Math-Net.Ru

Общероссийский математический портал

Т. И. Антипина, Нечеткая модель в задаче идентификации диктора по голосу,
Нечеткие системы и мягкие вычисления, 2018, том 13, выпуск 1, 37–43

DOI: <https://doi.org/10.26456/fssc39>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и
согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 3.89.197.203

26 апреля 2023 г., 17:27:08

НЕЧЕТКАЯ МОДЕЛЬ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ ДИКТОРА ПО ГОЛОСУ

Антипина Т.И.

Тверской государственный университет, г. Тверь

Поступила 09.01.2018, после переработки 13.02.2018.

В статье представлена система идентификации диктора по голосу, разработанная на основе нечеткой модели. Рассматривается структура нечеткой модели для решения задачи идентификации по голосу. Разработан алгоритм принятия решения о распознавании на основе выходных данных нечеткой системы. Приведены результаты работы на реальной речевой базе.

Ключевые слова: биометрическая идентификация, текстонезависимая идентификация диктора, распознавание по голосу, нечеткие модели.

Нечеткие системы и мягкие вычисления. 2018. Т. 13, № 1. С. 37–43.
<https://doi.org/10.26456/fssc39>

Введение

Инструменты нечеткой логики широко используются для решения задач обработки речевых сигналов, таких как распознавание речи [3] и распознавание голоса. Работа [5] является основным примером решения задачи распознавания диктора с применением теории нечеткой логики. В работе [5] решается задача текстозависимой идентификации голоса и применяется нечеткая логика второго типа в качестве вспомогательного инструмента для базового алгоритма на основе нейронных сетей. В данной работе рассматривается пример системы текстонезависимой идентификации голоса на основе нечеткой модели для решения задачи распознавания диктора по голосу. То есть основным отличием от работы [5] является применение теории нечетких множеств как основного и единственного подхода. Также в данной статье используется нечеткая модель на основе нечетких множеств 1-го типа, в отличие от [5], где применяются алгоритмы нечетких множеств 2-го типа. В статье приведены способы представления входного и выходного сигналов нечеткой системы Мамдани [1], решающей задачу идентификации голоса, а также алгоритм принятия решения по множеству ответов нечеткой системы на множестве векторов признаков неизвестного речевого сигнала. Результаты тестирования системы, разработанной по описанным в статье алгоритмам, достигают 98,25% точности распознавания, в отличие 96% успешного распознавания, приведенных в результатах работы [5].

Определение количества переменных и нечетких множеств для входного и выходного сигнала нечеткой системы представляет из себя нетривиальную задачу и, по сути, является задачей определения структуры нечеткой модели [1].

В статье предлагаются два типа нечетких систем – система с одним входом и одним выходом (SISO) и система с множеством входов и одним выходом (MISO). Приводится описание и сравнение результатов распознавания обоих типов для выбора оптимальной архитектуры.

Поскольку для неизвестного голосового сигнала извлекается множество векторов признаков, каждый из которых подается на вход нечеткой системы и генерирует ответный выход системы, то возникает проблема принятия решения по полученному множеству ответов нечеткой системы. В данной статье подробно описывается алгоритм принятия решения о распознавании по полученному множеству выходов нечеткой системы.

1. Архитектура нечеткой системы с одним входом и одним выходом (SISO)

Всюду в данной работе полагаем, что рассматриваемая система является нечеткой системой логического вывода Мамдани [1]. При моделировании системы типа SISO в качестве входного сигнала w предлагается использовать вектор признаков, вычисленных по записи голоса. В данной работе используются n -мерные вектора мел-частотных кепстральных коэффициентов. Таким образом, нечеткая система будет содержать одну входную переменную w_1 – вещественный n -мерный вектор X . В случае такой системы обучающие значения ее входов будут неравномерно распределены в n -мерном пространстве и искажены шумами измерений [2]. Поэтому предлагается предварительно использовать кластеризацию полученных измерений в n -мерном пространстве с помощью того или иного алгоритма кластеризации. В результате кластеризации имеются m четких кластеров для каждого диктора, характеризующихся центром и растяженностью. Полученные кластеры будут соответствовать нечетким множествам входной переменной $\{W_{ij}, i = 1, \dots, d, j = 1, \dots, m\}$, d – количество зарегистрированных дикторов. Каждое нечеткое множество характеризуется функцией принадлежности, которая, по сути, является мерой, обратной расстоянию до центра кластера V_i . Предлагается использовать евклидово расстояние $D_{ki} = \sqrt{\|X_k - V_i\|^2}$, $k = 1, \dots, m, i = 1, \dots, c$. Таким образом, функция принадлежности вектора X кластеру i будет иметь вид:

$$\pi(X) = 1 - (D_{ki})/d_{max}, \quad (1)$$

где d_{max} – самое большое расстояние между обучающими векторами, V_i – центр кластера i .

