**О** **ПОСТРОЕНИИ МНОЖЕСТВА МИНИМАЛЬНЫХ ТЕСТОВ ИЛИ ОТДЕЛЬНЫХ ТЕСТОВ**

**Ю. А. Бродская**

*1Саратовский государственный университет им. Н. Г. Чернышевского, Саратов, Россия*

Благодаря успехам, достигнутым в дискретной математике, алгебраической топологии, теории вероятностей и математической статистике, получены заметные теоретические и практические результаты в технической диагностике, распознавании рудных месторождений, прогнозировании исходов процессов в производстве, в биологических системах. Заметный вклад в теорию и практику распознавания внесли: Гренандер У., Фу К., Журавлев Ю.И., Загоруйко Н.Г., Кудрявцев В.Б., Ф.Розенблат, Богомолов А.М., Закревский А.Д., Вапник В.Н., Цыпкин Я.З., Горелик А.Л., Яблонский С.В. и др. Задачи распознавания образов и диагностирования – одни из наиболее важных; получены теоретические и практические результаты в технической диагностике, распознавании рудных месторождений, прогнозировании исходов процессов в производстве, в биологических системах. Под распознаванием образа объекта (объекта) понимается исследование, целью которого является определение класса объектов (предметов, явлений, процессов)  классификации, заданной на генеральной совокупности ℳнекоторой предметной области, которому принадлежит распознаваемый объект.

Определение принадлежности распознаваемого объекта к классу объектов основывается на оценке близости его свойств (признаков) к свойствам объектов класса. Описание класса представляет собой неупорядоченный набор описаний эталонных объектов, принадлежность которых к классу известна. Описание *эталонного* объекта – строка таблицы эталонных объектов, элементы которой являются значениями признаков, описывающих объекты; множество признаков (упорядоченное), описывающее объекты – единое для всех классов. В большинстве случаев рассматриваются классификации с непересекающимися классами (разбиения): ∅. Разбиение генеральной совокупности ℳ определено не полностью: существуют объекты ℳ, принадлежность которых классам не известна. Такие объекты называются *неизвестными*. На выборке эталонных объектов ℳ определяется упорядоченный набор признаков  задаваемый экспертами предметной области. Из этого набора, если он велик, исключают признаки с малой *информативностью*, определяемые формальными способами.

Однако, следует заметить, что в подавляющем большинстве теоретических и прикладных работ по распознаванию образов основное внимание уделяется разработке общих и частных моделей для построения и совершенствования решающих правил распознавания. Значительно меньше усилий прилагается к решению проблем формирования пространств признаков, обеспечивающих оптимальные соотношения между затратами ресурсов и времени и точностью распознавания. Наиболее актуальна задача формирования таких признаковых пространств в предметных областях, где велика интенсивность потока требований на распознавание и возможны резкие колебания этой интенсивности. К таким предметным областям, в первую очередь, следует отнести медицину, микробиологию, ветеринарию, криминалистику. В этих областях часто возникает дефицит ресурсов и времени и, соответственно, снижается качество диагностирования.

Отдельные минимальные тесты можно получить, если использовать для уточнения длины минимального теста следствие из теоремы 1, приведенной в данной работе.

Представляется, что проблеме уменьшения затрат ресурсов и времени при построении минимальных тестов уделяется существенно меньше внимания, чем другим проблемам распознавания (диагностики). Тесты используются при решении задач технической диагностики, распознавания образов, кодирования и др. [1, 2, 3]. Указанная проблема может быть разбита на две подпроблемы: а) проблему построения таких тестов для задач тестового распознавания, позволяющих уменьшить затраты ресурсов (в первую очередь, дефицитных) и времени при определении значений признаков распознаваемых объектов, т.е. в процессах распознавания [4], и б) проблему построения множеств тестов с уменьшенными затратами времени на построение тестов, что позволяет увеличить размерность используемых матриц.

Данная работа посвящена проблеме уменьшения затрат вычислительных ресурсов и времени в задачах построения минимальных тестов булевых матриц. Следует обратить внимание на то, что на основе минимальных тестов могут быть выполнены построения неминимальных тупиковых тестов.

Пусть задана таблица (матрица) *Q*, содержащая значения из множества {0,1} (булева матрица). В этой таблице строки (и столбцы) попарно различны.

Определение 1. [Соловьев Н.А.] Множество столбцов называется безусловным (диагностическим) тестом таблицы *Q*, если в подтаблице *Q*() нет одинаковых строк.

