

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
Санкт-Петербургский государственный университет
аэрокосмического приборостроения

КАФЕДРА ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ И СЕТЕЙ

РАСПОЗНАВАНИЕ
В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

Методические указания

Часть II

Санкт-Петербург
2013

Составитель: доцент кафедры вычислительных систем и сетей
 кандидат технических наук Н.В. Соловьев
Рецензент: профессор кафедры аэрокосмических компьютерных
 технологий, кандидат технических наук Н.А. Шехунова

Содержатся методические указания по второй части цикла лабораторных работ, выполняемых студентами при изучении дисциплины «Интеллектуальные системы». Описан процесс построения экспертной систем с применением условных вероятностей – алгоритм работы системы, разработка интерфейса и подготовка информации для базы данных, а также подробно рассмотрен тестовый пример работы экспертной системы.

Предназначено для студентов, обучающихся по направлению 230100.68 «Информатика и вычислительная техника».

Подготовлено к изданию кафедрой вычислительных систем и сетей по рекомендации редакционно-издательского совета Санкт-Петербургского государственного университета аэрокосмического приборостроения.

Содержание

Введение	3
1 Разработка структуры базы данных экспертной системы	5
2 Алгоритм работы экспертной системы	7
3 Диалоговый интерфейс экспертной системы	11
4 Заполнение базы данных экспертной системы	12
5 Тестирование экспертной системы	13
6 Задание к лабораторным работам	18
Рекомендуемая литература	20

Введение

Большинство задач, для решения которых используются системы искусственного интеллекта (СИИ), можно разделить на две группы.

Первая группа – поиск последовательности действий, которые приведут от начального состояния к цели. К ним относятся стратегические игры, поиск оптимального маршрута, планирование операций и т.п. Теоретическую основу СИИ, предназначенных для решения таких задач, составляет теория графов.

Вторая группа – отнесение объекта к одному из известных классов. К ним относятся экспертные системы в таких областях как медицина, техника, геология, химия, генетика, прогнозирование погоды и т.п., а также системы распознавания, идентификации и контроля различных объектов. Теоретическую основу СИИ, предназначенных для решения таких задач, составляет теория распознавания образов.

Одно из основных отличий систем распознавания от экспертных систем состоит в том, что в системе распознавания, как правило, распознаваемый образ представляется вектором признаков, в котором перед началом распознавания все значения элементов известны. В экспертных системах процесс распознавания совмещен с процессом получения информации об объекте, причем последовательность получения значений признаков меняется в зависимости от полученных ранее данных. Такой подход к распознаванию применяется при большом количестве признаков и классов, иерархической структуре классов и отсутствии в описании каждого класса всех признаков. Кроме этого, в экспертных системах часто требуются большие затраты на получение значений признаков. Например, в медицинской диагностике полное обследование пациента может включать в себя длительные и дорогостоящие анализы, проведение которых требуется далеко не при всех диагнозах.

В дальнейшем мы ограничимся рассмотрением одного из подходов к построению экспертных систем, в котором информация о распознаваемых объектах носит вероятностный характер.

Цель экспертной системы – на основании полученных данных о распознаваемом объекте сделать вывод о его принадлежности какому-то известному классу. В зависимости от области применения система может на основании симптомов поставить диагноз больному, по значениям параметров найти неисправность в устройстве, по геологическим характеристикам местности предсказать наличие полезных ископаемых, по психофизическим характеристикам человека подобрать подходящую сферу деятельности и т.д.

Рассматриваемая далее экспертная система состоит из двух достаточно независимых частей – программы-оболочки для хранения и обработки информации и базы данных с необходимыми для работы системы сведениями о конкретной области знаний. Наша задача – построение достаточно универсальной экспертной системы, способной диагностировать объект в некоторой области знаний после заполнения базы данных системы соответствующей информацией. Система должна последовательно задавать вопросы о состоянии объекта, причем количество и порядок вопросов могут меняться в зависимости от полученных ранее ответов, и после анализа имеющихся данных система должна принять решение, т.е. отнести объект к одному из известных системе классов.

Процесс создания экспертной системы включает в себя следующие этапы:

- разработка структуры базы данных;
- разработка алгоритмов выбора последовательности запросов и принятия решения;
- разработка диалогового интерфейса;
- заполнение базы данных;
- тестирование.

