимость ввода неверного результата, но признанного достоверным, в дальнейшую обработку (в зависимости от конкретной системы распознавания), введем как стоимость ошибки второго рода: $W_{ae} \in \mathrm{R}_+$. Для того, чтобы добиться максимальной эффективности правила отбраковки, необходимо минимизировать сумму оценок стоимости ошибок первого и второго рода:

$$P_{rc}W_{rc} + P_{ae}W_{ae} \to \min, \tag{3}$$

где P_{rc}, P_{ae} — апостериорные вероятности ошибок первого и второго рода соответственно.

Для построения комбинированного метода определения надежности распознавания необходимо определить, по каким признакам можно судить о том, насколько достоверен полученный результат, т. е. необходимо определить набор **признаков надежности** (или **предикторов**)

$$F = \langle f_1, f_2, ..., f_N \rangle, \quad f_l(\vec{a}) \in \mathrm{R}. \tag{4}$$

Предиктором может являться любая вещественнозначная функция от вектора альтернатив, по значению которой можно судить о достоверности результата распознавания. Самыми простыми признаками надежности являются значения оценок первой и второй альтернатив:

$$f_{fq}(\vec{a}) = g_1, \quad f_{sq}(\vec{a}) = g_2. \tag{5}$$

Чем больше значение оценки первой альтернативы, тем достовернее результат. Для второй альтернативы напротив — чем меньше значение ее оценки принадлежности, тем достовернее результат.

Еще одним рассматриваемым в этой работе признаком надежности является энтропийная функция оценок альтернатив. Пусть $G(\vec{a})$ — сумма значений оценок:

$$G(\vec{a}) = \sum_{k=1}^{M} g_k. \tag{6}$$

Тогда энтропийная функция оценок альтернатив выражается следующим образом:

$$f_{ent}(\vec{a}) = -\sum_{k=1}^{M} \frac{g_k}{G(\vec{a})} \ln \frac{g_k}{G(\vec{a})}. \tag{7}$$

Для обобщения будем считать, что $0 \ln 0 = 0$. В самом идеальном варианте, когда первая альтернатива имеет максимально возможную оценку принадлежности (единицу), а все остальные оценки альтернатив нулевые, энтропийная функция принимает нулевое значение. В самом «ненадежном» варианте, когда все оценки принадлежности альтернатив равны между собой, энтропийная функция альтернатив достигает своего максимального значения.

Исходными данными для задачи определения надежности распознавания является база результатов распознавания (обучающая база)B. Она представляет собой множество пар:

$$B = \{(i_c, \vec{a})\}, \tag{8}$$

где $i_c \in \{1, ..., M\}$ — истинный индекс класса, к которому принадлежит символ, $\vec{a}$ — вектор альтернатив.

Подмножество верных результатов:

