

УДК: 519.2, 612.087, 621.319.7

Иванова Н.А., Серикова Ю.И.

Синтез предсказателя, оценивающего ошибку вычисления математического ожидания для выборки из 32 примеров биометрического образа

В настоящее время в России и в Казахстане активно создаются нейросетевые преобразователи биометрия-код [1, 2]. Основой технологии является возможность автоматического обучения больших искусственных нейронных сетей [3]. Стандартизированный алгоритм обучения [3], построен на учете математических ожиданий биометрических параметров и их стандартных отклонений.

В силу того, что обучающая выборка примеров биометрического образа всегда мала, необходимо, при вычислении математического ожидания и стандартного отклонения, параллельно оценивать интервал возможных значений их ошибок. Покажем, что эта задача вполне корректна на примере выборки из 32 опытов.

Для синтеза простейшего предсказателя ошибки вычислений рассмотрим ситуацию, когда мы вычисляем единственное математическое ожидание по выборке из 32 опытов - $E_{32}(x)$, а затем вычисляем 31 математическое ожидание по выборке из 31 опыта (в каждой выборке удаляется один из опытов). Формально мы можем описать эту ситуацию следующим образом:

$$E_{31,i}(x) = (E_{32}(x) - \frac{x_i}{32}) \cdot \frac{32}{31} \quad (1).$$

Далее мы можем найти левую и правую границы действительного положения математического ожидания:

$$L = \min(E_{31,i}(x)) \quad (2),$$

$$R = \max(E_{31,i}(x)) \quad (3).$$

Очевидно, что синтез предсказателя ошибок вычисления математического ожидания малых выборок (3) сводится к написанию соответствующей программы. После программной реализации предсказателя может быть осуществлено тестирование достоверности предсказаний на выборке из 1000 опытов. Проведение подобного численного эксперимента показало, что программный предсказатель, реализующий вычисления по формулам (1), (2), (3), верно угадывает интервал возможных ошибок с вероятностью 0.697 и ошибается с вероятностью 0.307.

Вероятность ошибок может быть снижена, если удвоить амплитуду стандартного отклонения, заменив соотношение (1) на следующее:

$$E_{31,i}(x) = (E_{32}(x) - \frac{2x_i}{32}) \cdot \frac{32}{31} \quad (4).$$

Программный предсказатель интервала ошибок, реализующий формулы (4), (2), (3) ошибается с вероятностью 0.166.

Если утроить влияние отбрасываемых компонент:

$$E_{31,i}(x) = (E_{32}(x) - \frac{3x_i}{32}) \cdot \frac{32}{31} \quad (5),$$

то предсказатель ошибается с вероятностью 0.072.

Учетверение влияния отбрасываемых компонент:

$$E_{31,i}(x) = (E_{32}(x) - \frac{4x_i}{32}) \cdot \frac{32}{31} \quad (6),$$

приводит к снижению вероятности ошибок предсказания до 0.001.

В качестве примера работы предсказателя ниже даны три таблицы результатов, полученные для выборки из 32 опытов при реализации предсказателя по формулам (6), (2), (3).

№ =	0	1	2	3	4	5	6	7	8	9
L =	-0.164	-0.164	-0.302	-0.663	-0.199	-0.02	-0.16	-0.506	-0.036	-0.406

№ =	0	1	2	3	4	5	6	7	8	9
R =	0.605	0.605	0.623	0.092	0.477	0.566	0.484	0.216	0.581	0.263

№ =	0	1	2	3	4	5	6	7	8	9
(R - L) =	0.769	0.769	0.925	0.754	0.677	0.586	0.644	0.722	0.617	0.67

Рис. 1. Таблицы предсказания интервала возможного положения математического ожидания между левой границей – L и правой границей – R.

В нижней части рисунка 1 дана ширина интервалов возможного положения вычисляемого математического ожидания. Все интервалы имеют разную длину, то есть у нас появляется возможность сравнивать между собой качество выборок. Чем уже интервал, тем выше качество представленной для обучения выборки. Как видно из данных таблицы в нижней части рисунка 1 наихудшей является выборка № 2, а наилучшей является выборка №5, дающая самый узкий интервал возможного положения математического ожидания.

ЛИТЕРАТУРА:

1. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. //Ю.К.Язов (редактор и автор), соавторы В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника, 2012 г. 157 с.
2. Ахметов Б.С., Иванов А.И., Фунтиков В.А., Безяев А.В., Малыгина Е.А. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа. Монография, Казахстан, г. Алматы, ТОО «Издательство LEM», 2014 г. -144 с., находится в открытом доступе (<http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf>).
3. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».

Статья поступила 15.05.2016, опубликована 25.05.2016 по положительной рецензии д.т.н. Малыгина А.Ю.