

УДК 004.934.2

Е.Ю. Костюченко, Р.В. Мещеряков, А.Ю. Крайнов

Критерии информативности при обработке биометрических сигналов при помощи нейронных сетей

При решении задач идентификации пользователей по биометрическим характеристикам возникает проблема выбора параметров биометрических характеристик с целью их дальнейшей обработки. Проблеме построения критериев информативности и оценке применимости предлагаемого критерия посвящена настоящая статья.

Ключевые слова: идентификация биометрических сигналов, параметры сигнала, критерий информативности параметров, нейронные сети.

При обработке биометрических сигналов при помощи нейронных сетей возникает проблема выбора параметров для подачи на входы нейронной сети. На этапе выбора возникает проблема ранжировки параметров, при этом имеется необходимость учета результатов обработки биометрического сигнала нейронной сетью. Без учета этих результатов предсказать точность работы нейронной сети даже на бинарном уровне (пригоден параметр к использованию или нет) не представляется возможным.

Можно выделить следующие существующие критерии информативности [1]:

– EER (Equal error rate) – уровень равной ошибки. Уровень, при котором в системе частота появления ошибки первого рода равна частоте появления ошибки второго рода [2].

– CER (Classification error rate) – уровень ошибки классификации. Представляет собой общую частоту возникновения ошибок независимо от рода.

– MTER (Minimum total error rate) – уровень минимальной суммарной ошибки. Уровень, при котором сумма частот появления ошибок первого и второго рода минимальна.

– Критерий на основе коэффициента корреляции между полученным и эталонным критерием [3].

– Критерий Байеса – критерий минимального среднего риска [4].

На основе анализа перечисленных критериев можно выделить следующие недостатки существующих критериев:

– Зависимость значения критерия от встречаемости участков первого и второго рода.

– Не учитывается важность меньших ошибок по сравнению с большими. На практике возникает необходимость учета нелинейной зависимости критерия от ошибки.

– Нет учета приоритета ошибок первого или второго рода, на практике часто возникает необходимость задавать более важную ошибку при допустимости существенного роста ошибки другого рода.

– Частоты появления ошибок первого и второго рода считаются на разных непересекающихся участках, частота появления которых, как правило, не совпадает между собой. Как следствие, физический смысл критерия не выражен.

На основе анализа приведенных критериев можно предложить новый критерий, учитывающий выявленные недостатки:

$$M(m_1, m_2) = \begin{cases} M'(m_1, m_2), & M'(m_1, m_2) < 1, \\ 1, & M'(m_1, m_2) \geq 1, \end{cases} \quad (1)$$

где $k \in [1, \infty)$ – коэффициент, учитывающий взаимную значимость ошибок первого и второго рода.

$$M'(m_1, m_2) = ((m_1)^s + (k \times m_2)^s)^{1/s} \quad (2)$$

для случая, когда более важной является ошибка второго рода, или

$$M'(m_1, m_2) = ((k \times m_1)^s + (m_2)^s)^{1/s} \quad (3)$$

для случая, когда более важной является ошибка первого рода; $s \in (0, \infty)$ – коэффициент, учитывающий значимость больших ошибок по сравнению с меньшими.

Предложенный критерий учитывает выявленные недостатки посредством введения коэффициентов k и s . При определенных значениях k и s данный критерий преобразуется к уже существующим. Кроме того, соответствие существующим критериям может быть выявлено путем сопоставления результатов анализа с применением существующих кри-

териев и предложенного. Например, рассмотрим сопоставление предложенного критерия и корреляционного критерия применительно к анализу речевого сигнала.

В качестве исследуемых параметров речевого сигнала были выбраны статистические параметры спектра исследуемого речевого сигнала:

- mosp – среднее значение интенсивности на отсчете;
- dispr – дисперсия интенсивности на отсчете;
- asim – коэффициент асимметрии интенсивности на отсчете;
- asim1 – отношение медианы интенсивности к среднему значению на отсчете;
- cvar – вариация интенсивности на отсчете;
- lmosp – логарифм среднего значения интенсивности на отсчете;
- maxch – номер канала с максимальной интенсивностью на отсчете;
- medi – медиана интенсивности на отсчете;
- grand – случайная величина, равномерно распределенная в диапазоне $[-0,5; 0,5]$;

Рассмотрим оценки информативности параметров, получаемые для 10 различных нейронных сетей с применением предложенного критерия (табл. 1) и корреляционного критерия (табл. 2).

Таблица 1

Значения предложенного критерия информативности, $k = 1, s = 2$

№ эксп.	1	2	3	4	5	6	7	8	9	10
maxch	0,29372	0,29492	0,28918	0,29123	0,28434	0,29406	0,29277	0,28048	0,28548	0,2903
medi	0,41293	0,40849	0,39591	0,41441	0,40852	0,4027	0,40119	0,41756	0,40062	0,39704
asim	0,531	0,56406	0,51646	0,52031	0,48096	0,50831	0,4784	0,52363	0,5126	0,52309
asim1	0,34759	0,35506	0,3749	0,33655	0,36441	0,38043	0,3675	0,35304	0,36515	0,35221
mosp	0,42948	0,42656	0,43947	0,4348	0,42335	0,43777	0,43591	0,43547	0,43383	0,44218
lmosp	0,6041	0,64279	0,45521	0,48392	0,48701	0,49658	0,51851	0,46933	0,55425	0,48885
randpar	0,6899	0,70253	0,70626	0,69282	0,70088	0,6566	0,68737	0,69636	0,71356	0,70752
dispr	0,37245	0,40278	0,39703	0,3968	0,39665	0,38344	0,3806	0,39369	0,38247	0,40656
cvar	0,5289	0,47102	0,57454	0,4861	0,44846	0,62775	0,45169	0,57621	0,63072	0,47734