$$B_c = \{(i_c, \vec{a})\,|(i_c, \vec{a}) \in B \wedge i_1 = i_c\}, \tag{9}$$

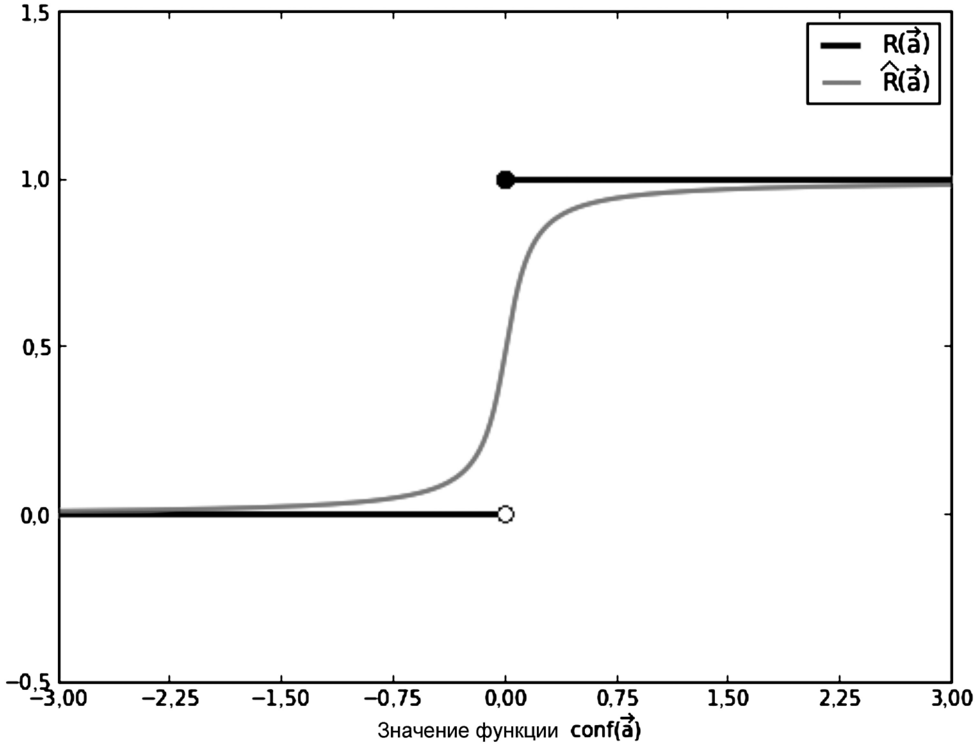

Рис. 1. Зависимость функций $R(\vec{a})$ и $\hat{R}(\vec{a})$ от значения функции $conf(\vec{a})$при$\omega = 8$

где i_1 — индекс класса первой альтернативы из вектора $\vec{a}$.

Соответственно, подмножество ошибочных результатов:

$$B_e = B \backslash B_c. \tag{10}$$

Определим $\nu : B \to \{0, 1\}$ — функцию проверки:

$$v(i_c, \vec{a}) = \begin{cases} 1, & \text{если } i_1 = i_c \\ 0, & \text{в противном случае} \end{cases} \tag{11}$$

где i_c — истинный индекс класса, i_1 — индекс класса первой альтернативы.

Определим искомую функцию надежности в виде:

$$conf(\vec{a}) = \beta + \sum_{l=1}^{N} c_l f_l(\vec{a}), \quad \beta \in \{-1, 0, 1\}, \quad c_l \in \mathrm{R} \tag{12}$$

где N — количество предикторов, $f_i(\vec{a})$ – i-й предиктор.

Определим также функцию (правило) отбраковки (рис. 1):

$$R(\vec{a}) = \begin{cases} 1, & \text{если } conf(\vec{a}) \geqslant 0, \\ 0, & \text{если } conf(\vec{a}) < 0. \end{cases} \tag{13}$$

Задача состоит в том, чтобы найти такие коэффициенты функции надежности $c_l, \quad l = 1, ..., N$, а также свободный член $\beta \in \{-1, 0, 1\}$, при которых минимизируется функционал стоимости:

$$\begin{aligned} &W(B, \beta, c_1, ..., c_N) = \\ &= \sum_{(i_c, \vec{a}) \in B} \big[W_{rc} \cdot \nu(i_c, \vec{a}) \cdot (1 - R(\vec{a})) + \\ &+ W_{ae} \cdot (1 - \nu(i_c, \vec{a})) \cdot R(\vec{a})\big] \to \min_{\substack{c_{l1} \leqslant c_l \leqslant c_{l2} \\ \beta \in \{-1, 0, 1\}}}. \end{aligned} \tag{14}$$

где c_{l1}, c_{l2} — заданные границы изменения l-го коэффициента функции надежности.

2. Минимизация функционала стоимости

Для минимизации функционала стоимости (14) использовался метод адаптивного ячеистого прямого поиска (MeshAdaptiveDirectSearch, MADS [4]), реализованный в программном пакете NOMAD [5]. Данный метод хорошо себя зарекомендовал на практике в качестве численного метода оптимизации трудно вычислимых функций и функций типа «черный ящик».

Начальное приближение $c_{10}, ..., c_{N0}$ и границы изменения коэффициентов функции надежности $c_{11}, c_{12}, ..., c_{N1}, c_{N2}$задаются в качестве исходных

