Концепция байесовского подхода и методов в современном анализе

Н. М. Дружинина
Финансовый университет при Правительстве Российской Федерации (Финуниверситет), Financial University
nikol_druginina@bk.ru

***Аннотация*. Байесовские методы создавались по причине множественных испытаний ученых установить проблемы статистического анализа поведения различных процессов и открыть их разрешение через использование основы байесовской методологии – теоремы Байеса. Применение этой теоремы обладает несколькими условиями, и главное из них – наличие некоторых отношений между вероятностями явлений, заключающих отличный характер и спецификации каждого явления в желательной степени.**

Ключевые слова: теорема Байеса; байесовский подход; вероятность; оценка; классификатор

Одним из главнейших достоинств байесовского подхода значится применение любой начальной (априорной) информации сравнительно параметров модели. Подобная информация проявляется в виде априорной вероятности или функции плотности вероятности. Далее исходные вероятности трансформируются с помощью выборочных данных, которые определяют собственное воспроизведение в виде апостериорного распределения оценок параметров или переменных модели.

Байесовская вероятность – это истолкование дефиниции вероятности, эксплуатируемая в байесовской теории. Вероятность интерпретируется как степень убежденности в достоверности суждения. Для обозначения степени убежденности в достоверности суждения при приобретении новой информации в байесовской теории применяется теорема Байеса.

Различные вариации байесовской интерпретации вероятности: субъективная вероятность и логическая вероятность.

- Субъективная вероятность – степень субъективной веры субъекта в вероятность установления определенного инцидента.
- Логическая вероятность – логическая связь между двумя суждениями, степень подтверждения гипотезы *H* свидетельством *E*.

Байесовская вероятность противополагается частотной, в каковой вероятность определяется сравнительной частотой образования случайного факта при довольно продолжительном мониторинге.

I. Инструменты баейсовского метода

1. теорема (формула) Баейса

Теорема Байеса (или формула Байеса) – одна из коренных теорем элементарной теории вероятностей, которая предоставляет возможность установить вероятность какого-нибудь события при условии, что случилось иное статистически взаимозависимое с ним событие. Иными словами, по формуле Байеса имеется возможность более верно перерассчитать вероятность, приняв в соображение как прежде известные данные, так и информацию, полученную от последних исследований. Формула Байеса имеет возможность быть выведена из важнейших аксиом теории вероятностей, а именно из условной вероятности. Специфика теоремы Байеса в том, что для её применения на практике нужно значительное число расчетов, вычислений, потому байесовские оценки начали активно пускать в ход только после прорыва в компьютерных и сетевых технологиях.

Формула Байеса:

$$P_A(H_i) = \frac{P(H_i) * P_{Hi}(A)}{\sum_{i=1}^{n} P(H_i) * P_{Hi}(A)},$$

где $P(H_i)$ – вероятности гипотез до опыта (априорные вероятности); $P_{Hi}(A)$ – условные вероятности возникновения события A при выборе i-й гипотезы; $P_A(H_i)$ – условная вероятность i-й гипотезы после возникновения события A (апостериорная или послеопытная вероятность).

Априорная вероятность включает в себя начальные ожидания о вероятности того или иного расчета (которые могут быть субъективны). Апостериорная вероятность исправляет начальные гипотезы в зависимости от проделанных опытов. При значительном количестве наблюдений апостериорная вероятность фактически не зависит от априорной вероятности (кроме вырожденных случаев, например P(H)=0). При небольшой статистике осмысленный отбор априорной вероятности обнаруживает главное затруднение при использовании байесовского вывода. По факту недостатка статистических данных об априорных вероятностях гипотез или наблюдениях о происхождении события A байесовскую методологию использовать нельзя, так как подобная «формализация» утрачивает физический смысл.

