ЛАБОРАТОРНАЯ РАБОТА №8

Байесовская стратегия оценки выводов

8.1. Назначение, возможности и принцип работы байесовской стратегии оценки выводов
Байесовская стратегия оценки выводов - одна из стратегий, применяемых для оценки достоверности выводов (например, заключений продукционных правил) в ЭС. Основная идея байесовской стратегии заключается в оценке вероятности некоторого вывода с учетом фактов, подтверждающих или опровергающих этот вывод.
Формулировка теоремы Байеса, известная из теории вероятностей, следующая.
Пусть имеется n несовместных событий H1, H2,...,Hn. Несовместность событий означает, что никакие из событий H1, H2,...,Hn не могут произойти вместе (другими словами, вероятности их совместного наступления равны нулю). Известны вероятности этих событий: P(H1), P(H2),...,P(Hn), причем P(H1)+P(H2)+...+P(Hn) = 1; это означает, что события H1, H2,...,Hn образуют полную группу событий, т.е. одно из них происходит обязательно. С событиями H1, H2,...,Hn связано некоторое событие E. Известны вероятности события E при условиях того, что какое-либо из событий H1, H2,...,Hn произошло: P(E/H1), P(E/H2),..., P(E/Hn). Пусть известно, что событие E произошло. Тогда вероятность того, что какое-либо из событий Hi (i=1,...,n) произошло, можно найти по следующей формуле (формула Байеса):

$$P(H_i/E) = \frac{P(E/H_i)\,P(H_i)}{P(E/H_1)P(H_1) + P(E/H_2)\,P(H_2) + \ldots + P(E/H_n)\,P(H_n)}$$

События H1, H2,...,Hn называются гипотезами, а событие E - свидетельством. Вероятности гипотез P(Hi) без учета свидетельства (т.е. без учета того, произошло событие E или нет) называются доопытными (априорными), а вероятности P(Hi/E) - послеопытными (апостериорными). Величина P(EHi) - совместная вероятность событий E и Hi, т.е. вероятность того, что произойдут оба события вместе. Величина P(E) - полная (безусловная) вероятность события E. Формула Байеса позволяет уточнять вероятность гипотез с учетом новой информации, т.е. данных о событиях (свидетельствах), подтверждающих или опровергающих гипотезу.
В ЭС формула Байеса может применяться для оценки вероятностей заключений продукционных правил на основе данных о достоверности их посылок. Заключения (выводы) в этом случае соответствуют гипотезам в теореме Байеса, а посылки - свидетельствам. Обычно посылка правила в ЭС содержит несколько условий. Вероятности P(Hi) и P(E/Hi) определяются на основе статистических данных с использованием формул теории вероятностей.
Основные из этих формул следующие.

Формула умножения вероятностей (вероятность того, что произойдет и событие A, и событие B):

$$P(AB) = P(A)P(B/A) = P(B)P(A/B),$$

где P(A), P(B) - вероятности событий A и B;
P(B/A) - условная вероятность события B, т.е. вероятность события B при условии, что произошло событие A;
P(A/B) - условная вероятность события A, т.е. вероятность события A при условии, что произошло событие B.

Если события A и B независимы (т.е. вероятность одного события не зависит от того, произошло ли другое событие), то формула умножения вероятностей записывается следующим образом:

$$P(AB) = P(A)P(B).$$

Формула умножения вероятностей для нескольких событий (вероятность того, что произойдут все указанные события вместе):

$$P(A1A2 \ldots An) = P(A1)\ P(A2/A1)\ P(A3/A1,A2) \ldots P(An/A1,A2,\ldots,An\text{-}1).$$

Формула сложения вероятностей (вероятность того, что произойдет хотя бы одно из событий):

$$P(A+B) = P(A) + P(B) - P(AB).$$

Если события A и B несовместны (т.е. не могут произойти вместе), то P(AB)=0, и формула сложения вероятностей принимает следующий вид:

$$P(A+B) = P(A) + P(B).$$

Формула сложения вероятностей для нескольких событий обычно записывается следующим образом:

$$P(A_1 + A_2 + \ldots + A_n) = 1 - P(\overline{A}_1 + \overline{A}_2 + .$$

где P(A1 + A2 + ... + An) - вероятность того, что не произойдет ни одного из событий A1, A2,...An. Эту величину можно найти, например, по формуле умножения вероятностей.

8.2. Пример применения байесовской стратегии оценки выводов

Пример. В ЭС для составления прогнозов погоды вероятность дождя на следующий день определяется с учетом трех факторов: ветер, влажность, облачность (в день наблюдения). За 173 дня (из них 53 дождливых) накоплены статистические данные, приведенные в табл.8.1.