Системы типа SISO имеют единственную выходную переменную o_1 . Для решения задачи идентификации диктора удобно использовать в качестве выходного сигнала o_1 номер идентифицированного диктора. Нечеткие множества $O_1, O_2, \dots, O_d$ выходной переменной o_1 будут характеризоваться триангулярной функцией распределения с модальным значением, равным порядковому номеру диктора и интервалом толерантности, равным 1: $O_i = Tr(i, 1)$, где $i \in \overline{1, d}$.

В рассмотренном случае правила базы правил будут строиться следующим образом: если $w_1 = W_{ij}$, то $o = O_i$.

2. Архитектура нечеткой системы с множеством входов и одним выходом (MISO)

Рассмотрим вариант нечеткой модели типа MISO для решения задачи иденти-

фикации диктора по голосу. Множество входов данной модели будут составлять n компонент вектора мел-частотных кепстральных коэффициентов $w_1, w_2, \dots, w_n$. Для выходного сигнала предлагается использовать переменную o_1 , аналогичную выходной переменной системы SISO, описанную выше.

Рассмотрим организацию нечетких множеств на примере первой компоненты вектора мел-частотных кепстральных коэффициентов. На экспериментальной базе было установлено, что первая мел-частотная компонента может изменяться в диапазоне $[-172.873; 371.15]$. В данном промежутке для каждого диктора значения первой компоненты неравномерно разбросаны. Поэтому предлагается рассматривать проекции кластеров, полученных после процедуры кластеризации множества обучающих векторов одного диктора. Таким образом, для каждого диктора, зарегистрированного в системе, получим m проекций кластеров, где m – число кластеров, на которые разбивается множество обучающих векторов одного диктора. Каждая проекция характеризуется 1-й координатой центра кластера c_i и длиной проекции h при необходимости. Данные проекции могут быть использованы для построения нечетких множеств $\{W_{1ij}, i = 1, \dots, d, j = 1, \dots, m\}$, с характеризующими их параметрами – центром и растяженностью. Остается лишь выбрать вид функций принадлежности для данных нечетких множеств. Для сравнения результатов идентификации в исследовании были использованы функции принадлежности разных типов: триангулярная, трапецевидная, нормальная и функция принадлежности π_2 , основанная на расстоянии до проекции центра кластера, аналогично функции (1). Наилучшие результаты были получены при использовании π_2 в качестве функции принадлежности числа x нечеткому множеству, задаваемому центром и длиной проекции:

$$\pi_2 = \begin{cases} 1 - \frac{|x - c_i|}{h_{i/2}}, & |x - c_i| \leq h_{i/2}, \\ 0, & \text{иначе,} \end{cases} \quad (1)$$

где c_i – i -я координата центра кластера, h_i – длина i -й проекции кластера.

Аналогичным образом определяем нечеткие множества для всех n входных переменных $\{w_1, w_2, \dots, w_n\}$: $\{W_{kij}, k = 1, \dots, n, i = 1, \dots, d, j = 1, \dots, m\}$. Тогда база правил системы будет состоять из правил следующего вида: если $w_1 = W_{1ij}$ и $w_2 = W_{2ik}$ и ... и $w_n = W_{nil}$, то $o = O_i$.

3. Алгоритм принятия решения о распознавании диктора по выходным значениям нечеткой системы

Как было сказано выше, для каждого незнакомого диктора по записи его голоса рассчитывается множество векторов признаков $\{v_j, j = 1, \dots, m\} = \{v_{j1}, v_{j2}, \dots, v_{jm}\}$ и каждый из полученных векторов подается на вход нечеткой системы. Полученное на выходе нечеткой системы четкое значение o'_j округляется до модального значения соответствующей функции принадлежности o_j . Таким образом, для каждого вектора признаков v_j , рассчитанного по записи незнакомого диктора, имеется итоговый результат в виде номера идентифицированного диктора o_j и соответствующее значение функции принадлежности $\pi(o'_j)$.

На основе множества полученных пар $O = (o_j, \pi(o'_j))$, $j = \overline{1, m}$ вычисляется множество $\{k_i, i = \overline{1, d}\}$, где k_i – количество пар $(i, \pi(o'_j)) \in O$, d – число голосов, используемых для обучения системы. Также вычисляется дополнительное множество $\{l_i, i = \overline{1, d}\}$, где l_i – количество пар $(i, \pi(o'_j) > p) \in O$, p – заданный

параметр. Вычисляем максимальный элемент $k_{max_1} \in \{k_i\}$. Итоговый результат вычисляется по следующей формуле:

1. $\max_1$, если $k_{max_1} > p_1$ и $l_{max_1} > p_2$, p_1 и p_2 — заданные параметры,
2. диктор не идентифицирован, иначе.