Все операции над вероятностями основываются на использовании двух правил:

1. Sum-rule

Пусть $A_1,\ldots,A_k$ взаимоисключающие случаи, одно из которых неизменно случается. Тогда $P(A_i \cup A_j) = P(A_i) + P(A_j)$

$$\sum_{i=1}^{k} P(A_i) = 1$$

Очевидное следствие (формула полной вероятности): $\forall B$ верно $\sum_{i=1}^{k} P(A_i|B) = 1$,

Откуда $\sum_{i=1}^{k} \frac{P(B|A_i)P(A_i)}{P(B)} = 1$

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

В интегральной форме

$$p(b) = \int p(b,a)da = \int p(b|a)p(a)da$$

2. Product-rule

Product rule гласит, что любую общую плотность неизменно можно разделить на множители

$$p(a,b) = p(a|b)p(b)$$

$$P(A,B) = (A|B)P(B)$$

Аналогично для многомерных совместных распределений

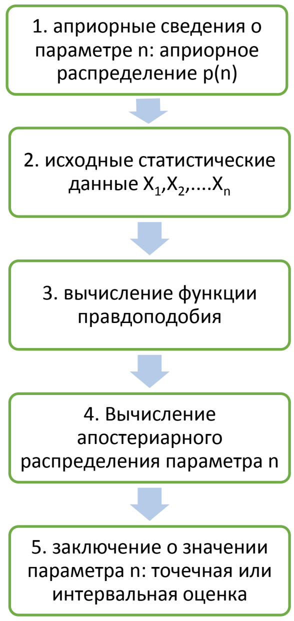
$$p(a_1,\ldots,a_n) = p(a_1|a_2,\ldots,a_n)p(a_2|a_3,\ldots,a_n)\ldots p(a_{n-1}|a_n)p(a_n)$$

Можно показать, что Sym- и Product-rule это единственными допустимыми операциями, дозволяющими исследовать вероятности как переходную ступень между истиной и ложью.

Положительные стороны байесовского подхода:

- Все параметры и величины значатся непроизвольными.
- Истинная величина параметров распределения исследователь не знает, значит, они непроизвольны с соображения нашего неведения.
- Методы Байеса возможны при объеме выборки 0! При данных значениях апостериорное распределение тождественно априорному.
- В виде оценок неведомых параметров значатся апостериорные распределения, т.е. решить задачу оценивания некой величины, это найти ее апостериорное распределение.
- Главным инструментом служит формула Байеса, в совокупности с sum- и product-rule.

Общая логическая схема байесовского оценивания



Отрицательные стороны байесовского подхода:

С 1930 гг. байесовские методы претерпевали большой объем критики и почти не применялись по причинам, которые изложены ниже:

- в байесовских методах подразумевается, что априорное распределение ясно до начала исследований и не представляется возможностей его отбора;
- утверждение решения при применении байесовских методов в нестандартных случаях вызывает огромное количество вычислительных операций. Все это связано с численным интегрированием в многомерных пространствах;
- Фишером была представлена приемлемость метода максимального правдоподобия, а значит – невозможность изобрести что-то лучшее;
- с начала 1990 гг. байесовские методы становятся более популярны, потому что с помощью них стало возможным решение многих серьезных проблем статистики и машинного обучения.

II. Применение байесовского подхода в современном анализе

Для различных задач применение байесовских методов означает получение наилучшего результата, в отличие от методов, которые основаны на частотной вероятности.

Байесовская теория применяется как метод приспособления нынешних вероятностей к только что полученным экспериментальным данным. Байесовская теория применяется для конструирования интеллектуальных фильтров, которые в современном мире используются, например, в системе фильтрации спама электронной почты. Первый фильтр был создан Полем Грахемом. «MSBN» фирмы Microsoft и «Hugin» являются одними из наиболее известных во всем мире программных систем, которые используют теорию байесовских сетей доверия,. Hugin – это программа, которая включает в себя системы принятия решений, которые базируются на байесовских сетях доверия. В ней подразумеваются две версии – Pro и Explorer. Hugin работает в среде OS Windows, а также имеет версию, разработанную под UNIX. Эта программа включает в себя два основных режима работы: режим редактирования и построения причинно-следственной сети, а также заполнения таблиц условных вероятностей; режим расчёта вероятностных оценок для принятия решения по всем событиям, входящим в причинно- следственную сеть. В этой программе расчеты происходят на основе классической теории Байеса, и на основе методов теории возможностей. Hugin имеет все важные функции информационной системы, даже такие как: хранение данных, вывод на принтер всех элементов ЭС, диагностика ошибок в работе. Баейсовы сети в основном применяются в экспертных системах, из-за того, что ЭС невозможны без средств оперирования с вероятностями.