В процессе тестирования проверяется реакция системы при введении информации о заранее известных объектах. Если интерфейс экспертной системы соответствует требованиям пользователя и тестирование системы дает удовлетворительные результаты, то система принимается в опытную эксплуатацию. Во время опытной эксплуатации экспертной системы информация, заложенная в базу данных, корректируется с целью получения адекватных ответов.

1. Разработка структуры базы данных экспертной системы

При построении экспертной системы одной из основных проблем является формализация представления знаний о связях между признаками (в дальнейшем будем называть их симптомами) и классами образов (в дальнейшем будем называть их диагнозами). Заметим, что часто эти связи носят вероятностный характер, т.е. одно и то же значение симптома может с разной вероятностью соответствовать нескольким диагнозам. Например, в медицинской сфере боль в горле является симптомом, как ангины, так и наличия инородного тела, или в сфере технической диагностики отсутствие света фар у автомобиля может быть следствием, как севшего аккумулятора, так и неисправности проводки.

Проблему представляет и количественное описание симптомов. Например, если с температурой, весом, ростом или давлением проблем на первый взгляд не будет, то такие характеристики как уровень головной боли или износ покрышек количественно оценить труднее. В таком случае возможно задание нескольких диапазонов. Например, при вопросе об уровне боли ответ может выбираться из трех вариантов – легкая боль, средняя, острая. Симптомы могут быть и бинарными, т.е. иметь только два возможных значения – да или нет, например, наличие света фар.

Рассмотрим один из возможных способов представления в экспертной системе знаний о конкретной области ее применения.

Симптомы X_j ($j = 1 \dots n$), где n – число симптомов, удобно представить в виде списка вопросов, которые система будет задавать в процессе работы, т.е. всего один вопрос на каждый симптом. Например, «Ваш вес?», «Уровень боли в горле», «Включаются ли фары автомобиля?» и т.п.

Диагнозы W_i ($i = 1 \dots k$), где k – число диагнозов, можно представить в виде списка со строками разной длины. Каждая строка списка описывает один диагноз и состоит из следующих полей:

- название диагноза W_i ;
- априорная вероятность диагноза $p_a(W_i)$;

- связь данного диагноза с симптомами в виде массива из трех элементов – номер строки симптома X_j , условная вероятность наличия данного симптома при данном диагнозе $p(X_j/W_i)$, условная вероятность наличия данного симптома при любом другом диагнозе $p(X_j/\text{no } W_i)$.

Под априорной вероятностью диагноза понимается вероятность наличия данного диагноза у объекта при отсутствии информации о симптомах. Например, у людей ангина встречается чаще, чем инфаркт, а у автомобилей севший аккумулятор встречается чаще, чем обрыв в проводке.

Заметим, что сумма последних двух вероятностей не равна единице. Например, при гриппе головная боль присутствует в 80%, а в 30% случаев присутствие головной боли вызвано другим диагнозом. Сумма априорных вероятностей по всем диагнозам должна быть меньше единицы, т.к. показывает вероятность наличия у объекта какого-либо диагноза (неисправности, болезни). Последнее справедливо только при наличии у одного объекта одного диагноза.

Между симптомами не должно быть корреляции, т.е. они должны быть статистически независимы. Только при соблюдении этого условия можно применять приведенные далее формулы для вычисления условных вероятностей диагнозов. Два симптома X_1 и X_2 статистически независимы, если для i -го диагноза условная вероятность $p(X_1 \& X_2 / W_i) = p(X_1 / W_i) * p(X_2 / W_i)$. Например, пусть из 100 автомобилей у 20-ти не горят фары, а у 10-ти не заводится двигатель. Эти события можно считать статистически независимыми, если у 2-х автомобилей одновременно не горят фары и не заводится двигатель ($20/100 * 10/100 = 2/100$). При выборе симптомов необходимо проверять их статистическую независимость. Например, существует корреляция между симптомами «наличие озноба» и «высокая температура». Использование таких симптомов делает некорректным использование статистических формул для вычисления условных вероятностей диагнозов, связанных с этими симптомами.

При необходимости можно дополнить базу данных и списком рекомендаций по каждому диагнозу, которые будут выдаваться пользователю после принятия системой решения о наиболее вероятном диагнозе.

2. Алгоритм работы экспертной системы

Алгоритм работы экспертной системы является циклическим. В каждом цикле анализируется один симптом (напомним – в нашей структуре базы данных имеется один вопрос на каждый симптом). Для этого выполняется следующая последовательность шагов:

- выбирается наиболее информативный симптом и задается соответствующий ему вопрос;
- полученный ответ используется для вычисления вероятностей диагнозов, в которых он учитывается, а вопрос (симптом) исключается из дальнейшего рассмотрения;
- по результатам вычисления некоторые диагнозы могут исключаться;
- если вероятность какого-то диагноза достаточна для принятия решения, то система заканчивает работу, если нет – то цикл повторяется.