Определение 2. [Соловьев Н.А.] Тест называется тупиковым, если никакое его собственное подмножество не есть тест.

Определение 3. Тест наименьшей длины называется минимальным. Очевидно, всякий минимальный тест – тупиковый.

Особенности предлагаемого метода построения множества минимальных тестов булевой матрицы (все элементы которой определены однозначно – неопределенных элементов нет) являются, в отличие от известных

1. Определяется длина минимального теста с помощью эвристического метода, возможно, с погрешностью, зависящей от размерности (m×n) матрицы и степенью ее разряженности. Под этой степенью здесь понимается абсолютная величина разности между количеством единиц и нулей в столбцах матрицы (всех или некоторых).
2. Производится построение тестов с вычисленной длиной минимального теста: значение этой длины (при обнаружении погрешности) корректируется, опираясь на эвристические и точные способы проверки.

Предполагается, что большие затраты на построение минимальных тестов возникают в связи с использованием точных методов при построении тестов и отсутствием проверки.

Любой тупиковый тест булевой матрицы (в том числе, минимальный) содержит хотя бы один элемент (столбец), разделяющий (различающий) каждую пару строк [1]. Существуют (могут существовать) пары строк, разделяемые только одним столбцом. Очевидно, такие столбцы обязательно должны входить в любой тест матрицы: нетупиковый, тупиковый, минимальный. Как показано в [1], целесообразно разбить множество строк матрицы на подмножества-классы с помощью обязательных элементов. В каждом классе содержатся строки, имеющие одинаковые значения обязательных элементов тестов. В разных классах все (или некоторые) обязательные элементы содержат разные значения. Такое разбиение уменьшает количество пар строк, которые должны анализироваться при построении тестов. Легко показать, что пары строк, различаемых одним обязательным элементом тестов, принадлежат множеству пар строк таких, что строки различаются по количеству единиц в них ровно на одну единицу. Предварительный анализ показал, что количество пар строк, различающихся по количеству единиц в строках ровно на единицу в неразряженных матрицах составляет ≈31%.

Чтобы уменьшить затраты времени при поиске обязательных элементов тестов необходимо: а) определить количество единиц в каждой строке; б) упорядочить (отсортировать) матрицу по возрастанию значений двоичных чисел, отображаемых в строках; в) выявить пары строк, различающихся количеством единиц на одну единицу; г) выявить из пар строк, выявленных в п. «в», пары строк, различаемых (каждая) одним столбцом не менее (2-3 пары); д) произвести разбиение множества строк матрицы на подматрицы-классы; е) в сформированных подматрицах-классах произвести поиск остальных обязательных элементов; ж) если найдены дополнительные обязательные элементы в п. «е», произвести разбиение ранее созданных подматриц с помощью новых обязательных элементов.

Краткое описание метода множества минимальных тестов булевой матрицы

1. Сортировка строк (если тесты столбцовые) или столбцов (если тесты строковые) – по возрастанию значений двоичных чисел в строках (столбцах).
2. Поиск обязательных элементов тестов.
3. Разбиение булевой матрицы на подматрицы-классы с помощью обязательных элементов по п.2 (если они выявлены).
4. Определение длины минимальных тестов с помощью эвристической формулы на матрице и подматрицах, а также анализа пар строк в подматрицах в соответствии с теоремами.
5. Построение минимальных тестов а) выявление множеств столбцов, мощностью , не являющихся тестами; б) определение количества минимальных тестов и составление списка минимальных тестов; в) проверка результатов и, при необходимости, корректировка.

Примечание. Эвристическая формула определения длины минимального теста описана в [5] с использованием некоторых корректировок.

Теорема 1. Если в булевой матрице существует набор из *k* столбцов, образующий хотя бы одну пару одинаковых подстрок, то такой набор из *t* столбцов тестом матрицы не является.

Доказательство. Набор из *k* столбцов булевой матрицы, который образует хотя бы одну пару одинаковых подстрок, тестом не является, т.к. такой набор столбцов не соответствует определению теста (определение 1). □

Следствие 1. Если в булевой матрице все наборы из *k* столбцов образуют (каждый из них) не менее одной пары одинаковых подстрок, то в этой матрице длина минимального теста и, если, к тому же, существует хотя бы один тест с длиной , то длина минимального теста в матрице .