Рассмотрим множества возможных значений для используемых параметров p , p_1 и p_2 . Параметр p используется для оценки значений функции принадлежности, и поэтому должен быть больше 0.5. p_1 — минимальное количество входных сигналов, приводящее к правильному ответу системы. Очевидно, что данное число должно быть больше, чем половина количества всех векторов, подающихся на вход системы. По описанному выше способу работы нечеткой системы все множество векторов, полученное по голосовой записи, разбивается на k кластеров с помощью алгоритма k-means, и центры этих кластеров используются в качестве входных сигналов нечеткой системы. Во время тестирования системы максимально высокие показатели распознавания были достигнуты при $k = 16$. Таким образом, p_1 должен быть больше 8. p_2 — параметр, отвечающий за минимальное количество векторов, поданных на вход нечеткой системе, приводящее к правильному ответу системы с принадлежностью большей p . Значение этого параметра может быть меньше p_1 , но больше 0.

Точные значения параметров были получены экспериментальным путем во время тестирования разработанной программы: $p = 0.76$, $p_1 = 10$, $p_2 = 3$.

4. Анализ работы систем типа SISO и MISO

Программное обеспечение для идентификации диктора по голосу, разработанное в соответствии с описанной структурой нечеткой системы, было реализовано на языке C++ в среде разработки Visual Studio 2013. Использовались следующие параметры, упомянутые в описании системы: $n = 19$, $m = 16$ и алгоритмы: алгоритм k-means в качестве алгоритма кластеризации [4], и метод первого максимума в качестве метода дефазификации.

Критериями качества работы систем идентификации являются вероятности ошибок первого и второго рода. Ошибка первого рода (FRR — False Rejection Rate) возникает в случае, если принимается решение о том, что диктор не принадлежит списку зарегистрированных пользователей, в то время как на самом деле он там присутствует. Ошибка второго рода (FAR — False Acceptance Rate) напротив, возникает при ложном допуске незарегистрированного пользователя. В данной работе значения показателей FRR и FAR рассчитывались как отношения числа неудачных тестов к общему числу тестов. Помимо этих критериев предлагается сравнить для различных методов такой показатель качества распознавания как процент точной верификации. Под данным критерием понимается процентное соотношение количества правильно допущенных пользователей с количеством допущенных и правильно распознанных пользователей.

Для обучения систем использовалось 10 аудио файлов, каждый из которых представляет запись речи одного человека с частотой дискретизации 12 кГц. Каждый диктор произносит произвольную фразу от одного до пяти произвольных слов. Слова произносились как на русском, так и на английском языке. Для тестирования всех систем таким же образом было записано 20 аудио файлов, по одному для каждого диктора. 10 тестовых записей принадлежат зарегистрированным пользователям, а оставшиеся 10 представляют записи речи неизвестных

дикторов. Записи аудио-файлов зарегистрированных пользователей для тестирования систем были сделаны на несколько месяцев позже, чем записи тех же пользователей для обучения систем. Для исследования была записана речь 15 мужчин и 15 женщин в возрастном диапазоне от 20 до 55 лет. Мел-частотные кепстральные коэффициенты состоят из 19 компонент. Запись производилась с помощью микрофона Soundking EH031 на компьютере Lenovo в операционной системе Windows 7. Таким образом, стоит отметить, что в работе сравнение всех методов и алгоритмов проводится в самом сложном случае текстонезависимой идентификации, когда речь подвержена временным, фонетическим и лексическим изменениям.

Результаты проведенных работ представлены в Таблице 1.

Таблица 1: Эффективность распознавания системами типов MISO и SISO

Тип системы	Процент точной верификации	FRR	FAR
MISO	98,25%	0	1,75%
SISO	97,75%	0	2,25%

Полученные данные говорят о том, что количество входных переменных незначительным образом влияет на итоговый результат распознавания. Тем не менее, система типа MISO с n входами превосходит систему с одним входом.

Сложность алгоритма формирования базы правил для приведенных выше систем типа MISO и SISO равна $O(dm^{2k+1} \log m)$, где d – количество дикторов, зарегистрированных в системе, m – количество векторов признаков, извлеченных из голосового сигнала одного диктора, k – число кластеров, на которые нужно разбить множество всех обучающих векторов. Эта оценка объясняется сложностью выполнения алгоритма кластеризации k-means $O(m^{2k+1} \log m)$ [4], умноженной на количество зарегистрированных дикторов d , так как алгоритм кластеризации применяется для векторов каждого диктора, чтобы получить соответствующие правила.

Рассмотрим сложность работы системы на неизвестном входном сигнале. Нечеткая система типа SISO работает таким образом, что для входной переменной вычисляется значение функции принадлежности для каждого нечеткого множества $C_{ij} \ i = 1, \dots, k, j = 1, \dots, d$, полученного на этапе обучения. Функция принадлежности вычисляется алгоритмом порядка $O(n)$, где n – размерность пространства векторов признаков. То есть вся операция вычисления принадлежности всем нечетким множествам входной переменной будет $O(dkn)$. Поскольку каждое правило базы правил контроллера содержит только одну переменную в условии, то и итоговая сложность будет равна $O(dkn)$.