Как уже отмечалось, особенностью экспертных систем является то, что не все признаки (симптомы) используются в описании каждого класса (диагноза). При таком подходе можно считать наиболее информативным вопрос о симптоме, ответ на который изменит вероятность максимального числа диагнозов. Вычислить этот показатель довольно легко – для каждого еще не заданного вопроса определяется количество еще не исключенных диагнозов, в которых он учитывается. Вопрос с максимальным значением считается самым информативным в текущем цикле.

В зависимости от вида признака (симптома) ответ на вопрос может иметь разный вид (целочисленный, перечисляемый, бинарный), но желательно перед дальнейшей обработкой привести его к универсальному виду. Рассмотрим один из вариантов такого приведения.

Будем считать, что ответ на j -й вопрос X_j изменяется в пределах $X_j = -5 \dots +5$. При этом -5 означает твердое «нет», т.е. отсутствие симптома, $+5$ – твердое «да», т.е. безусловное наличие симптома, 0 – отсутствие информации о симптоме. Промежуточные значения можно интерпретировать как степень уверенности в

наличии (положительные значения) или отсутствии (отрицательные значения) симптома.

Если для такого симптома как наличие боли в горле ответ хорошо интерпретируется и его можно сразу использовать для дальнейших вычислений, то для некоторых симптомов полученный ответ необходимо обработать. Например, для такого симптома как давление крови общепринятый ответ в форме пары целых чисел следует перевести в заданный диапазон по заранее разработанной шкале.

Другой подход к универсализации ответов заключается в том, чтобы переложить оценку уровня симптома на пользователя, задав в вопросе возможные варианты ответа. Например, варианты ответа о давлении крови могут быть: сильно пониженное (-5), пониженное (-3), нормальное (0), повышенное (+3), высокое(+5).

Получив ответ на заданный вопрос, необходимо использовать его для вычисления вероятностей диагнозов, в описании которых он присутствует. При этом подтверждать диагноз может как положительный, так и отрицательный ответ. Для вычисления условной вероятности диагноза при наличии данного симптома используется формула Байеса

$$p(W_i / X_j) = \frac{p(W_i) * p(X_j / W_i)}{p(W_i) * p(X_j / W_i) + (1 - p(W_i)) * p(X_j / noW_i)}, \quad (1)$$

где $p(W_i)$ – вероятность диагноза W_i на текущем шаге. До получения ответа на первый вопрос $p(W_i) = p_a(W_i)$, а после обработки ответа на j -й вопрос можно считать $p(W_i) = p(W_i / X_j)$. Однако подставлять в формулу (1) непосредственно значения $p(X_j / W_i)$ и $p(X_j / noW_i)$ можно только в том случае, когда ответ интерпретирован как твердое «да».

При твердом «нет»

$$p(W_i / X_j) = \frac{p(W_i) * (1 - p(X_j / W_i))}{p(W_i) * (1 - p(X_j / W_i)) + (1 - p(W_i)) * (1 - p(X_j / noW_i))} \quad (2).$$

Учесть возможность получения ответа X_j в заданном диапазоне (± 5) можно линейной аппроксимацией значения условной вероятности. В таком случае формулы (1) и (2) усложняются (X_j берется со своим знаком):

$$\begin{aligned} \text{если } X_j > 0 \quad p(W_i / X_j) &= p(W_i) \left(1 + \left(\frac{p(X_j / W_i)}{p(X_j)} - 1 \right) \frac{X_j}{5} \right); \\ \text{если } X_j < 0 \quad p(W_i / X_j) &= p(W_i) \left(1 + \left(p(X_j / W_i) - \frac{(1 - p(X_j / W_i)) * p(X_j)}{1 - p(X_j)} \right) \frac{X_j}{5} \right); \end{aligned} \quad (3)$$

если $X_j = 0$ $p(W_i / X_j) = p(W_i)$;

где $p(X_j) = p(W_i) * p(X_j / W_i) + (1 - p(W_i)) * p(X_j / \text{no}W_i)$ – вероятность положительного ответа на вопрос j -го симптома.

