

Лабораторная работа № 7

Тема 7. Решение задач классификации байесовскими методами

Цель: приобрести опыт использования байесовских правил при решении задач классификации.

Общая часть

Как известно, теорема Байеса определяет способ получения апостериорной вероятности гипотезы при условии, что некоторое событие произошло, или имеют место наблюдения некоторого признака. При этом апостериорная вероятность определяется по априорной вероятности гипотезы и вероятности события, если гипотеза имеет место. Байесовские правила системы ЭКО основаны на этом факте. Их использование состоит в следующем. Под гипотезой понимается соотнесение проблемной ситуации с одним из интересующих нас классов, которые представляются как утверждения, в данном случае, – с классом состояния здоровья.

Вообще говоря, требуется задание априорного распределения ситуаций по классам, что в системе ЭКО задается путем указания у утверждения значения «по умолчанию». Затем к каждому утверждению (гипотезе) приписывается таблица, доставляющая основу байесовского правила, в которой указываются утверждения (наблюдения), определенности которых влияют на определенность гипотезы, и связанные с ними два веса. Первый вес определяет степень влияния на гипотезу в случае, если значение коэффициента определенности больше нуля, второй – в противоположном случае. Веса задаются в диапазоне $(-100, +100)$. В результате применения правила определенность утверждения (гипотезы) вычисляется как сумма влияния определенности наблюдений. Суммирование осуществляется по формулам, вытекающим из теоремы Байеса.

Байесовские правила применяются для вычисления коэффициентов определенности тех утверждений, об истинности которых можно судить по выполнению ряда факторов (симптомов), имеющих разную значимость.

В случае байесовского правила действие имеет вид:

```
<действие правила> ::= <описание фактора> {<описание фактора>}  
<описание фактора> ::= <ссылка на утверждение> <разделитель>  
                        <вес> <разделитель> <вес> <разделитель>  
<разделитель> ::= <пробелы> | <запятая>  
<вес> ::= <число> | <имя числового атрибута>
```

Действие байесовского правила задается в виде таблицы с тремя столбцами. В первом столбце перечисляются утверждения-факторы, коэффициенты определенности которых влияют на коэффициенты определенности целевого утверждения. Во втором столбце указываются подтверждающие, а в третьем – опровергающие веса факторов. Подтверждающий вес утверждения-фактора должен быть равен такому коэффициенту определенности, который, по мнению эксперта, следует приписать целевому утверждению в случае истинности утверждения-фактора. Опровергающий вес равен коэффициенту, приписываемому целевому утверждению экспертом в случае ложности утверждения-фактора.

Использование байесовских правил удобно в тех случаях, когда имеется некоторое число признаков, появление или отсутствие которых изменяет веру в гипотезу, а четкие логические зависимости неизвестны.

В нашем случае за признаки, влияющие на определение состояния здоровья, условно примем значения атрибутов «комплексия» и «возр_катег», а за гипотезы – утверждения «[здоровье.хорошее]», «[здоровье.удовл]» и «[здоровье.плохое]».

Вспомогательная часть

Прежде чем приступить к выполнению лабораторной работы, рекомендуется изучить модель ЦВЕТОК (PLANT) в системе ЭКО.

Задача PLANT предназначена для поиска причин заболеваний комнатных растений. В задаче рассматриваются три группы возможных причин: неправильный уход, неудовлетворительное окружение и внешние факторы. Во время консультации для каждой причины вычисляется коэффициент определенности, при этом используются правила вывода байесовского типа. **Эти правила позволяют вычислять коэффициенты определенности гипотез** (в данном случае гипотез о причине заболевания) **на основании коэффициентов определенности подтверждающих и/или опровергающих гипотезу наблюдений** (симптомов, в данном случае ~ симптомов заболевания растения). Рассмотрим, например, гипотезу о том, что причиной заболевания является слишком сухой воздух (см. символьный атрибут «неудовлетворительное окружение»). Данная гипотеза выводится с помощью правила Bda. Наблюдения (симптомы) «листья стали коричневыми по краям», «листья свернулись», «цветы быстро вянут» и «цветы и/или бутоны опали» в равной степени подтверждают данную гипотезу, что представлено с помощью соответствующих весов в правиле. Симптом «листья имеют желтые или коричневые пятна», наоборот, опровергает гипотезу.

Во время вывода решения система задает пользователю вопросы о симптомах и получает соответствующие коэффициенты определенности. По умолчанию, в тех случаях, когда пользователь не смог ответить на вопрос системы, принимается нулевой коэффициент, при котором симптом не оказывает влияния на дальнейший вывод. Затем по байесовским правилам оцениваются все гипотезы о возможных причинах заболевания. Коэффициент определенности для группы причин находится как максимальный коэффициент по этой группе. В результате консультации выдается сообщение о коэффициенте определенности групп; пользователь может выбрать интересующую его группу и ознакомиться с коэффициентами определенности всех входящих в нее причин.

Методика выполнения

В качестве основы для выполнения работы возьмите одну из моделей, использующих эмпирические (полученные опытным путём) формулы и полученных в результате выполнения темы 6. Удалите из модели все правила, вычисляющие атрибут «здоровье». Введите к каждому утверждению атрибута «здоровье» байесовское правило, например:

правило к утверждению:	зп
имя правила:	Б_зп1
действие:	кп, +50, -30
	кх, +10, 0
	кн, -70, +30
	вкм, -20, +10
	вкп, -10, 0
	вксп, -5, 0

Правила для утверждений о том, что здоровье хорошее и удовлетворительное определите сами.

Задание:

На основе одной из моделей, использующей эмпирические формулы и полученной в результате выполнения темы 6, проведите классификацию состояний здоровья с применением байесовского правила. При классификации используются атрибуты «возр_катег» и «комплексия». Проверьте влияние априорных определенностей на результат: осуществите подборку весов в правиле. Введите различные байесовские правила для «мужчин» и «женщин». Подбирая веса правил, добейтесь «разумного», с Вашей точки зрения, поведения системы

Введите ко всем утверждениям атрибутов «здоровье», «комплексия» и «возр_катег» априорные определенности (значения по умолчанию), откомпилируйте полученную модель и проанализируйте, как изменились вырабатываемые системой решения.

Проанализируйте байесовские правила. Посмотрите, какой результат получает система при ответе на все вопросы «НЕ ЗНАЮ» (**F2**). Посмотрите, как влияют Ваши ответы на какой-то один вопрос и выдаваемые системой результаты.