РАСПОЗНАВАНИЕ ЧЕЛОВЕКА ПО ГОЛОСУ С ИСПОЛЬЗОВАНИЕМ ИСКУССТЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Акунович А.А.

Цурко А.В. – ассистент кафедры ИРТ

Задача распознавания человека по голосу была поставлена несколько десятилетий тому назад, с тех пор она не потеряла своей актуальности. Задача распознавания по голосу может найти применение в так областях как криминалистика, разведка, защита несанкционированного доступа, в банковской сфере и финансах и т.д.

В работах по распознаванию диктора по голосу наибольшую популярность приобрёл метод кепстрального преобразования спектра. Схема этого метода следующая: на интервале времени в 10 – 20 мс вычисляется текущий спектр мощности, а затем применяется обратное преобразование Фурье от логарифма этого спектра (кепстр), и находятся коэффициенты этого кепстра:

$$c_n = \frac{1}{\Theta}\int_0^{\Theta} \log\left|S(j\omega,t)\right|^2 e^{-jn\Omega\omega} d\omega$$

Число кепстральных коэффициентов n зависит от требуемого сглаживания спектра, и находится в пределах от 20 до 40.

Если используется гребёнка полосовых фильтров, то коэффициенты дискретного кепстрального преобразования вычисляются как:

$$c_n = \sum_{m=1}^{M}[\log Y(m)]\cos\left[\frac{\pi m}{M}\left(m-\frac{1}{2}\right)\right]$$

где $Y(m)$ – выходной сигнал *m*-го фильтра, c_n – *n*-й коэффициент кепстра.

Свойства слуха учитываются путем нелинейного преобразования шкалы частот, обычно в шкале *мел*. Эта шкала формируется исходя из присутствия в слухе так называемых критических полос, таких, что сигналы любой частоты в пределах критической полосы неразличимы. Шкала *мел* вычисляется как

$$M(f) = 1125\ln(1+ f/700)$$

где *f* - частота в Гц, M – частота в мелах.

Коэффициенты кепстрального преобразования формируют пространство, в котором и производится распознавание диктора. Эти коэффициенты сокращенно обозначаются как MFCC – Mel Frequiency Cepstral Coefficients. Число используемых коэффициентов от 10 до 30. Часто используются первые и вторые разности по времени кепстральных коэффициентов, что втрое увеличивает размерность пространства принятия решений, но улучшает эффективность распознавания диктора .

Искусственная нейронная сеть (ИНС, нейронная сеть) – математическая модель, имитирующая работу биологической нейронной сети живых организмов. Нейронная сеть состоит из нейронов и связей (синапсов) между этими нейронами (рисунок 1).

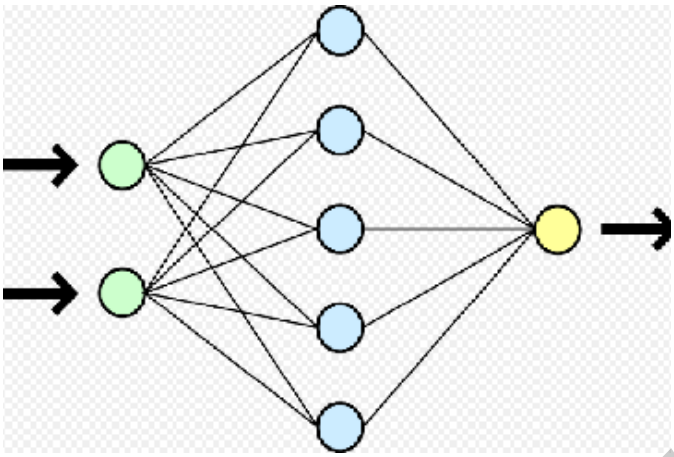

Рисунок 1 – Пример структуры нейронной сети

Каждый нейрон может хранить сигнал, а также передавать сигнал другим нейронам или принимать сигнал от других нейронов. Объединив нейроны в большие сети можно выполнять довольно сложные задачи, в частности – распознавание.

Одним из главных преимуществ нейронных сетей перед традиционными алгоритмами – это возможность обучения. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В процессе обучения коэффициенты связей подстраиваются таким образом, чтобы на выходе нейронной сети получались значения в зависимости от данных, поданных на её вход. Таким образом, в процессе обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат, на основании данных, которые отсутствовали в обучающей выборке, а также неполных или «зашумлённых», частично искажённых данных. В контексте задачи идентификации человека по голосу, успешно обученная сеть, получив на вход вектор признаков, выдаст на выходе вектор, поставленный в соответствие определённому человеку, которого нужно распознать. Таким образом, нейронная сеть в силах отличить образец голоса одного диктора от множества образцов голосов других дикторов, записанных в базу данных, а также, распознавать образцы голоса, которые не принадлежат не одному из дикторов. Из этого можно сделать вывод, что нейронные сети – это отличный и надёжный инструмент для нашей задачи идентификации человека по голосу.

Таким образом, был разработан способ идентификации человека по голосу с использованием искусственных нейронных сетей. Метод кепстрального преобразования спектра позволяет выделять речевые признаки, характерные для человека, а искусственная нейронная сеть позволяет выполнить распознавание голоса диктора и отличить его от множества других голосов.

Список использованных источников:

1. Huang X., Acero A., Hon H.-W. (2001). Spoken Language Processing: a Guide to Theory, Algorithm, and System Development. Prentice-Hall, New Jersey.
2. Farrell K., Mammone R., Assaleh K. (1994). Speaker recognition using neural networks and conventional classifiers. IEEE Trans. Speech Audio Process., v.2, N1, 194–205.