

Ю. Н. МАТВЕЕВ

ОЦЕНКА ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ОБЩЕГО РЕШЕНИЯ АНСАМБЛЯ КЛАССИФИКАТОРОВ

Предложен алгоритм оценки доверительного интервала общего решения ансамбля классификаторов, выходом каждого из которых является логарифмическое отношение правдоподобия.

Ключевые слова: доверительный интервал, общее решение, ансамбль классификаторов, идентификация дикторов, голосовая биометрическая система.

Введение. При построении голосовых биометрических систем для повышения надежности идентификации личности по голосу (идентификации диктора) часто используется набор автоматических, полуавтоматических и экспертных методов исследования фонограмм, основанных на признаках речи различной природы. В качестве надежного итогового решения используется обобщающее, построенное на базе решений каждого из перечисленных методов идентификации.

По результатам исследования фонограмм для каждого метода выдается мера доказательности, которая строится на основе оценки степени тождества/различия дикторов. В качестве меры доказательности принят логарифм отношения правдоподобия (*LR*-оценка, Likelihood Ratio) каждого метода:

$$\log(LR) = \log \left(\frac{P(X|H_0)}{P(X|H_1)} \right), \quad (1)$$

где $P(X|H_0)$ — вероятность получения данных исследования X при истинности гипотезы H_0 , $P(X|H_1)$ — вероятность получения данных исследования X при истинности гипотезы H_1 .

Для автоматических методов идентификации оценка значений $P(X|H_0)$, $P(X|H_1)$ производится без участия эксперта. Для получения зависимости вероятностей $P(X|H_0)$, $P(X|H_1)$ от расстояния между векторами признаков сравниваемых сигналов на этапе разработки используются большие речевые базы данных, на которых устанавливается статистическая связь вероятностей с расстоянием. Тем не менее представительность этих баз данных ограничена условиями, в которых собирались фонограммы. В силу этого для условий, отличных от тех, на которые алгоритм был точно настроен, реальные значения вероятности истинности того или иного идентификационного решения становятся меньше. Наиболее важными факторами, влияющими на надежность решения автоматической системы, являются различия в свойствах канала записи эталонной (образцовой) и исследуемой (спорной) фонограмм, различное фи-

зиологическое и эмоциональное состояние диктора, несопоставимый речевой материал или условия внешней среды.

Для учета влияния этих факторов следует выявить их наличие и ввести к LR -оценке в формуле (1) степенную поправку, например, в форме степенного показателя Q :

$$LR = \left(\frac{P(X | H_0)}{P(X | H_1)} \right)^Q.$$

Эта степенная поправка учитывает сопоставимость исследуемых фонограмм с теми факторами, на которые ориентирован выбранный метод идентификации. Значение $Q=1$ соответствует тому, что качество фонограмм полностью удовлетворяет заявленным требованиям. Значение Q , близкое к нулю, соответствует тому, что качество фонограмм значительно ниже того, при котором метод сохраняет работоспособность; LR -оценка стремится к единице, а значит, невозможно принять решение по идентификации диктора.

Формула получения итогового обобщающего решения, т.е. итоговой оценки, наиболее часто реализуется в виде логарифмического отношения правдоподобия:

$$LLR = \log(LR) = \sum_i \log(LR_i), \quad (2)$$

где $LR_i = FRR_i / FAR_i$ — оценка отношения правдоподобия (LR -оценка) i -го метода идентификации, FRR_i и FAR_i — оценки вероятностей ошибок первого и второго рода i -го метода идентификации соответственно.

Согласно руководству [1], результат измерения является только аппроксимацией или оценкой значения измеряемой величины и, таким образом, будет полным, только когда дополняется установлением неопределенности этой оценки. В соответствии с этим ставится задача установления неопределенности итоговой LLR -оценки.

На практике существует много возможных источников неопределенности, например, неполное определение и несовершенная реализация определения измеряемой величины; нерепрезентативная выборка измерений; неполное представление о влиянии условий окружающей среды на измерения или несовершенное измерение параметров окружающей среды; недостоверные значения констант и других параметров, полученных из внешних источников и используемых в алгоритме обработки данных; аппроксимации и предположения, используемые в методе измерения и измерительной процедуре и т.д. [2, раздел 3.3.2]. Практически все перечисленные источники неопределенности присутствуют при идентификации дикторов.