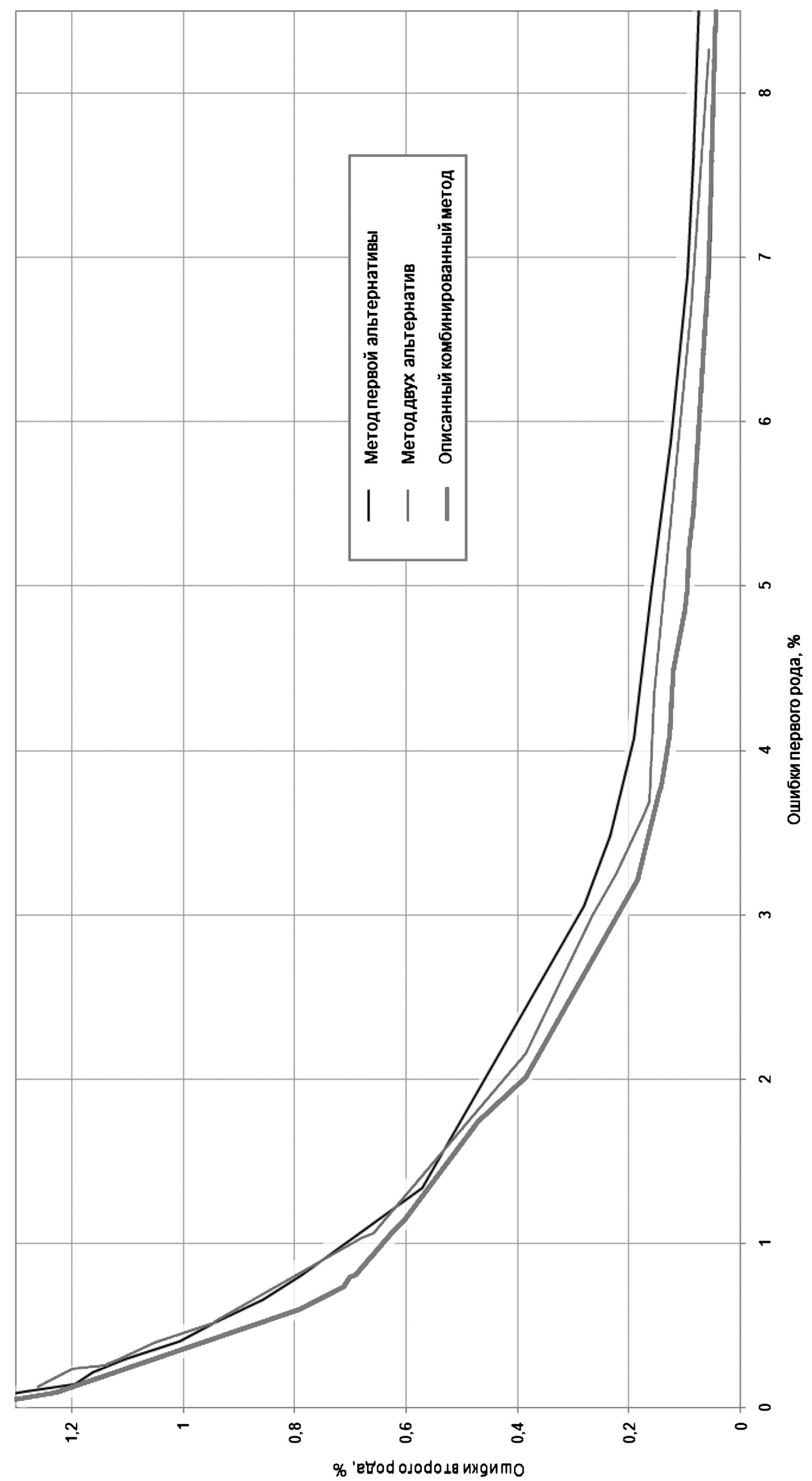

Рис. 2. ER-кривые рассмотренных методов определения надежности

данных. Также предусмотрена возможность задать параметры $r_c \in Z^0_+$, $r_d \in R_+$ - количество дополнительных случайных начальных приближений и отклонение. Если эти параметры заданы, тогда дополнительно генерируется r_c начальных приближений в виде:

$$\hat{c}_{l0} = c_{l0} + u(r_d), \tag{15}$$

где $u(r_d)$ — случайное значение в интервале $(-r_d, r_d)$ с равномерным законом распределения.

В любой точке пространства аргументов $\beta, c_1, ..., c_N$ существуют направления, вдоль которых знак функции надежности (12) не меняется. Как следствие, не меняется значение функции (13), а значит и функционала стоимости (14). Это создает трудности при численном решении задачи. Для того чтобы этого избежать, аппроксимируем функцию (13) следующей функцией:

$$R(\vec{a}) \approx \hat{R}(\vec{a}) = \frac{1}{\pi}\left(arctg\,(\omega \cdot conf(\vec{a})) + \frac{\pi}{2}\right), \tag{16}$$

где ω — заранее заданный коэффициент жесткости. Очевидно, что $\lim\limits_{\omega\to\infty} \hat{R}(\vec{a}) = R(\vec{a}), conf(\vec{a}) \neq 0$ при фиксированных коэффициентах функции надежности.

Теперь вместо исходного функционала стоимости (14) будем минимизировать аппроксимированный функционал стоимости:

$$\begin{aligned} W(B,\beta,c_1,...,c_N) &\approx \hat{W}(B,\beta,c_1,...,c_N) = \\ &= \sum_{(i_c,\vec{a})\in B} \left[W_{rc}\cdot \nu(i_c,\vec{a})\cdot(1-\hat{R}(\vec{a}))+\right. \\ &\left. W_{ae}\cdot(1-\nu(i_c,\vec{a}))\cdot \hat{R}(\vec{a})\right] \to \min_{\substack{c_{l1}\leqslant c_l \leqslant c_{l2} \\ \beta\in\{-1,0,1\}}} . \end{aligned} \tag{17}$$

3. Результаты эксперимента

Исходная экспериментальная обучающая база содержала результаты распознавания символов номеров пластиковых карт при помощи трехслойного персептрона. В качестве вектора признаков для персептрона использовался структурный тензор, построенный по входному изображению [3].