Таблица 8.1

Погода в день наблюдения		Количество случаев дождливой погоды на следующий день	Количе погоды на след
Ветер	Слабый	19	
	Умеренный	27	
	Сильный	7	
Влажность	Высокая	35	
	Средняя	12	
	Низкая	6	
	Ясно	5	

Приведенные в таблице данные означают, например, следующее: за период наблюдений (173 дня) слабый ветер наблюдался 71 день (71=19+52). В 19 случаях на следующий день погода была дождливой, в 52 случаях – без осадков. В некоторый день наблюдается следующая погода: сильный ветер, высокая влажность, облачно. Требуется найти вероятность дождливой погоды на следующий день.

В данном случае в качестве гипотез рассматриваются состояния погоды на следующий день: H1 - дождь, H2 - погода без осадков. Свидетельством здесь является сочетание трех факторов, характеризующих погоду в день наблюдения: ветер, влажность и облачность (можно сказать, что в данном случае используются три свидетельства); обозначим эти факторы как E1, E2, E3.

Обозначим наблюдаемое сочетание факторов (сильный ветер, высокая влажность, облачно) как событие E.

Определим вероятности, необходимые для расчетов по формуле Байеса.

Априорные вероятности гипотез (т.е. вероятности дождя и погоды без осадков без учета наблюдаемого состояния погоды):

$$P(H1)=53/173=0{,}306;$$
$$P(H2)=120/173=0{,}694.$$

Наблюдаемое свидетельство (состояние погоды в день наблюдения) представляет собой сочетание трех событий, наблюдаемых вместе: сильного ветра, высокой влажности и облачности. Считая эти события независимыми (т.е. считая, например, что влажность не зависит от облачности, и т.д.), можно найти условные вероятности свидетельства по формуле умножения вероятностей:

$$P(E/Hi) = P(E1,E2,E3/Hi) = P(E1/Hi)\ P(E2/Hi)\ P(E3/Hi),\quad i=1,2.$$

Найдем величины, необходимые для применения формулы умножения вероятностей:

$$P(E1/H1)=7/53=0{,}132;\ \ P(E2/H1)=35/53=0{,}66;\ \ (E3/H1)=8/53=0{,}151;$$
$$P(E1/H2)=24/120=0{,}2;\ P(E2/H2)=18/120=0{,}15;\ \ P(E3/H2)=27/120=0{,}225.$$

Здесь, например, P(E1/H1) - вероятность того, что в текущий день наблюдается сильный ветер, при условии, что следующий день будет дождливым. Эта величина

показывает, насколько часто наблюдается сильный ветер в дни, предшествующие дождливой погоде.
Подставляя найденные величины в формулу умножения вероятностей, получим:

$$P(E/H1) = 0{,}132 \cdot 0{,}66 \cdot 0{,}151 = 0{,}0132;$$
$$P(E/H2) = 0{,}2 \cdot 0{,}15 \cdot 0{,}225 = 0{,}00675.$$

Здесь, например, величина P(E/H1) - вероятность наблюдаемого состояния погоды (сильного ветра, высокой влажности и облачности) при условии, что следующий день будет дождливым.
Найдем вероятность дождливой погоды на следующий день при наблюдаемом состоянии погоды (апостериорную вероятность):

$$P(H_1/E) = \frac{P(E/H_1)P(H_1)}{P(E/H_1)P(H_1) + P(E/H_2)P(H_2)} = \frac{0{,}0132 \cdot 0{,}306}{0{,}0132 \cdot 0{,}306 + 0{,}0067}$$

Эта величина является более точной оценкой вероятности дождя на следующий день, чем априорная вероятность P(H1), рассчитанная на основе статистических данных без учета погоды в текущий день.
Следует также отметить, что полученная апостериорная вероятность (0,463) больше, чем априорная (0,306). Это означает, что наблюдаемые свидетельства с(ильный ветер, высокая влажность, облачность) подтверждают гипотезу о том, что погода на следующий день будет дождливой.
Примечание. Байесовская стратегия вывода может также применяться в ЭС, в которых для оценки достоверности выводов применяются не вероятности, рассчитанные по статистическим данным, а субъективные оценки достоверности коэффициенты уверенности), указанные экспертами.

Порядок выполнения работы

По заданию, выданному преподавателем, рассчитать вероятность указанной гипотезы на основе байесовской стратегии оценки.

Контрольные вопросы

1. Назначение байесовской стратегии оценки выводов.
2. Теорема Байеса.
3. Пример оценки достоверности гипотезы на основе байесовской стратегии (на основе задания, выполненного в ходе лабораторной работы).
4. Интерпретация вероятностей, получаемых в ходе расчетов на основе байесовской стратегии.