Согласно руководству [1], неопределенность — параметр, связанный с результатом измерения (оценкой), характеризующий дисперсию значений, которые могли быть обоснованно приписаны измеряемой величине. Параметром может быть, например, стандартное отклонение или полуширина интервала, имеющего установленный доверительный уровень. Стандартные отклонения оценивают из предполагаемых распределений вероятностей, основанных на опыте или другой информации.

Доверительное оценивание. В настоящее время в системах распознавания (верификации и идентификации) диктора все чаще применяется концепция доверительности [3—7]. В этом случае дополнительно определяется доверительная вероятность (доверительный интервал), указывающая на надежность полученного результата (оценки).

В соответствии с ГОСТ доверительный интервал с заданной вероятностью накрывает неизвестное значение оцениваемого параметра распределения. Границы доверительного интервала называют доверительными границами. Оцениванием с помощью доверительного интервала называют способ оценки, при котором с заданной доверительной вероятностью устанавливают границы доверительного интервала.

Несмотря на то что доверительные интервалы обычно применяют при оценке одного числового параметра, для многих двухпараметрических и трехпараметрических распределений (в задачах распознавания дикторов — нормальных, бинормальных, гамма-распределений) обычно используют точечные оценки и построенные на их основе доверительные границы для каждого из параметров отдельно.

Исходя из формулы (2) алгоритм оценки доверительного интервала итоговой *LLR*-оценки состоит из следующих шагов.

1) Оценка доверительных интервалов значений FAR_i и FRR_i для отдельных методов идентификации дикторов.

2) Оценка доверительных интервалов LR_i -оценок отдельных методов идентификации дикторов.

3) Оценка доверительного интервала итоговой *LLR*-оценки.

Основные подходы к оценке доверительных интервалов FAR и FRR. В литературе рассматривается несколько подходов к оценке значений FAR и FRR . В общем случае эти подходы делятся на параметрические и непараметрические методы оценки.

Непараметрические методы, наиболее популярными из которых являются методы бутстрепа, „блочного“ бутстрепа и т.д. [3, 8], требуют проведения множества тестов на различных выборках, что не подходит для экспертных и полуавтоматических методов идентификации дикторов из-за большой трудоемкости.

В параметрических методах используется предположение о виде распределения значений FAR и FRR при определении доверительных интервалов. Наиболее часто в работах по распознаванию дикторов делаются следующие предположения [9]:

- о бинормальном распределении,
- о нормальном распределении,
- о биномиальном распределении.

Допустим, что определены некоторое выборочное распределение D и некоторый доверительный порог δ . В предположении о бинормальном распределении генерирование доверительных интервалов и их границ производится с использованием доверительных границ Хотеллинга [3].

В предположении о нормальном распределении генерирование доверительных интервалов и границ производится путем расчета среднего μ и стандартного отклонения σ распределения D . Затем ищется статистическая константа z двусторонней доверительной границы δ для распределения размерности $|D|$, что дает доверительный интервал $\mu \pm z\sigma$.

В предположении о биномиальном распределении дисперсия рассчитывается как $V = \mu(1-\mu)$, что дает доверительный интервал $\mu \pm z\sqrt{V/|D|}$.

На практике наиболее часто используется предположение о биномиальном распределении. Решение, приведенное в работе [10], основывается на параметрической оценке доверительных интервалов значений FRR и FAR по методу, описанному в работе [11].

Оценка доверительных интервалов FAR и FRR оценок отдельных методов идентификации дикторов. Введем следующие обозначения: P_k — априорная вероятность появления целевой личности (клиента), $P_3 = 1 - P_k$ — априорная вероятность появления злоумышленника (импостера).

Значение этих параметров зависит от типа приложения. Например, в случае идентификации диктора при радиопередаче предполагается $P_k < 1$, в то время как в системах контроля доступа предполагается $P_3 \ll 1$.

Определим доверительные интервалы для значений FAR и FRR в предположении о биномиальном характере распределения этих ошибок.

При большом числе попыток N распределение биномиальной случайной величины будет близко к нормальному. Учитывая, что дисперсия биномиальной случайной величины

равна $Np(1-p)$, получаем для ее математического ожидания Np приближенные доверительные границы для значения p (см. раздел 2.2.4 работы [12]):

$$\bar{p} - z\sqrt{\frac{\bar{p}(1-\bar{p})}{N}} \leq p \leq \bar{p} + z\sqrt{\frac{\bar{p}(1-\bar{p})}{N}}, \quad (3)$$

где \bar{p} — выборочная оценка p , а z — квантиль нормального распределения, равная хвосту кривой распределения $\alpha/2$ (например, $z = 1,96$ для $\alpha = 0,05$). Используя формулу (3), можно вычислить доверительный интервал частоты FAR (при $\bar{p} = FAR$ и $N=N_3$) и FRR (при $\bar{p} = FRR$ и $N=N_k$).

