

Оценка информативности параметров речевого сигнала при обработке с использованием нейронных сетей

Костюченко Е.Ю.

Коцубинский В.П.

Авдеев А.А.

Людвиг К.В.

Тюменцев И.В.

Томский государственный университет систем управления и радиоэлектроники
634050 г. Томск, пр. Ленина, 40.
Тел. (факс) (3822) 413-426;
E-mail: office@keva.tusur.ru

В работе представлены результаты оценки коэффициента корреляции между параметрами и его влияния на совместную информативность применяемых параметров для исследования речевого сигнала. При исследовании критерия, основанного на результатах обработки тестовой выборки обученной нейронной сетью, сформулированы и показаны подходы к выбору и применению конкретных параметров наблюдаемой выборки значений критерия информативности в зависимости от количества планируемых экспериментов. Экспериментально показана независимость ранжирования параметров по информативности от используемых характеристик выборки. Экспериментально показана зависимость ранжирования параметров по информативности от класса решаемой задачи относительно приоритета ошибки первого или второго рода и конкретных значений коэффициентов при вычислении значения информативности

При использовании в качестве анализируемого речевого сигнала исходный сигнал подаётся на систему фильтров с целью выделения параметров для дальнейшего анализа. В данной работе в качестве системы фильтров применялась система нерекурсивных фильтров, учитывающих особенности слуховой системы человека, с целью выделить параметры речевого сигнала в соотношении, описывающем восприятие сигнала после обработки его слуховой системой [1]. После обработки речевого сигнала на выходе системы фильтров получается спектр, разбитый по каналам, соответствующим критическим полосам слуховой системы человека, представленный в виде матрицы размером $N_{ch} \times \text{LenSign}$, где N_{ch} — количество каналов анализа, LenSign — длина исследуемого сигнала в отсчетах. Далее производится формирование собственно параметров для обучения нейронной сети. В качестве параметров для исследования выбраны различные статистические характеристики вектора интенсивностей по всем каналам в рассматриваемый момент времени.

Исходя из требований к речевому сигналу и его параметрам, предназначенным для анализа с применением искусственных нейронных сетей прямого распространения, были выбра-

ны следующие условия для проведения экспериментальной части работы по оценке информативности параметров:

1) частота дискретизации сигнала 12 кГц, теоретический максимум анализа составляет в этом случае 6 кГц, на практике получается строить спектральную картину до 4 кГц, что позволяет в случае необходимости получать формантную структуру сигнала до третьей форманты [2], что является достаточным при решении задач идентификации диктора [3]);

2) обучение проводилось применительно к решению задачи разбиения сигнала на вокализованные и невокализованные участки;

3) 256 каналов анализа;

4) размер окна выбран исходя из минимальной продолжительности вокализованного участка для речевого сигнала 360 отсчётов, что соответствует 30 мс;

5) количество входов нейронной сети определяется размерностью параметров, при этом количество параметров в условиях данного эксперимента не должно превышать 3, поскольку обучение на 4 параметрах, что соответствует 1440 входам, не представляется возможным ввиду чрезмерно большой ошибки уже на этапе обучения (итоговая ошибка не менее 0,25);

6) исследуемые параметры:

- *mosp* — среднее значение интенсивности на отсчёте;
- *dispr* — дисперсия интенсивности на отсчёте;
- *asim* — коэффициент асимметрии интенсивности на отсчёте;
- *asim1* — отношение медианы интенсивности к среднему значению на отсчёте;
- *svar* — вариация интенсивности на отсчёте;
- *lmosp* — логарифм среднего значения интенсивности на отсчёте;
- *maxch* — номер канала с максимальной интенсивностью на отсчёте;
- *medi* — медиана интенсивности на отсчёте;
- *prand* — случайная величина, равномерно распределенная в диапазоне [-0,5; 0,5];
- *disprcorr* — параметр, представляющий собой линейную комбинацию параметров *dispr* и *prand*.