Если вероятность $p(W_i / X_j)$, вычисленная по (3), близка к нулю, то i -й диагноз можно исключить из дальнейшего рассмотрения. Если вероятность $p(W_i / X_j)$ близка к единице, то i -й диагноз можно считать истинным и закончить работу системы. Порог близости устанавливается заранее и его точное значение подбирается экспериментально в процессе настройки системы.

После того как будут пересчитаны вероятности всех диагнозов, в которых учитывается симптом, связанный с заданным вопросом, необходимо для каждого из них оценить возможные максимальную и минимальную вероятности. Для этого у каждого диагноза просматривается список еще не рассмотренных симптомов и по каждому из них выполняется следующая последовательность действий (при первом цикле эта последовательность выполняется для всех диагнозов).

Во-первых, необходимо определить какой ответ – положительный или отрицательный, увеличивает вероятность i -го диагноза. Если $p(X_j / W_i) > p(X_j / \text{no}W_i)$, то положительный ответ увеличивает $p(W_i / X_j)$ и при оценке возможных вероятностей диагноза расчетные условные вероятности $p'(X_j / W_i)$ и $p'(X_j / \text{no}W_i)$ не изменяются, т.е. $p'(X_j / W_i) = p(X_j / W_i)$ и $p'(X_j / \text{no}W_i) = p(X_j / \text{no}W_i)$. Если же $p(X_j / W_i) < p(X_j / \text{no}W_i)$, то $p(W_i / X_j)$ увеличивается при отрицательном ответе и для оценки следует использовать обратные значения условных вероятностей, т.е. считать

$$p'(X_j / W_i) = 1 - p(X_j / W_i), \quad p'(X_j / \text{no}W_i) = 1 - p(X_j / \text{no}W_i). \quad (4)$$

Если между симптомами нет корреляции, т.е. они статистически независимы, то максимально и минимально возможные вероятности диагноза вычисляются как произведение условных вероятностей по всем еще не заданным вопросам,

связанным с симптомами данного диагноза, при условии соответственно положительных и отрицательных ответов на них.

Во-вторых, с учетом (4) вычисляются четыре условные вероятности (m – число еще не рассмотренных симптомов i -го диагноза):

- вероятность i -го диагноза при всех положительных ответах

$$p(Wi / X_{\Pi}) = \prod_{j=1}^m p'(Xj / Wi);$$

- вероятность не i -го диагноза при всех положительных ответах

$$p(\text{no}Wi / X_{\Pi}) = \prod_{j=1}^m p'(Xj / \text{no}Wi);$$

- вероятность i -го диагноза при всех отрицательных ответах (5)

$$p(Wi / \text{no}X_{\Pi}) = \prod_{j=1}^m (1 - p'(Xj / Wi));$$

- вероятность не i -го диагноза при всех отрицательных ответах

$$p(\text{no}Wi / \text{no}X_{\Pi}) = \prod_{j=1}^m (1 - p'(Xj / \text{no}Wi)).$$

В-третьих, используя результаты вычислений (5), по формуле Байеса определяются максимально возможная вероятность i -го диагноза

$$p_{\max}(Wi) = \frac{p(Wi) * p(Wi / X_{\Pi})}{p(Wi) * p(Wi / X_{\Pi}) + (1 - p(Wi)) * p(\text{no}Wi / X_{\Pi})} \quad (6)$$

и минимально возможная вероятность i -го диагноза

$$p_{\min}(Wi) = \frac{p(Wi) * p(Wi / \text{no}X_{\Pi})}{p(Wi) * p(Wi / \text{no}X_{\Pi}) + (1 - p(Wi)) * p(\text{no}Wi / \text{no}X_{\Pi})} \quad (7).$$

Если $p_{\max}(Wi) < p(Wi)$, то i -й диагноз можно исключить из дальнейшего рассмотрения, т.к. его максимально возможная вероятность при всех положительных симптомах меньше вероятности наличия данного диагноза у объекта до получения каких-либо данных о еще не рассмотренных симптомах.

После оценки возможных максимальных и минимальных вероятностей всех еще не исключенных из рассмотрения диагнозов выполняется поиск наиболее вероятного диагноза. Для этого находится диагноз d , у которого минимально воз-

можная вероятность максимальна, т.е. $p_{\min}(Wd) \geq p_{\min}(Wi)$ для всех $i = 1 \dots k$. Если минимально возможная вероятность этого диагноза больше максимально возможной вероятности любого другого диагноза, т.е. $p_{\min}(Wd) \geq p_{\max}(Wi)$ для всех $i = 1 \dots k, i \neq d$, то диагноз d считается наиболее вероятным и сообщается пользователю. Если такой диагноз не находится, то цикл повторяется.

