

Окулич В. К.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Плотников Ф. В.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Шилин В. Е.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Кабанова А. А.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Бабака Н. К.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Какойченкова А. К.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

Пинчук А. Н.

*Витебский государственный ордена Дружбы народов медицинский университет,
Беларусь*

ОБОСНОВАНИЕ ВЫБОРА ТЕСТОВ ДЛЯ ИДЕНТИФИКАЦИИ СТРЕПТОКОККОВ

Идентификация — отнесение исследуемого образца на основе определенного правила к какому-либо классу при исследовании его характеристик. Эти характеристики являются описаниями классов.

В данном случае классы — это 44 микроба рода *Streptococcus*, характеристики — 24 субстрата. В качестве правила отнесения к какому-либо классу используется расчет и выбор наименьшего Евклидова расстояния между исследуемым образцом и этими классами в 24-мерном пространстве [1, 2].

Цель: выбор комбинации 24 субстратов из таблицы Берджи, позволяющих наилучшим образом проводить идентификацию. Выбор необходимо провести объективно: с использованием математического аппарата. В качестве объективного критерия выбираем информативность. Информативность субстрата должна показать, насколько он характеризует состояние класса и всей выборки в целом, то есть насколько от него зависит принятие решение при идентификации.

Методы исследования. В теории информации существуют различные методы расчета информативности: метод Шеннона, Кульбака, метод накопленных частот (МНЧ), метод Фишера. Выбор метода зависит от области применения, конкретной постановки задачи, а также свойств анализируемых признаков объектов: способа их кодировки, объема выборки, количества градаций признаков. В медицине большой популярностью пользуются методы Кульбака, Шеннона, МНЧ [3].

Информативность, определяемая всеми тремя методами — величина положительная, однако в МНЧ и методе Кульбака она не является нормированной, поэтому об информативности, определенной этими методами, можно говорить только в относительном плане: более высокая или более низкая по сравнению с информативностью другого признака. Метод Шеннона дает оценку информативности как нормированной величины, которая изменяется от 0 до 1. Поэтому об информативности признака, определенной методом Шеннона, можно говорить в абсолютном плане: ближе к 1 — высокая; ближе к 0 — низкая.

На основании вышеизложенного, а также учитывая, что у нас имеется обучающая выборка из 44 классов, в качестве метода расчета информативности выбираем метод Шеннона, так как он позволяет рассчитывать информативность для любого количества классов, а результаты являются нормализованными, т. е. лежат в пределах 0–1 и легко поддаются анализу.

Результаты и обсуждение. Метод Шеннона предлагает оценивать информативность как средневзвешенное количество информации, приходящиеся на различные градации признака. Информативность i -го признака рассчитывается по формуле:

$$I(X_i) = 1 + \sum_{i=1}^G (P_i \sum_{k=1}^K P_{i,k} * \log_k P_{i,k}), \quad (1)$$

где G — количество градаций признака; K — количество классов; P_i — вероятность i -й градации признака.

$$P_i = \frac{\sum_{k=1}^K m_{i,k}}{N}, \quad (2)$$

где $m_{i,k}$ — частота появления i -й градации в K -м классе; N — общее число наблюдений.

$$P_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^K m_{i,k}}, \quad (3)$$

где $P_{i,k}$ — вероятность появления i -й градации признака в K -м классе.

Для данного метода исследователю необходимо самому задать число градаций признака i , а также границы интервалов каждой градации. Так как метод Шеннона не зависит от числа градаций, то при выборе данных величин нужно исходить из диапазона самих выборок.

Для вычисления информативности по Шеннону был разработан алгоритм, который затем был реализован в виде отдельной функции программы идентификации.

Результаты расчетов представлены в таблице.

Информативность отдельных диагностических тестов

<i>bGAR</i>	0.706291762281806	<i>GLYG</i>	0.534687097322296
<i>GTA</i>	0.689752412250478	<i>CDEX</i>	0.523734630367105
<i>TAG</i>	0.684615328673894	<i>PUL</i>	0.522626612532188
<i>bGAL</i>	0.650169209906056	<i>ADH</i>	0.516538511445277
<i>PAL</i>	0.640059167015037	<i>TRE</i>	0.513505748041421
<i>bNAG</i>	0.633789309161167	<i>PYRA</i>	0.500866662504714
<i>MbDG</i>	0.598121125159903	<i>LAC</i>	0.454817216869662
<i>RAF</i>	0.592284744371048	<i>bMAN</i>	0.447543102656959
<i>SOR</i>	0.587734769087419	<i>LARA</i>	0.441567308945729

<i>HIP</i>	0.584763186308815	<i>bGUR</i>	0.338215034436164
<i>aGAL</i>	0.582955477826106	<i>APPA</i>	0.301741696756098
<i>RIB</i>	0.544423947997944	<i>URE</i>	0.29329742405358
<i>MAN</i>	0.544055388473566	<i>SAC</i>	0.291669476953424
<i>bGLU</i>	0.543612041209095	<i>MAL</i>	0.252704401530065
<i>MEL</i>	0.541946821518661	<i>MLZ</i>	0.250758655304904
<i>VP</i>	0.541343205119348	<i>DARL</i>	0.169421771296366

В результате полученных данных, а также с учетом коммерческих характеристик реагентов отобран список субстратов для идентификации микробов рода *Streptococcus*: L-аргинин, резорифин-βD-глюкопиранозид, резорифин-βD-галакто-пиранозид, резорифин-βD-глюкуронид, 4-нитрофенил-αD-галактопиранозид, 4-нитрофенил-βD-галактопиранозид, D-рибоза, D-маннит, D-лактоза, D-трегалоза, D-раффиноза, D-сахароза, L-арабиноза, α-циклодекстрин, натрия пируват, 2-нафтил-βD-галактопиранозид, пироглютаминат-β-нафтиламид, натрия гиппурат, D-мальтоза, пуллулан, D-мелибиоза, D-мелицитоза, метил-βD-глюкопиранозид, D-тагатоза.

Выводы:

1. Оптимальным методом для определения информативности диагностических тестов при идентификации микроорганизмов является метод Шеннона.
2. Метод Шеннона позволяет определить информативность предложенных тестов для идентификации стрептококков.
3. Отобранные с помощью метода Шеннона диагностические тесты обладают достаточной информативностью при идентификации микроорганизмов и могут быть использованы для создания диагностических тест-систем.

ЛИТЕРАТУРА

1. Годин, А. М. Статистика : учеб. / А. М. Годин. Москва : Дашков и К°, 2012. 451 с.
2. *Bergey's Manuals. Understanding bacteria* by Sheela Srivastava. Springer, 2003. P. 40.
3. Чернавский, Д. С. Синергетика и информация : динамическая теория информации / Д. С. Чернавский. Москва : КД Либроком, 2013. 304 с.