

МЕТОД ФИЛЬТРАЦИИ ВЕКТОРОВ ПРИЗНАКОВ В ЗАДАЧЕ ТЕКСТОНЕЗАВИСИМОЙ ИДЕНТИФИКАЦИИ ДИКТОРА

Н.С. Клименко

Институт проблем искусственного интеллекта
МОН Украины и НАН Украины, Киев, Украина

В статье предложена модификация метода идентификации при помощи фильтрации множества векторов акустических признаков как обучающей выборки, так и выборки тестируемого образца. Результаты численных исследований показали, что данная модификация позволяет повысить вероятность текстонезависимой идентификации на основе одного классификатора. Вместе с тем, объем вычислений сокращается в сравнении с использованным ранее решающим правилом, полученным бустинг-алгоритмом.

Введение

Данная работа является продолжением [1], где была рассмотрена адаптация алгоритма бустинга к задаче текстонезависимой идентификации диктора. Текстонезависимой является идентификация, выполняемая по произвольному фрагменту звукозаписи речи без информации о ее текстовом содержимом. Эта возможность усложняет реализацию текстонезависимых систем, понижает их надежность и скорость распознавания.

Бустинг – метод, основанный на комбинировании нескольких «слабых» классификаторов в один «сильный». Решающее правило, полученное бустинг-алгоритмом AdaBoost, показывает лучшую вероятность распознавания в сравнении с единичными классификаторами.

Работа заключается в повышении эффективности единичного классификатора для решения проблемы значительно возросшего объема вычислений, что привело к невозможности работы с большой акустической базой дикторов в режиме реального времени.

Описание

Для выбора подхода фильтрации был рассмотрен ряд работ, посвященных как фильтрации множества векторов признаков, так и оптимизации подмножества самих признаков [2-4]. Метод минимальной избыточности и максимальной релевантности (minimum redundancy-maximum relevance, mRMR) успешно применяется для фильтрации пространства признаков в задачах распознавания образов [4].

Критерий максимальной релевантности находит такое подмножество признаков, что оно максимизирует среднее значение взаимной информации признака и класса по всем признакам, включенным в подмножество. Критерий был приспособлен для решения задачи фильтрации акустических признаков следующим образом:

$$mRMR(S) = \max_S \left[\frac{1}{|S|} \sum_{x^i \in S} I(x^i, c_j) - \frac{1}{|S|^2} \sum_{x^i, x^j \in S} I(x^i, x^j) \right],$$

где x^i – вектор множества акустических признаков S ;

c_j – акустическая модель диктора;

S – подмножество признаков;

$I(x^i, c_j)$ – расстояние в пространстве признаков или критерий корреляции.

В качестве акустических признаков в данной работе выступают Мел-частотные кепстральные коэффициенты, зарекомендовавшие себя компактными и информативными признаками во множестве современных исследований как по распознаванию речи, так и диктора [5, 6]. Вектор признаков дополнен производными от аналогичных коэффициентов, полученных по соседним окнам, для учета динамической составляющей. Таким образом, общая длина вектора признаков равна 39.

В рамках данной работы для классификации использован метод гауссовых смесей. Этот классификатор может представлять собой аналогию скрытой марковской модели с одним состоянием. Модели гауссовых смесей отображают классы в пространстве признаков в виде многомерного вероятностного распределения. Класс описывается как взвешенная сумма M нормальных распределений. Таким образом, модель описывается векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели [5]. Способом оценки параметров модели является метод максимизации правдоподобия.

Модификацией метода гауссовых смесей и главной особенностью предложенной схемы идентификации является учет широких фонетических классов (ШФК) – множества, характеризующего акустическое пространство голоса диктора. ШФК в данной работе представляют акустические события, на основании которых идет дальнейшая работа методов классификации. ШФК описывают определенные фонетические события (гласные, фрикативные и т. д.), по которым в отдельности формируются акустические модели, в совокупности представляющие «комплексную» модель диктора. По результатам численных исследований был использован состав из четырех ШФК: глухие, звонкие, сонорные и гласные. Такой состав признан оптимальным для звуков речи русского языка по параметрам межклассового и внутриклассового разбросов векторов признаков в ходе проведенного численного исследования [1].