Процесс выполняется до тех пор, пока не будет найден наиболее вероятный диагноз, или не рассмотрены все симптомы, или не исключены все диагнозы. В двух последних случаях система не принимает никакого решения, но может выдавать пользователю вероятности диагнозов после последнего цикла.

3. Диалоговый интерфейс экспертной системы

Минимальный диалоговый интерфейс экспертной системы должен обеспечивать возможность задавать пользователю вопросы, получать от него ответы в требуемой форме и сообщать результат работы – наиболее вероятный диагноз и соответствующую рекомендацию, если она имеется. Полезной для пользователя будет информация о количестве уже заданных и еще оставшихся вопросов, число и наименование исключенных из рассмотрения диагнозов, а по окончании работы системы – наименования ближайших диагнозов, со значениями их вероятностей.

Желательно давать пользователю подсказки о возможных вариантах ответов. Например, «Курите ли Вы? Для ответа введите число выкуриваемых Вами в день сигарет деленное на 5 (максимальное число +5, -5 означает, что Вы не курите, 0 означает, что вы курите нерегулярно)».

Для анализа работы системы и внесения необходимых изменений в алгоритм и базу данных следует обеспечить в отладочном режиме на каждом шаге возможность просмотра результатов вычислений и их сохранения при необходимости, а также фиксации последовательности заданных вопросов и полученных ответов. Эта информация облегчает выявление причин неадекватной реакции системы, внесение изменений в базу данных и повторное прохождение последовательности вопросов и ответов.

Для облегчения ввода априорных вероятностей диагнозов и условных вероятностей симптомов можно предусмотреть ввод только начальных статистических данных, переложив на систему вычисления соответствующих вероятностей и проверку корректности введенных данных. Например, вводится общее число обследованных объектов, число объектов имеющих данный диагноз и число объектов имеющих данный симптом при заданном диагнозе. Из этой информации по известным статистическим зависимостям можно вычислить априорные вероятности диагнозов и условные вероятности, связывающие симптомы и диагнозы, а также проверить корректность исходных данных.

4. Заполнение базы данных экспертной системы

Заполнение базы данных, пожалуй, наиболее ответственный, сложный и трудоемкий этап создания экспертной системы. От его результатов в основном зависит правильность работы экспертной системы. Основные трудности заключаются в сборе и обработке статистической информации. Например, на первый взгляд, при подготовке экспертной системы в области медицины достаточно взять справочник по болезням и выписать из него диагнозы с соответствующими симптомами. Однако вряд ли Вы найдете в справочнике данные об априорных вероятностях заболеваний и тем более условные вероятности симптомов по каждому диагнозу.

При подготовке вопросов по симптомам наибольшую сложность представляет требование статистической независимости симптомов. Именно это требование заложено в основу оценки возможных вероятностей диагнозов (6) и (7). Два события независимы, если вероятность их совместного наступления равна произведению вероятностей наступления каждого события в отдельности. Собрав статистическую информацию, как по отдельным симптомам, так и по их сочетанию, можно оценить их независимость, например, как показано в разделе 1. Если между симптомами есть корреляция, то приходится задавать условные вероятности всех возможных сочетаний симптомов для каждого диагноза, что резко увеличивает объем базы данных.

Важно правильно сформулировать связанный с симптомом вопрос, а также дать возможные варианты ответов. Бывает полезно представить последовательность вопросов, в которой каждый последующий вопрос уточняют предыдущий. Например, первый вопрос – «Бывают ли у Вас боли в груди?» можно дополнить вопросами о частоте, месте и уровне болей. При этом не следует забывать, что каждый вопрос должен быть связан хотя бы с одним диагнозом.

Формирование списка, связывающего диагноз с симптомами, с указанием условных вероятностей также представляет собой серьезную проблему. Для каждого диагноза необходимо максимально полно учесть связанные с ним симптомы. Не следует забывать, что на повышение вероятности диагноза влияют как положительный, так и отрицательный ответы. Таким образом, для одного и того же вопроса положительный ответ может быть связан с одним диагнозом, а отрицательный – с другим.

Можно предложить следующий порядок подготовки базы данных системы – на первом этапе формируется список диагнозов с соответствующими им симптомами, а по нему составляется список симптомов с соответствующими вопросами. На втором этапе собирается необходимая статистическая информация, что требует изучения имеющихся источников и возможно проведения специальных исследований.