Оценка доверительных интервалов LR -оценок отдельных методов идентификации дикторов. В соответствии с формулой (2) следующим этапом является оценка доверительных интервалов LR_i -оценок отдельных методов идентификации дикторов, а точнее логарифма этих значений — LLR_i .

Оценка доверительного интервала значений LLR_i определяется по методу распространения ошибок [13]. Для случая функциональной зависимости вида $f = \frac{x}{y}$

$\left(LR_i = \frac{FRR_i}{FAR_i} \right)$ оценка доверительного интервала f вычисляется по следующей формуле:

$$\frac{\delta_f}{f} = \sqrt{\left(\frac{\delta_x}{x}\right)^2 + \left(\frac{\delta_y}{y}\right)^2}. \quad (4)$$

Однако, поскольку при получении оценки LLR_i используется операция логарифмирования, то, как отмечено в работе [13], для этого должен использоваться другой подход к вычислению итоговой оценки, без вычисления выражения (4).

Вычисление доверительного интервала итоговой LLR -оценки. Для вычисления итоговой LLR оценки будем использовать абсолютные значения изменения величин FAR_i и FRR_i , т.е. ΔFAR_i и ΔFRR_i , соответственно. Предполагается, что величины FAR_i и FRR_i изменяются в диапазоне $(FRR_i - \Delta FRR_i; FRR_i + \Delta FRR_i)$ и $(FAR_i - \Delta FAR_i; FAR_i + \Delta FAR_i)$ соответственно.

Задача состоит в определении доверительного интервала величины LLR ($LLR - \Delta_d LLR$; $LLR + \Delta_u LLR$), где $\Delta_d LLR$ и $\Delta_u LLR$ — абсолютное значение изменения величины LLR в сторону уменьшения и увеличения соответственно.

Воспользуемся формулой из работы [13]:

$$\Delta LLR = \sqrt{\sum_i (\Delta LLR_{FRR_i}^2 + \Delta LLR_{FAR_i}^2)},$$

где ΔLLR_{FRR_i} и ΔLLR_{FAR_i} — значение изменения LLR при изменении FRR_i и FAR_i , если остальные величины остаются неизменными.

Значение LLR определяет принадлежность тестового произнесения искомому диктору. При $LLR > 0$ тестовое произнесение принадлежит искомому диктору, $LLR < 0$ — нецелевому (злоумышленнику). Модуль значения LLR соответствует степени уверенности (большее значение по модулю соответствует большей уверенности).

Следовательно, при $LLR > 0$ наибольший интерес представляет величина $\Delta_d LLR$, которую можно найти по формуле:

$$\Delta_L LLR = \sqrt{\sum_i (\Delta_L LLR_{FRR_i}^2 + \Delta_L LLR_{FAR_i}^2)},$$

где

$$\begin{aligned}\Delta_d LLR_{FRR_i} &= \sum_j w_j \ln \left(\frac{FRR_j}{FAR_j} \right) - \sum_{j \neq i} w_j \ln \left(\frac{FRR_j}{FAR_j} \right) - w_i \ln \left(\frac{FRR_i - \Delta FRR_i}{FAR_i} \right) = \\ &= w_i \ln \left(\frac{FRR_i}{FAR_i} \right) - w_i \ln \left(\frac{FRR_i - \Delta FRR_i}{FAR_i} \right) = w_i \ln \left(\frac{FRR_i}{FRR_i - \Delta FRR_i} \right), \\ \Delta_d LLR_{FAR_i} &= w_i \ln \left(\frac{FRR_i}{FAR_i} \right) - w_i \ln \left(\frac{FRR_i}{FAR_i - \Delta FAR_i} \right) = w_i \ln \left(\frac{FAR_i - \Delta FAR_i}{FAR_i} \right), \\ &w_i = D_i.\end{aligned}$$

Аналогично при $LLR < 0$ наибольший интерес представляет величина $\Delta_u LLR$, которую можно найти по формуле:

$$\Delta_u LLR = \sqrt{\sum_i (\Delta_u LLR_{FRR_i}^2 + \Delta_u LLR_{FAR_i}^2)},$$

