

ОПТИМАЛЬНЫЙ НАБОР ИНФОРМАТИВНЫХ ПРИЗНАКОВ НА ОСНОВЕ АНАЛИЗА АКУСТИЧЕСКИХ СИГНАЛОВ

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

Трус И.И., Бурак А.И.

Давыдов И.Г. – к.т.н., доцент

Сердечно-сосудистые заболевания в настоящее время являются одной из самых распространённых причин инвалидности и ранней смертности населения. Во избежание развития стойких нарушений и осложнений, приводящих к столь печальным последствиям, принципиально важным является обнаружение патологий на ранних стадиях. Фонокардиограмма – это наиболее простой и доступный метод диагностики состояния сердца, основанный на регистрации звуковых колебаний, сопровождающих сердечные сокращения, с помощью микрофона и последующем их анализе. В ходе построения автоматизированных систем анализа фонокардиограммы критически важным этапом является определение набора признаков, позволяющих в наиболее полной мере дифференцировать фонокардиограммы больных и здоровых людей. В работе рассматриваются методы для осуществления отбора признаков.

Существует большое число всевозможных характеристик сигнала в частотной, временной, частотно-временной областях, которые могут использоваться в качестве информативных признаков для осуществления классификации. Тем не менее, необходимо дополнительное исследование применимости признаков для каждой конкретной задачи, так как зачастую многие признаки не обладают необходимой дифференцирующей способностью. Кроме того, использование большого набора информативных признаков ведет к увеличению вычислительной сложности системы.

Задача отбора оптимального набора признаков заключается в том, чтобы выбрать такое подмножество признаков из исходного набора признаков, что точность классификатора, обученного на этом подмножестве признаков, будет максимальной (по всем подмножествам исходного множества признаков).

Существуют 3 группы методов выбора информативных признаков:

- 1) методы-фильтры;
- 2) методы-обёртки;
- 3) встроенные методы [1].

Методы-фильтры не требуют взаимодействия с алгоритмом обучения. Они осуществляют отбор оптимального набора признаков, действуя только информацией, полученную из обучающей выборки. Методы-фильтры выполняются на этапе предварительной обработки, до выполнения алгоритма обучения. Методы-фильтры рассматривают каждую переменную независимо и, в некоторой степени, изолированно, оценивая ее по тому или иному показателю (информационные или статистические критерии, минимальная избыточность при максимальной релевантности MRMR и др.). Среди рассматриваемых подходов методы-фильтры обладают наименьшей вычислительной сложностью. Также они масштабируемы и просты в применении.

Методы-обёртки осуществляют оценку информативности конкретных подмножеств признаков, используя алгоритм обучения.

Принцип методов обёртки состоит в следующем:

- 1) выполняется поиск по пространству подмножеств исходного множества признаков (как правило, в задачах классификации акустических сигналов невозможно выполнить прямой перебор всевозможных подмножеств исходного набора признаков, поэтому используют различные алгоритмы поиска);
- 2) для каждого шага поиска используется информация о качестве обучения на текущем подмножестве признаков.

Итак, для применения методов-обертки необходимо, во-первых, выбрать стратегию поиска по пространству подмножеств исходного множества признаков, и, во-вторых, выбрать функцию оценки качества обучения на текущем подмножестве признаков [2].

Встроенные методы выполняют выбор подмножества признаков в качестве одного из этапов обучения, и поэтому специфичны для конкретной модели. Достоинства этих методов заключаются в следующем:

- 1) наилучшим образом приспособлены к конкретной модели;
- 2) отсутствует необходимость выделять специальную тестовую выборку, на которой тестируется порог функции ранга для методов-фильтров или выполняется поиск наилучшего подмножества для методов-обертки;
- 3) как следствие из предыдущего пункта, при использовании этого метода вероятность переобучения классификатора значительно меньше.

Рассмотренные группы методов применимы для формирования оптимального набора информативных признаков для классификации акустических сигналов.

Список использованных источников:

1. Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. [Электронный ресурс]. – Режим доступа: <https://ranalytics.github.io/data-mining>. – Дата доступа: 20.03.2019.
2. Ron Kohavi, George H. John. Wrappers for feature selection // Artificial Intelligence 97. – 1997. – P. 273–324.