5. Тестирование экспертной системы

Тестирование экспертной системы производится для обнаружения недостатков и внесения необходимых изменений. На первом этапе проверяется работоспособность системы, т.е. корректность работы алгоритма и удобство интерфейса, на втором – полнота и непротиворечивость информации, заложенной в базу данных.

Для проверки работоспособности системы база данных заполняется тестовой информацией минимально необходимого объема. Данные подготавливаются таким образом, чтобы можно было на каждом шаге следить за процессом решения и проверить реакцию системы в критических ситуациях.

Правильность заполнения базы данных проверить значительно сложнее, т.к. количество возможных сочетаний симптомов не позволяет проверить реакцию системы на них в полном объеме. Например, пусть имеется всего 100 симптомов и каждый может принимать 10 значений. Тогда число возможных комбинаций симптомов с разными значениями составляет 10^{100} . Конечно реальное число комбинаций несколько меньше этой оценки, но в целом порядок сохраняется.

Разработчики экспертных систем вынуждены очень тщательно готовить исходные данные и проверять правильность их внесения в систему. После проверки некоторого количества комбинаций симптомов с заранее известным результатом система переводится в опытную эксплуатацию, в процессе которой могут продолжаться выявляться некорректные результаты.

Рассмотрим пример работы системы, построенной по изложенному выше алгоритму. В качестве исходных данных возьмем не связанную с конкретной областью абстрактную информацию, приведенную в таблице 1.

Таблица 1 – Исходные значения $p(X_j/W_i)$ и $p(X_j/\text{no}W_i)$

	Диагноз 1	Диагноз 2	Диагноз 3
$p_a(W_i)$	0,3	0,2	0,1
Симптом 1	0,6/0,01	-	0,8/0,03
Симптом 2	0,7/0,02	0,8/0,01	0,02/0,1
Симптом 3	0,1/0,2	0,9/0,02	0,9/0,01
Симптом 4	-	0,01/0,07	0,7/0,04

В нашей базе данных всего четыре симптома или признака ($j = 1 \dots 4$) и три возможных диагноза или класса ($i = 1 \dots 3$). Первый диагноз связан с первыми тремя симптомами, второй – с последними тремя, а третий – со всеми четырьмя симптомами. Напомним, что $p_a(W_i)$ – априорная вероятность диагноза W_i , $p(X_j/W_i)$ – условная вероятность наличия симптома X_j при диагнозе W_i , $p(X_j/\text{no}W_i)$ – условная вероятность наличия симптома X_j при отсутствии диагноза W_i .

Наша информация не является статистически полной, т.к. сумма априорных вероятностей всех диагнозов меньше единицы и для некоторых комбинаций «диа-

гноз – симптом» информация отсутствует. Например, неизвестна статистика по первому симптому при первом диагнозе, что неэквивалентно отсутствию первого симптома при первом диагнозе, т.к. такая информация (отсутствие симптома) подтверждала бы первый диагноз. В нашем случае для этой пары отсутствуют статистические данные, что не мешает разработке экспертной системы, но может сказаться на результатах ее работы. В случае появления недостающей статистики она может быть добавлена в базу данных, и после переобучения результаты работы, скорее всего, улучшатся.

До начала первого цикла «вопрос-ответ» текущая вероятность каждого диагноза принимаем равной его априорной вероятности $p(W_i) = p_a(W_i)$. Далее определим максимально и минимально возможные вероятности диагнозов.

Для этого при условии независимости симптомов вычислим по (5) условные вероятности для диагноза W_1 (в данный момент еще не задано ни одного вопроса и ни один диагноз не исключен из рассмотрения):

$$p(W_1/X_{II}) = 0,6*0,7*0,9 = 0,378; p(\text{no } W_1/X_{II}) = 0,01*0,02*0,8 = 0,00016;$$

$$p(W_1/\text{no } X_{II}) = 0,4*0,3*0,1 = 0,012; p(\text{no } W_1/\text{no } X_{II}) = 0,99*0,98*0,2 = 0,19404.$$

Заметим, что у третьего симптома согласно (4) используются обратные условные вероятности $p'(X_3/W_1)$ и $p'(X_3/\text{no } W_1)$, т.к. $p(X_3/W_1) < p(X_3/\text{no } W_1)$.

Далее согласно (6) и (7) найдем:

$$p_{\max}(W_1) = (0,3*0,378)/(0,3*0,378+0,7*0,00016) = 0,999;$$

$$p_{\min}(W_1) = (0,3*0,012)/(0,3*0,012+0,7*0,19404) = 0,02582.$$

Аналогично вычисляются p_{\max} и p_{\min} для двух других диагнозов. Результаты расчетов приведены в таблице 2.