С учетом предложенного критерия фильтрации на основе имеющихся разработок было сформировано множество классификаторов для проведения численного исследования. В него включены классификаторы на основе одного и всех широких фонетических классов, а также решающее правило на основе бустинг-алгоритма. Звукозаписи дикторов были взяты из общедоступной базы VoxForge [7]. Была получена выборка звукозаписей 50 дикторов. Звукозаписи представлены в виде wav-файлов с глубиной квантования 16 бит и частотой дискретизации 16-44,1 кГц. Все дикторы записаны динамическим микрофоном в спокойном эмоциональном состоянии, 38 дикторов — мужчины, 12 — женщины.

Исследования проводились на обучающих выборках с разным количеством распознаваемых классов (дикторов) от 2 до 50. Обучение проводилось на фрагментах речи дикторов продолжительностью от 8 до 35 секунд. Распознавание проводилось по фрагментам дикторов той же сессии звукозаписи продолжительностью 5-7 секунд.

Проведен сравнительный анализ эффективности идентификации диктора классификаторами по отдельным ШФК, комплексным классификатором на их основе и композицией классификаторов. Зависимость вероятности распознавания от размера обучающей выборки (количества моделей диктора) $2 \leq N \leq 50$ позволяет установить:

– фильтрация увеличивает вероятность распознавания в среднем на $3,9 \pm 1,81\%$ ($p < 0,05$) при $10 \leq N \leq 50$ по сравнению с полным множеством признаков;

– повышения вероятности распознавания возможно добиться обязательной фильтрацией обучающей выборки, в то время как фильтрация тестовых векторов признаков дает несущественный прирост (за исключением шипящих и глухих ШФК).

Сравнение эффективности комплексного классификатора с решающим правилом, построенным алгоритмом бустинга на множестве тех же классификаторов показало превосходство бустинга по доле верно идентифицированных дикторов в среднем на $1,1 \pm 0,6\%$ ($p < 0,05$) при $10 \leq N \leq 50$.

Выводы

Предложено использование критерия фильтрации признаков для отсева векторов акустических признаков применительно к задаче текстонезависимой идентификации диктора, включающей предварительную сегментацию по широким фонетическим классам (ШФК). Показано, что фильтрация увеличивает вероятность распознавания в среднем на $3,9 \pm 1,81\%$ ($p < 0,05$) при $10 \leq N \leq 50$ по сравнению с полным множеством признаков.

Получил дальнейшее развитие робастный метод классификации на основе гауссовых смесей. Его применения с возрастанием количества моделей дикторов показывает значительное сокращение объема вычислений по сравнению с решающим правилом на основе бустинг-алгоритма.

Список литературы

1. Клименко, Н.С. Исследование эффективности бустинга в задаче текстонезависимой идентификации диктора / Н.С. Клименко, И.Г. Герасимов // Искусственный интеллект. – 2014. – №4(66). – С. 191-201.
2. Isabelle Guyon. An introduction to variable and feature selection / Isabelle Guyon, Andr e Elisseeff // Journal of Machine Learning Research 3 (2003): P. 1157-1182. 2003.
3. Molina L.C., Belanche L., Nebot A. Feature Selection Algorithms: A Survey And Experimental Evaluation // Proceedings of the 2002 IEEE International Conference on Data Mining, IEEE Computer Society: P. 306-313. 2002.
4. Peng H. C., Long F., Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy // IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): P. 1226-1238. 2005.
5. Сорокин В.Н. Верификация диктора по спектрально-временным параметрам речевого сигнала / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – Т. 10, № 2. – С. 87-104.
6. Садыхов Р.Х. Модели гауссовых смесей для верификации диктора по произвольной речи / Р.Х. Садыхов, В.В. Ракуш // Доклады БГУИР. – 2003. – №4. – С. 95-103.
7. Downloads - voxforge.org [Электронный ресурс]. – Режим доступа: <http://voxforge.org/ru/downloads>. – Дата доступа: 12.02.2016.