Для исследования выбраны параметры:

- множество возможных параметров является открытым. Всегда можно синтезировать новый параметр, не используемый ранее;
- процедура выделения данных параметров имеет линейную сложность относительно количества каналов и размера окна анализа после операции получения спектра сигнала. Данное соображение является актуальным, поскольку получение сигнала после фильтрации имеет высокую вычислительную сложность и добавление дополнительных вычислительно ёмких процедур приведёт к существенному общему росту общего времени вычислений;
- первоначальный выбор параметров не влияет на алгоритм определения оптимальных параметров интерпретатора и на алгоритм их ранжирования по информативности.

После составления обучающей выборки на основе наблюдаемого диапазона изменения параметров диапазон приводился к [-1; 1] (нормировка). Данный этап необходим для корректного обучения нейронной сети и обеспечения

влияния параметров, не зависящего от разницы их абсолютных значений, кроме того нормализация позволяет увеличить скорость обучения нейронной сети в случае значительного разброса значений входных параметров.

Эталонная обучающая выборка получена после выполнения последовательности следующих действий, представляющей собой алгоритм формирования наборов для нейронной сети:

- для получения обучающей выборки записаны 25 сигналов, представляющих собой запись произнесённых диктором фраз, выбранных на основе [4];
- проведена ручная сегментация данных сигналов по признаку наличия или отсутствия вокализации на сегменте;
- проведена фильтрация полученных речевых сигналов системой фильтров, учитывающих особенности речевой системы человека [5];
- из результата фильтрации рассчитаны значения параметров для обучения нейронной сети для всего речевого сигнала;
- из значений параметров сформирована обучающая выборка путём вырезки из полученных параметров речевого сигнала окон длиной 360 отсчетов, что соответствует 30 мс. Нарезка таких окон осуществлялась с шагом 180 отсчетов (15 мс), за исключением «буферной зоны», составляющей участки сигнала, удалённые от границы сегмента менее чем на 180 отсчетов (15 мс). Это сделано с целью, чтобы на протяжении окна эталонной выборки не происходили переходные процессы, происходящие при смене атрибутов между сегментами;
- каждому из полученных таким образом примеров сопоставлено значение атрибута, соответствующее значению этого атрибута на сегменте, которому принадлежит это окно.

После выполнения данной последовательности действий получена обучающая выборка объёмом $N_{об} = 2038$ примеров.

Для получения тестовой выборки осуществлена сегментация дополнительно 25 сигналов, представляющих собой запись произнесённых диктором фраз, выбранных на основе [4], и не участвовавших при формировании тестовой выборки. После этого к данным сигналам был применён алгоритм формирования наборов для нейронной сети с тем отличием, что шаг при нарезке составил 60 отсчетов (5 мс) и нарезка наборов производилась, в том числе, и с «буферных зон». Объём тестовой выборки составил $N_T = 7608$ примеров.

В таблице 1 приведено значение достигнутых уровней ошибки для различного количества нейронов в промежуточном слое и количество циклов обучения, за которое достигается ошибка, равная 110% от конечного значения. Шаг при оценке был равен 25 циклам обучения. Обучение проводилось на примере параметра $maxch$ — номер канала с максимальной интенсивностью.

Таблица 1

Зависимость обучаемости нейронной сети от количества нейронов в промежуточном слое для параметра $maxch$

	50	Циклы	90	Циклы	100	Циклы	110	Циклы	150	Циклы
1	0,009201	1150	0,009174	1125	0,008838	725	0,007926	1025	0,008023	825
2	0,009834	1550	0,006487	1350	0,006869	925	0,006385	925	0,00666	825
3	0,00884	1375	0,007231	1125	0,006379	750	0,006474	875	0,007192	1150
4	0,007135	1400	0,00825	925	0,006484	1125	0,007846	950	0,005629	775
5	0,008138	975	0,009114	975	0,005419	1000	0,005783	975	0,006237	950
Ср.	0,00863	1290	0,008051	1100	0,006798	905	0,006883	950	0,006748	905