При выборе наиболее информативного симптома оказывается, что второй и третий симптомы присутствуют в описании всех диагнозов, т.е. задаваемый вопрос приходится выбирать случайным образом. Пусть это будет вопрос о втором симптоме и на него получен ответ «5», т.е. $X_2 = 5$.

Исключим второй симптом из списка и вычислим согласно (1), т.к. ответ – твердое «да», текущие значения вероятностей тех диагнозов, в описании которых он входит:

$$p(W_1/X_2) = (0,3*0,7)/(0,3*0,7+0,7*0,02) = 0,9375;$$

$$p(W_2/X_2) = (0,2*0,8)/(0,2*0,8+0,8*0,01) = 0,95238;$$

$$p(W_3/X_2) = (0,1*0,02)/(0,1*0,02+0,9*0,1) = 0,02174.$$

После обработки ответа на второй вопрос наибольшая вероятность у второго диагноза, но если принять порог достоверности 0,99, то вероятность $p(W_2)=0,95238$ недостаточна для принятия решения.

Таблица 2 – Начальные значения p_{\max} и p_{\min}

	Диагноз 1	Диагноз 2	Диагноз 3
$p(W_i/X_{II})$	0,378	0,7128	0,49392
$p(\text{no}W_i/X_{II})$	0,00016	0,00019	0,00119
$p(W_i/\text{no}X_{II})$	0,012	0,0002	0,00012
$p(\text{no}W_i/\text{no}X_{II})$	0,19404	0,06791	0,09219
$p_{\max}(W_i)$	0,99901	0,99893	0,97878
$p_{\min}(W_i)$	0,02582	0,00074	0,00014

Повторим вычисление p_{\max} и p_{\min} для всех диагнозов, не забывая, что второй симптом исключен. Результаты расчетов приведены в таблице 3.

Из таблицы 3 видно, что пока ни один диагноз нельзя исключить. Максимальное значение p_{\min} имеет первый диагноз, но $p_{\max}(W_2) > p_{\min}(W_1)$, следовательно, первый диагноз пока нельзя считать наиболее вероятным.

Из оставшихся трех симптомов наиболее информативным будет третий. Пусть на него получен ответ «-5», т.е. $X_3 = -5$.

Исключим третий симптом из списка и вычислим согласно (2), т.к. ответ – твердое «нет», текущие значения вероятностей тех диагнозов, в описании которых он входит:

$$p(W_1/X_3) = (0,9375*(1-0,1))/(0,9375*(1-0,1)+(1-0,9375)*(1-0,2)) = 0,94406;$$

$$p(W_2/X_3) = (0,95238*(1-0,9))/(0,95238*(1-0,9)+(1-0,95238)*(1-0,02)) = 0,67114;$$

$$p(W_3/X_3) = (0,02174*(1-0,9))/(0,02174*(1-0,9)+(1-0,02174)*(1-0,001)) = 0,0022.$$

После обработки ответа на третий вопрос наибольшая вероятность у первого диагноза, но порог достоверности не преодолен.

Таблица 3 – Значения $p(W_i)$, p_{\max} и p_{\min} после ответа на вопрос № 2

	Диагноз 1	Диагноз 2	Диагноз 3
$p(W_i)$	0,9375	0,95238	0,02174
$p(W_i/X_{II})$	0,54	0,891	0,504
$p(\text{no}W_i/X_{II})$	0,008	0,0186	0,00001
$p(W_i/\text{no}X_{II})$	0,04	0,001	0,006
$p(\text{no}W_i/\text{no}X_{II})$	0,198	0,0686	0,92189
$p_{\max}(W_i)$	0,99901	0,99896	0,99893
$p_{\min}(W_i)$	0,75188	0,22573	0,00015

Повторим вычисление p_{\max} и p_{\min} для всех диагнозов, не забывая, что второй и третий симптомы исключены. Результаты расчетов приведены в таблице 4.