Легко показать, что сложность алгоритма MISO также равна $O(dkn)$. Дело в том, что принадлежность будет вычисляться $O(c)$, но это будет происходить n раз для каждого правила из базы, так как в условиях используются n переменных. Количество правил в базе равно dk , то есть совпадает с размерностью базы правил системы SISO. Таким образом, получаем итоговую сложность $O(dkn)$.

Полученные результаты позволяют сделать вывод о том, что по сложности вычислений системы SISO и MISO не отличаются.

Заключение

Результаты, приведенные в Таблице 1 для типа MISO, показывают, что подход

к решению задачи распознавания голоса с использованием инструментов нечеткой логики подходит для дальнейшего исследования с целью повышения точности распознавания диктора по голосу. Также данные результаты позволяют использовать описанное в статье программное обеспечение самостоятельно или в качестве дополнительного в системах распознавания диктора по голосу.

Список литературы

- [1] Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 287 с.
- [2] Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия, 2006. 452 с.
- [3] Al-Jawadi A.M.A. Speech Recognition and Retrieving using Fuzzy Logic System. Mosul, Iraq: Technical College, Foundation of Technical, Engineering. 2009. Pp. 1–10.
- [4] Arthur D., Vassilvitskii S. K-means++: the advantages of careful seeding // Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2007). New Orleans, Louisiana, USA, January 7-9, 2007. Pp. 1027–1035.
- [5] Melin P., Urias J., Solano D., Soto M., Lopez M., Castillo O. Voice recognition with neural networks, type-2 fuzzy logic and genetic algorithms // Engineering Letters. 2006. Vol. 13, № 2. Pp. 108–116.

Образец цитирования

Антипина Т.И. Нечеткая модель в задаче идентификации диктора по голосу // Нечеткие системы и мягкие вычисления. 2018. Т. 13, № 1. С. 37–43. <https://doi.org/10.26456/fssc39>

Сведения об авторах

1. Антипина Татьяна Игоревна

аспирант кафедры информационных технологий Тверского государственного университета.

Россия, 170100, г. Тверь, ул. Желябова, д. 33, ТвГУ.

E-mail: tatyana.buchneva@mail.ru

FUZZY MODEL IN THE PROBLEM OF THE SPEAKER IDENTIFICATION BY VOICE

Antipina Tatyana Igorevna

PhD student at Information Technologies department,
Tver State University
Russia, 170100, Tver, 33 Zhelyabova str., TverSU.
E-mail: tatyana.buchneva@mail.ru

Received 09.01.2018, revised 13.02.2018.

The article presents the speaker identification system developed on the basis of fuzzy model. The article provides the structure of a fuzzy model for solving the problem of voice identification. The method of decision-making is proposed. This method bases on the output data of the fuzzy system. Data are given about the results of identification for real speech base.

Keywords: biometric identification, text-independent speaker identification, speaker recognition, fuzzy sets, fuzzy model.

Nechetkie Sistemy i Myagkie Vychisleniya [Fuzzy Systems and Soft Computing], 2018, vol. 13(1), pp. 37–43. <https://doi.org/10.26456/fssc39>

References

- [1] Osovskij S. *Nejronnye Seti Dlya Obrabotki Informatsii* [Neural Networks for Information Processing]. Finansy i Statistika Publ., Moscow, 2002. 287 p. (in Russian)
- [2] Rutkovskaya D., Pilin'skij M., Rutkovskij L. *Nejronnye Seti, Geneticheskie Algoritmy i Nechetkie Sistemy* [Neural Networks, Genetic Algorithms and Fuzzy Systems]. Goryachaya Liniya Publ., Moscow, 2006. 452 p. (in Russian)
- [3] Al-Jawadi A.M.A. *Speech Recognition and Retrieving using Fuzzy Logic System*. Technical College, Foundation of Technical, Engineering, Mosul, Iraq, 2009. Pp. 1–10.
- [4] Arthur D., Vassilvitskii S. K-means++: the advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*. New Orleans, Louisiana, USA, January 7-9, 2007. Pp. 1027–1035.
- [5] Melin P., Urias J., Solano D., Soto M., Lopez M., Castillo O. Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms. *Engineering Letters*, 2006, vol. 13(2), pp. 108–116.

Citation

Antipina T.I. Fuzzy model in the problem of the speaker identification by voice. *Nechetkie Sistemy i Myagkie Vychisleniya* [Fuzzy Systems and Soft Computing], 2018, vol. 13(1), pp. 37–43. (in Russian) <https://doi.org/10.26456/fssc39>