При проведении экспериментов было замечено значительное ухудшение обучаемости при снижении количества нейронов с 100 до 50. Это было выявлено на основании снижения частоты достижения обученности нейронной сети до минимального приемлемого уровня до значения 0,5. Даже без учёта данных грубых обучений итоговая ошибка для нейронной сети со 100 нейронами в скрытом слое всё равно более чем на 20% выше, чем в случае нейронной сети с 50 нейронами в скрытом слое. При увеличении количества нейронов существенных изменений уровня обученности сети не наблюдается, однако существенно возрастает время обучения нейронной сети за счёт увеличения времени, затрачиваемого на 1 цикл обучения. Кроме того, за время эксперимента не наблюдалось останова процесса обучения за счёт достижения предельного количества обучений в 5000 циклов и было принято решение оставить это значение в качестве порогового критерия останова в случае неустойчивого обучения нейронной сети, поскольку оказывается быстрее сгенерировать и обучить новую нейронную сеть. На основании проведённого эксперимента была выбрана нейронная сеть с количеством нейронов в промежуточном слое равным 100 и предельным количеством циклов обучения равным 5000.

На основе разработанной методики по минимизации критерия, зависящего от результатов обработки тестовой выборки, проведена серия из 20 экспериментов для каждого из параметров, в рамках каждого из которых проводилось создание и обучение новой нейронной сети, с целью получения выборки оценок информативности на основе анализируемого критерия с целью последующей статистической обработки. Результаты экспериментов для сегментации сигнала на вокализованные и невокализованные участки для 10 экспериментов представлены в таблице 2.

Таблица 2

Значения критерия информативности для выбранных параметров

№ Эксп.	1	2	3	4	5	6	7	8	9	10
maxch	0,4149	0,4122	0,4074	0,4063	0,3971	0,4111	0,4096	0,3894	0,3978	0,4082
medi	0,5743	0,5627	0,5446	0,5716	0,5634	0,5534	0,5509	0,5754	0,5501	0,5431
mosp	0,5981	0,5944	0,6155	0,6055	0,5965	0,5989	0,5982	0,6094	0,6054	0,617
1mosp	0,7912	0,827	0,6437	0,6682	0,6797	0,6997	0,7328	0,6585	0,74	0,6841
asim	0,7276	0,7711	0,708	0,7127	0,6602	0,7136	0,6682	0,718	0,6999	0,7161
asim1	0,4868	0,4928	0,5207	0,4557	0,5	0,5319	0,5018	0,4876	0,5087	0,4839
dispr	0,5054	0,5559	0,5478	0,5534	0,5342	0,5276	0,5131	0,5308	0,5111	0,5488
cvar	0,7105	0,6446	0,7955	0,6605	0,6087	0,8221	0,6328	0,8043	0,8602	0,6536

Общее число обучений нейронной сети для каждого из параметров составило 20 раз. На основе этих данных определены такие параметры значений критерия, как среднее значений mean, дисперсия disp, медиана median и минимальное наблюдаемое значение min.

Для исследования влияния линейной зависимости параметров была рассчитана матрица коэффициентов корреляции. Примеры значений каждого из анализируемых параметров на окне размеров в 360 отсчётов представлены для параметров на рисунке.