Таблица 2

Значения корреляционного критерия информативности

№ эксп.	1	2	3	4	5	6	7	8	9	10
maxch	0,66198	0,66064	0,66748	0,66768	0,66024	0,66361	0,6717	0,66339	0,66417	0,66116
medi	0,58739	0,58952	0,60125	0,58854	0,5903	0,59615	0,59738	0,58187	0,59867	0,60511
asim	0,40331	0,36248	0,45477	0,44342	0,46723	0,4062	0,4751	0,39872	0,44174	0,4272
asim1	0,59272	0,58044	0,56427	0,61794	0,56692	0,57709	0,58365	0,59582	0,58593	0,58803
mosp	0,59757	0,60729	0,58044	0,58519	0,60742	0,58569	0,58792	0,57953	0,59395	0,57544
lmosp	0,556	0,49618	0,57912	0,55262	0,56914	0,53231	0,52822	0,58718	0,54749	0,55643
randpar	0,26523	0,23701	0,2486	0,23361	0,25733	0,24518	0,27976	0,24943	0,21029	0,26193
dispr	0,64544	0,60974	0,62383	0,62062	0,62466	0,63659	0,64849	0,62916	0,64084	0,60625
cvar	0,43236	0,50747	0,41887	0,44517	0,54301	0,33076	0,4824	0,43959	0,31653	0,49758

Для оценки связи между полученными значениями воспользуемся коэффициентом ранговой корреляции Спирмена [5]:

$$R = 1 - \frac{6 \sum_{i=1}^n (kr_{1i} - kr_{2i})^2}{n(n^2 - 1)}, \quad (4)$$

где R – значение коэффициента корреляции Спирмена; n – количество значений, по которым ведется расчет коэффициента корреляции; kr_{1i} , kr_{2i} – значения рангов для первого и второго критериев для каждого из параметров.

Применив формулу (4) к исходным данным, выполнив ранжировку, получим оценку коэффициента корреляции $R = -0,9036$.

Отрицательная корреляция говорит об противоположном характере критериев: согласно предложенному критерию параметр тем информативнее, чем его значение меньше, для корреляционного критерия верно обратное утверждение – параметр тем информативнее, чем его значение больше.

Проведем оценку значимости коэффициента корреляции. Для этого воспользуемся критерием Стьюдента [5]. Оценим эмпирическое значение критерия по формуле (5):

$$t_3 = |R| \sqrt{\frac{n-2}{1-R^2}}, \quad (5)$$

где R – значение коэффициента корреляции Спирмена; n – количество значений, по которым ведется расчет коэффициента корреляции; t_3 – эмпирическое значение критерия Стьюдента.

Подставив имеющиеся значения, получим

$$t_3 = 0,9036 \sqrt{\frac{90-2}{1-0,9036^2}} \approx 19,8.$$

Из таблиц найдем критическое значение коэффициента Стьюдента для уровня значимости 0,001 $t_{кр}=3,4$.

Так как $t_3 \gg t_{кр}$, то между критериями существует сильная линейная корреляция, что подтверждает достоверность предложенного подхода к определению критерия информативности.

Вышеприведенные выкладки показывают соответствие предложенного критерия существующим, с другой стороны, выявленные недостатки существующих критериев могут быть скомпенсированы за счет подбора коэффициентов критерия k и s .

Литература

1. Костюченко Е.Ю. Обработка естественной информации на основе аппарата нейронных сетей // Докл. Том. гос. ун-та систем управления и радиоэлектроники. – Томск: Изд-во ТУСУР, 2009. – № 1(19), ч. 2. – С. 54–56.
2. Нгуен М.Т. Разработка алгоритмов построения оценок достоверности для систем распознавания речи: автореф. дис. ... канд. тех. наук. – М., 2008. – 22 с.
3. Спортивная метрология / Под ред. В.М. Зациорского. – М.: ФиС, 1982. – 256 с.
4. Берсуцкий Я.Г. Принятие решений в управлении экономическими объектами: методы и модели: учеб. пособие / Я.Г. Берсуцкий, Н.Н. Лепа, А.Я. Берсуцкий. – Донецк: ИЭП, 2002. – 276 с.
5. Крамер Г. Математические методы статистики: пер. с англ. – 2 изд. – М., 1975. – 648 с.

Костюченко Евгений Юрьевич

Ассистент каф. комплексной информационной безопасности электронно-вычислительных систем ТУСУРа
Тел.: (382-2) 41-34-26
Эл. адрес: key@keva.tusur.ru

Мещеряков Роман Валерьевич

Доцент каф. комплексной информационной безопасности электронно-вычислительных систем ТУСУРа
Тел.: (382-2) 41-34-26
Эл. адрес: mrv@security.tomsk.ru

Крайнов Алексей Юрьевич

Доктор физ.-мат. наук, профессор Томского государственного университета
Тел.: (382-2) 52-98-45
Эл. адрес: akrainov@ftf.tsu.ru

E.Y. Kostyuchenko, R.V. Mescheriakov, A.U. Krainov

Information criteria for processing of biometric signals with neural networks

The problem of parameters biometric characteristics selection for the purpose of their further working appears with solution problems of the biometric users identification. To the problem of the informativeness criteria construction and to the proposed criterion applicability estimation is dedicated present article.

Keywords: the biometric signals identification, the signal parameters, the parameters informativeness criterion, the neural networks.