Таблица 4 – Значения $p(W_i)$, p_{\max} и p_{\min} после ответа на вопрос № 3

	Диагноз 1	Диагноз 2	Диагноз 3
$p(W_i)$	0,94406	0,67114	0,00022
$p(W_i/X_{II})$	0,6	0,99	0,56
$p(\text{no}W_i/X_{II})$	0,01	0,93	0,012
$p(W_i/\text{no}X_{II})$	0,4	0,01	0,06
$p(\text{no}W_i/\text{no}X_{II})$	0,99	0,07	0,9312
$p_{\max}(W_i)$	0,99901	0,68479	0,51160
$p_{\min}(W_i)$	0,87209	0,22573	0,00015

На этот раз максимальное значение p_{\min} снова имеет первый диагноз, но теперь уже $p_{\max}(W_2) < p_{\min}(W_1)$ и $p_{\max}(W_3) < p_{\min}(W_1)$, следовательно, первый диагноз можно считать наиболее вероятным.

Если на каждый вопрос отвечать (+5), то экспертная система задаст все вопросы и будет вынуждена выбрать в качестве наиболее вероятного третий диагноз, хотя его вероятность почти совпадает с вероятностью второго диагноза (после окончания работы $p(W_3) = 0,99893$; $p(W_2) = 0,99228$).

Необходимо отметить, что наша экспертная система представляет собой довольно грубую модель реальной экспертной системы, способную распознавать только значительно отличающиеся друг от друга диагнозы при отсутствии корреляции симптомов. Реальные экспертные системы должны уметь различать близкие диагнозы при условии корреляции некоторых симптомов, что требует более сложной структуры программы-оболочки с подстройкой ее под конкретную область знаний.

6. Задание к лабораторным работам

Лабораторные работы №1 – №4 относятся к другому направлению интеллектуальных систем – распознаванию образов в пространстве признаков. Их описание приведено в первой части методических указаний.

6.1. Лабораторная работа № 5 – Разработка оболочки экспертной системы

1. Изучить теоретические основы проектирования экспертной системы с вероятностным характером информации о распознаваемых объектах.
2. Разработать исполняемые модули программы, реализующей оболочку экспертной системы, в среде визуального программирования на одном из объектно-ориентированных языков высокого уровня по выбору студента.

6.2. Лабораторная работа № 6 – Разработка интерфейса экспертной системы

Разработать интерфейс для реализации экспертной системы в форме Windows-приложения. Приложение должно обеспечивать:

1. Ввод и отображение информации, подготовленной для базы данных экспертной системы.
2. Отображение текущего вопроса, задаваемо пользователю, и фиксацию полученного ответа.
3. Вывод принятого решения с указанием вероятности диагноза.
4. Возможность пошагового наблюдения за процессом принятия решения.

6.3. Работа №7 – Подготовка базы данных экспертной системы

Подготовить информацию для базы данных экспертной системы в какой-либо конкретной области знаний. Область знаний выбирается студентом самостоятельно после консультации с преподавателем. Например, медицинская диагностика, поиск неисправности в устройстве, выбор профессиональной деятельности, прогнозирование событий и т.п.

Минимальные требования к объему базы данных:

- количество признаков (симптомов) – не менее 20;
- количество классов (диагнозов) – не менее 10;
- количество симптомов в каждом классе (диагнозе) – не менее 5.

Особое внимание следует уделить вопросу статистической независимости выбранных признаков (симптомов).

6.4. Работа №8 – Исследование работы экспертной системы

1. Провести эксперименты по применению подготовленной базы данных для разработанной оболочки экспертной системы. Система должна принимать правдоподобные решения при соответствующих ответах на вопросы. При необходимости следует внести изменения в исходные данные.
2. Продемонстрировать изменение порядка запроса признаков (симптомов) при изменении их значений в процессе ввода.
3. Сделать подробный анализ полученных результатов экспериментов.

Все лабораторные работы, оформляются отчетами и сдаются преподавателю по отдельности. С разрешения преподавателя допускается представление итогового отчета, объединяющего все выполненные этапы разработки экспертной системы, с условием сдачи отдельных этапов (лабораторных работ) преподавателю в устной форме без представления отдельных отчетов, но с демонстрацией результатов по каждому этапу.

Рекомендуемая литература

1. Соловьев Н.В. Введение в системы искусственного интеллекта. – СПб.: ГУАП, 2008. – 104с.
2. Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем.: Пер. с англ. – М.: Вильямс, 2003. – 864с.
3. Джексон П. Введение в экспертные системы.: Пер. с англ. – М.: Вильямс, 2001. – 624с.
4. Нейлор К. Как построить свою экспертную систему.: Пер. с англ. – М.: Энергоатомиздат, 1991. – 286с.
5. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. – СПб.: Лань, 2002. – 256с.
6. Все об экспертных системах. <http://www.expertsys.ru>