Доказательство. Если выполняются условия в следствии, то в матрице в соответствии с теоремой 1, не существует ни одного теста длиной и тест T: -есть минимальный тест. □

Теорема 2. Любые столбцов булевой матрицы, которые образуют одинаковых строк, не могут принадлежать одному тесту с длиной , построенному из k этих столбцов и любого другого столбца матрицы.

Доказательство. Для того, чтобы в булевой матрице различить одинаковых подстрок, необходимо, чтобы эти подстроки были попарно различимы. Число *B* пар подстрок, в которых должны быть различимы *p* одинаковых подстрок: .

Количество пар элементов с различными значениями элементов в паре, составленных из *p* элементов одного столбца, обеспечивается при , если *p* – четное число; при (или ), если *p* - нечетное число, где - количество единиц (нулей) в столбце. Максимальное количество пар различных по значению элементов в одном столбце, взятых из *p* элементов: , где: - целая часть действительного числа *R*. Минимальное количество пар одинаковых подстрок, которых останутся после добавления (*k*+1)-го столбца к набору из *k* столбцов: . Так, при *p*=3: ; при *p*=4: ; при *p*=5: .

Из выше изложенного можно сделать вывод: если в булевой матрице существуют одинаковых подстрок с длиной (где *n* – число столбцов в матрице), то все эти подстроки нельзя сделать попарно различными с помощью одного столбца: останется хотя бы одна пара одинаковых подстрок. □

Справедливо также следующее утверждение. Тупиковому тесту булевой матрицы не могут принадлежать два столбца, связанные отношением взаимно однозначного соответствия. Действительно, предположим, что в тесте булевой матрицы имеются два столбца, связанные отношением взаимно однозначного соответствия. Такие столбцы содержат непременно: а) одинаковые двухэлементные подстроки матрицы, точно соответствующие одинаковым элементам в каждом из этих столбцов; б) различающиеся двухэлементные подстроки, точно соответствующие разным элементам в каждом из этих столбцов. Два таких столбца разделяют (различают) строки в одних и тех же парах строк в булевой матрице и, следовательно, один из этих столбцов является излишним для тупикового теста.

Теорема 3. Для каждого столбца , принадлежащего тупиковому тесту булевой матрицы, существует, хотя бы одна пара строк : , различаемых элементами столбца и никакого другого столбца , принадлежащего этому же тупиковому тесту :

.

Доказательство. Предположим, что утверждение, сформулированное в теореме ложно и верно следующее утверждение:

.

Выполнив преобразования этого утверждения, получим:

Полученное утверждение противоречит определениям тупикового теста и теста булевой матрицы. Следовательно, утверждение, сформулированное в данной теореме истинно. □

Теорема 4. Тест *T* булевой матрицы, имеющей длину является минимальным тестом тогда и только тогда, когда все тесты с длиной данной матрицы являются тупиковыми.

, (1)

Где *X* – множество столбцов матрицы *Q*; - тупиковый тест; - минимальный тест;

- длина минимального теста.

Доказательство. 1) Достаточность. Пусть условие в левой части утверждения (1) выполняется, а условие в правой части – не выполняется. Возможны следующие варианты:

а) , что противоречит определению 3.

б) – противоречит левой части (1) в теореме. Следовательно, утверждение в теореме верно.

Необходимость. Пусть условие в левой части утверждения (1) не выполняется то есть существует хотя бы один тест : , не являющийся тупиковым, и существует, следовательно, тупиковый тест ; , что противоречит утверждению теоремы. □

СПИСОК ЛИТЕРАТУРЫ

1.*Чегис И.А., Яблонский С.В.* Логические способы контроля электрических схем.// Труды математического института им. В.А. Стеклова. 1958. Т.51.- С.226-269.

2. *Соловьев Н.А.* Тесты: (теория, построение, применение). Новосибирск: Наука, 1978.- 189 с.

3. *Кудрявцев В. Б., Андреев А.Е., Гасанов Э.Э.* Теория тестового распознавания // М: изд-во «Физматлит», 2007, 320 с.

Т. 1. С. 172–176.

4. *Бродская Ю.А.* Методы распознавания объектов с заданными ограничениями. Автореферат диссертации на соискание ученой степени канд. физ.-мат. наук. – Саратов, 2002. – 16 с.

5. *Бродская Ю.А.* Формирование кратчайших покрытий небулевых матриц при решении задач диагностики / Автоматизация проектирования дискретных систем./ Материалы шестой международной конференции. – Минск, 2007, т.1, С. 138-145.