В рамках работы проведено вычисление коэффициентов функции надежности результатов распознавания эмбоссированных символов номера пластиковых карт по экспериментальной базе. Также было проведено сравнение эффективности правила отбраковки, основанного на построенной функции надежности (12), и методов первой альтернативы и двух альтернатив, описанных в работе [1].

На рис. 2 представлены ER-кривые, построенные для трех рассматриваемых методов. Кривая ER отражает зависимость между количеством ошибок первого и второго рода при использовании некоторого правила отбраковки. Точки, принадлежащие ER-кривой, получаются изменением контрольного параметра этого правила отбраковки. Контрольным параметром для метода первой альтернативы является пороговое значение $T(\theta)$, контрольными параметрами для метода двух альтернатив являются пороговые значения для первой и второй альтернативы $T(\theta)$ и $\hat{T}(\theta)$. Контрольным значением для комбинированного метода на основе функции надежности является соотношение между стоимостями ошибок второго и первого рода W_{ae}/W_{rc}.

На рис. 2 показано, что ER-кривая для комбинированного метода, основывающегося на построенной функции надежности, лежит ниже ER-кривых для традиционных методов первой альтернативы и двух альтернатив. Особенно важно заметить, что с помощью этого метода удалось достигнуть более низкого уровня ошибок второго рода в диапазоне ошибок первого рода 4–6 %, который является допустимым рабочим диапазоном для рассматриваемой системы распознавания символов. Из этого следует, что правило отбраковки результатов распознавания на основе построенной функции надежности можно применять на практике в качестве компонента системы распознавания.

Заключение

В работе показан метод построения линейной функции надежности распознавания по полученным значениям альтернатив классификатора для системы распознавания тисненых символов.

Предложенный метод построения функции надежности результатов распознавания был протестирован на системе распознавания эмбоссированных символов номера банковских и дисконтных карт при помощи нейронной сети. Проведен сравнительный анализ эффективности методов, описанных в работе [1] и предложенного в данной работе метода. Показано, что построенная функция надежности дает на практике лучшие результаты, чем традиционно используемые метод первой альтернативы и метод двух альтернатив.

Метод имеет важное значение для построения правил отбраковки результатов распознавания одиночных символов. На основе построенных правил отбраковки можно в дальнейшем строить правила отбраковки результатов распознавания полей, а также построения критериев отказа.

Таблица 1

Характеристики экспериментальной обучающей базы

Характеристика	Обозначение	Значение	Примечания
Количество классов	M	10	Классы соответствуют символам '0', '1', ..., '9'
Размер базы	$\|B\|$	28225	
Количество верно распознанных символов	$\|B_c\|$	27814	98.54 %
Количество символов, распознанных с ошибкой	$\|B_e\|$	411	1.46 %

Литература

1. *Арлазаров В. В.* Структурирование визуальных представлений информационной среды и методы определения надежности распознавания. Дисс. ... канд. техн. наук. М.: Московский государственный институт стали и сплавов (Технологический университет), 2004. 120 с.
2. *Мерков А. Б.* Распознавание образов. Введение в методы статистического обучения. М.: URSS, 2011. 256 с.
3. *Knutsson H.* Representing local structure using tensors // Proceedings 6th Scandinavian Conf. on Image Analysis. Oulu: OuluUniversity, 1989. с. 244–251.
4. *Audet Ch.*, *Dennis J. E.*, *Jr.* Mesh Adaptive Direct Search Algorithms for Constrained Optimization // SIAM Journal on Optimization, 2006. Vol. 17. Issue 1. P. 188–217.
5. *Le Digabel S.* Algorithm 909: NOMAD: Nonlinear Optimization with the MADS algorithm // ACM Transactions on Mathematical Software. 2011. Vol. 37(4). P. 44:1–44:15.

Арлазаров Владимир Викторович. С. н. с. ИСА РАН. К. т. н. Окончил в 1999 г. МИСиС. Кол-во печатных работ: 6. Область научных интересов: распознавание образов, обработка изображений, системы массового обслуживания. E-mail: vva777@gmail.com

Булатов Константин Булатович. Аспирант НИТУ МИСиС. Окончил НИТУ МИСиС в 2013 г. Область научных интересов: распознавание образом, машинное обучение, информационные системы. E-mail: hpbuko@gmail.com

Карпенко Семен Михайлович. Научный сотрудник ИППИ РАН. Окончил в 2002 г. МГУ им. М. В. Ломоносова. Количество печатных работ: 15. Область научных интересов: статистическое машинное обучение, обработка изображений. E-mail: simon@iitp.ru