где

$$\begin{aligned}\Delta_d LLR_{FRR_i} &= w_i \ln \left(\frac{FRR_i + \Delta FAR_i}{FAR_i} \right) - w_i \ln \left(\frac{FRR_i}{FAR_i} \right) = w_i \ln \left(\frac{FRR_i + \Delta FRR_i}{FRR_i} \right), \\ \Delta_d LLR_{FAR_i} &= w_i \ln \left(\frac{FRR_i}{FAR_i - \Delta FAR_i} \right) - w_i \ln \left(\frac{FRR_i}{FAR_i} \right) = w_i \ln \left(\frac{FAR_i}{FAR_i - \Delta FAR_i} \right), \\ &w_i = D_i.\end{aligned}$$

Заключение. В статье описан алгоритм оценки доверительного интервала для общего решения ансамбля из нескольких классификаторов (методов идентификации дикторов): автоматических, полуавтоматических и экспертных методов исследования фонограмм, основанных на признаках речи различной природы.

Описанный алгоритм оценки общего доверительного интервала основан на определении доверительных интервалов ошибок первого и второго рода (FAR и FRR) различных методов идентификации, составляющих ансамбль, в предположении о биномиальном характере распределения этих ошибок, а также оценке методом распространения ошибок доверительного интервала общего решения ансамбля по доверительным интервалам ошибок первого и второго рода (ΔFAR_i и ΔFRR_i) каждого из методов ансамбля.

СПИСОК ЛИТЕРАТУРЫ

1. Руководство по выражению неопределенности измерения / Пер. с англ., под науч. ред. проф. В. А. Слава. ВНИИМ им. Д. И. Менделеева, 1999.
2. Походун А. И. Экспериментальные методы исследований. Погрешности и неопределенности измерений: Учеб. пособие. СПб: СПбГУ ИТМО, 2006. 112 с.
3. Vogt R., Sridharan S., Mason M. Making confident speaker verification decisions with minimal speech // Proc. of Interspeech. Brisbane, Australia, 2008. P. 1405—1408.
4. Campbell W., Reynolds D., Campbell J., Brady K. Estimating and evaluating confidence for forensic speaker recognition // Proc. of ICASSP. Philadelphia, PA, USA, 2005. Vol. 1. P. 717—720.
5. Huggins J. G. M. Confidence metrics for speaker identification // Proc. of ICSLP. Denver, Colorado, USA, 2002. P. 1381—1384.
6. Richiardi J., Prodanov P., Drygajlo A. Speaker verification with confidence and reliability measures // Proc. of ICASSP. Toulouse, France, 2006. Vol. 1. P. 641—644.
7. Richiardi J., Drygajlo A., Prodanov P. Confidence and reliability measures in speaker verification // J. of the Franklin Institute. 2006. Vol. 343, N 6. P. 574—595.

8. Koval S., Lokhanova A. Confidence Bounds Curves as a Tool for Evaluation of Automatic Speaker Recognition Results Uncertainty // Proc. 14th Intern. Conf. on Speech and Computer. SPECOM 2011. Kazan, 2011. P. 284—289.
9. Wu J. C., Martin A. F., Kacker R. N. Measures, Uncertainties, and Significance Test in Operational ROC Analysis // J. of Research of the National Institute of Standards and Technology. 2011. Vol. 116, N 1. P. 517—537.
10. Biosecure Tool. Performance evaluation of a biometric verification system, version 1.0. France, Aurelien Mayoue: GET-INT. 2007.
11. Bolle R. M., Ratha N. K., Pankanti S. Error analysis of pattern recognition systems — the subsets bootstrap // Computer Vision and Image Understanding. 2004. Vol. 93, N 1. P. 1—33.
12. Мятлев В. Д., Панченко Л. А., Терехин А. Т. Основы математической статистики. М.: МАКС Пресс, 2002.
13. Lab Reference Manual (LR09): Propagation of Uncertainty [Электронный ресурс]: <<http://www.physics.pomona.edu/sixideas/labs/LRM/LR09.pdf>>.

Сведения об авторе

Юрий Николаевич Матвеев — д-р техн. наук; ООО „ЦРТ-инновации“, Санкт-Петербург; главный научный сотрудник; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра речевых информационных систем; профессор;
E-mail: matveev@mail.ifmo.ru

Рекомендована кафедрой
речевых информационных систем

Поступила в редакцию
22.10.12 г.