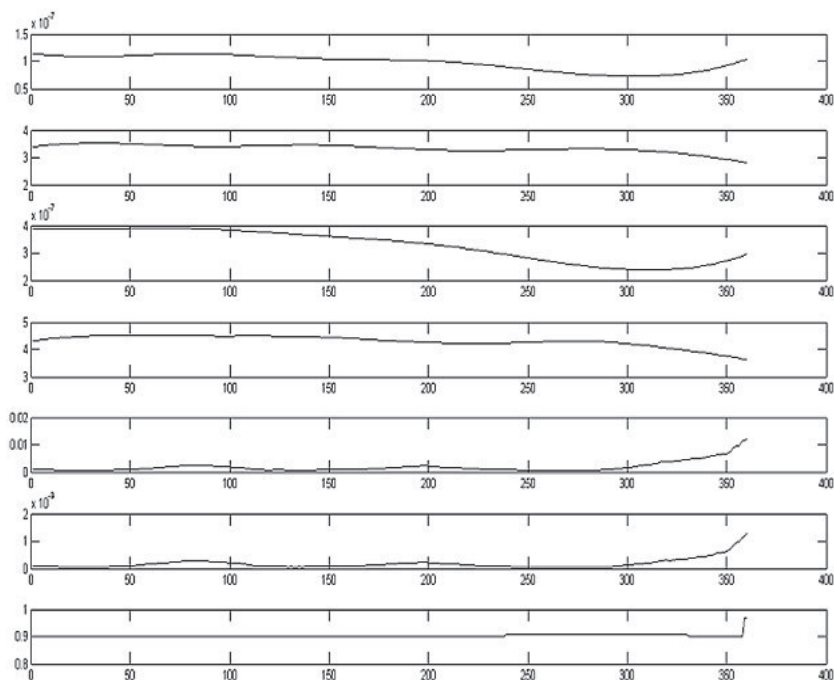


Рис. Значения выделяемых параметров речевого сигнала на примере одного окна анализа (360 отсчётов)

Идея эксперимента заключалась в выявлении более важного влияния на нейронную сеть при совместном использовании нескольких (трёх) параметров. Выбор осуществлялся между информацией, содержащейся в параметрах и коэффициенте корреляции. Значения коэффициентов корреляции между каждой парой параметров, полученные на 50 сигналах, представлены в таблице 3.

Таблица 3

Значения коэффициентов корреляции между параметрами

	mosp	dispr	asim	asim1	cvar	1mosp	maxch	medi
mosp	1	0,9303	-0,1383	-0,0672	-0,0537	0,4423	-0,1942	0,795
dispr	0,9303	1	-0,0327	-0,1624	0,0894	0,4543	-0,1924	0,5774
asim	-0,1383	-0,0327	1	-0,2307	0,7464	-0,1357	0,2234	-0,1706
asim1	-0,0672	-0,1624	-0,2307	1	-0,6792	-0,5296	0,1099	0,1558
cvar	-0,0537	0,0894	0,7464	-0,6792	1	0,1749	0,0351	-0,1778
1mosp	0,4423	0,4543	-0,1357	-0,5296	0,1749	1	-0,2757	0,2635
maxch	-0,1942	-0,1924	0,2234	0,1099	0,0351	-0,2757	1	-0,1157
medi	0,795	0,5774	-0,1706	0,1558	-0,1778	0,2635	-0,1157	1

После ранжировки на основе коэффициента корреляции был произведён выбор трёх параметров для обучения нейронной сети на основании критерия минимального значения максимального модуля коэффициента корреляции среди параметров выбранной группы. На основании этого были выбраны параметры *asim1*, *maxch* и *medi*. Кроме того была сформирована контрольная группа параметров на основании обратного критерия (максимальное значение минимального модуля) из параметров *mosp*, *dispr* и *medi*. На основании сформированной выборки было проведено обучение нейронной сети

и оценка критерия информативности при равной значимости ошибок первого и второго родов и значении коэффициента нелинейности s равным 1 (сумма частот встречаемости ошибок первого и второго рода). На основании этого по результатам 20 обучений были получены значения критерия $M=0,3591$ для параметров, выбранных с учётом корреляции и $M=0,4201$ для значений без учёта данного коэффициента. Уменьшение значения критерия информативности за счёт использования коэффициента корреляции составило 17%, что показывает необходимость учёта взаимной корреляции между параметрами при выборе входных параметров нейронной сети.

ЛИТЕРАТУРА

1. *Bondarenko V. P., Moor V. R., Chabanets A. N.* The analysis of speech perception mechanisms on the models of auditory system // Proceedings XIth ICPHS. Tallinn, 1987. V. 2. P. 77–80.
2. *Михайлов В. Г., Златоустова Л. В.* Измерение параметров речи / Под ред. М.А. Сапожкова. М.: Радио и связь, 1987. 168 с.
3. *Сапожков М.А.* Речевой сигнал в кибернетике и связи. М.: Государственное издательство литературы по вопросам связи и радио, 1963. 450 с.
4. ГОСТ Р 50840-95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости.
5. *Конев А. А. Мещеряков Р. В.* Алгоритм сегментации речевого сигнала на вокализованные и невокализованные участки // Сборник трудов XIX сессии Российского акустического общества. Т. III. М.: ГЕОС, 2007. С. 56–60.