Другие отрасли применения баейсовых сетей: это медицина – Система PathFinder (Heckerman, 1990) (создана для диагностики заболеваний лимфатических узлов, включает в себя 60 различных вариантов диагноза и 130 переменных. Версия этой программы – PathFinder-4 – получила коммерческое распространение). А также множество других разработок (Child, MUNIN, Painulim, SWAN и др.) применяются (и очень успешно) в различных направлениях медицины; космические применения – Система поддержки принятия решений Vista (Eric Horvitz), используется в Центре управления полетами NASA (NASA Mission Control Center) в Хьюстоне. Она была разработана для анализа телеметрических данных и для идентификации информации при выделении на диагностических дисплеях; финансы и экономика – байесовы методики применяются для оценки риска и прогноза доходности портфелей финансовых инструментов. В данном случае главными положительными чертами байесовых сетей в в сфере разрешения финансовых задач являются возможность коллективного учета количественных и качественных рыночных показателей, динамическое поступление новой информации и явные зависимости между существенными факторами, влияющими на финансовые показатели.

III. Классификатор Naive Bayes («Наивный» Байес)

Наивный классификатор Байеса – простой вероятностный классификатор, который основан на использовании теоремы Байеса с сильными (наивными) предположениями о независимости. Более описательным термином для базовой вероятностной модели будет «независимая модель объекта». В зависимости от точного характера вероятностной модели наивные классификаторы Байеса могут быть очень эффективно обучены в контролируемой системе обучения. Во многих практических приложениях оценка параметров для наивных моделей Байеса использует метод максимального правдоподобия. Несмотря на их наивный дизайн и, по-видимому, чрезмерно упрощенные предположения, наивные классификаторы Байеса часто работают намного лучше во многих сложных ситуациях реального мира, чем можно было бы ожидать. В последнее время тщательный анализ байесовской проблемы классификации показал, что существуют некоторые теоретические причины явно необоснованной эффективности наивных классификаторов Байеса. Преимущество классификатора Naive Bayes заключается в том, что для оценки параметров (средств и дисперсий переменных), необходимых для классификации, требуется небольшое количество данных обучения. Поскольку предполагаются независимые переменные, необходимо определить только дисперсии переменных для каждого класса, а не всю ковариационную матрицу.

Оценка параметров

Все параметры модели (например, приоритеты класса и возможности распределения вероятностей) могут быть аппроксимированы относительными частотами из набора тренировок. Это оценки вероятностей максимального правдоподобия. Сначала необходимо дискретировать недискретные функции. Дискретизация может быть неконтролируемой (специальный выбор бункеров) или контролироваться (биннинг, ориентированный на информацию в учебных данных). Если заданный класс и характеристика никогда не встречаются вместе в наборе тренировок, то оценка вероятности на основе частоты будет равна нулю. Это проблематично, поскольку оно уничтожит всю информацию в других вероятностях, когда они будут умножены. Поэтому часто желательно включать коррекцию малых выборок во всех вероятностных оценках, так что никогда не будет установлено, что вероятность равна нулю. Так как наивные методы Байеса представляют собой набор контролируемых алгоритмов обучения, нужно учитывать переменную класса y и зависимый вектор признаков x_1 через x_n, прикотором теорема Байеса утверждает следующее соотношение:

$$P(y|x_1, \ldots., x_n) = \frac{P(y)P(x_1, \ldots.x_n|y)}{P(x_1, \ldots., x_n)}$$

Используя предположение наивной независимости, что

$$P(x_i|y, x_1, \ldots., x_{i-1}, x_{i+1}, \ldots., x_n) = P(x_i|y)$$

для всех i это соотношение упрощается до

$$P(y|x_1, \ldots., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, \ldots., x_n)}$$

Поскольку $P(x_1, \ldots., x_n)$ постоянна с учетом ввода, мы можем использовать следующее правило классификации:

$$P(y|x_1, \ldots., x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i|y)$$

и мы можем использовать оценку Maximum A Posteriori (MAP) для оценки P (y) и P (x_i \ mid y); первая – это относительная частота класса y в обучающем наборе. Различные наивные классификаторы Байеса отличаются главным образом предположениями, которые они делают относительно распределения P (x_i \ mid y). Несмотря на их явно чрезмерно упрощенные предположения, наивные классификаторы Байеса хорошо работали во многих реальных ситуациях, классифицировали классификацию документов и фильтрацию спама. Для оценки необходимых параметров требуется небольшое количество данных для обучения. Изучающие и классификаторы Naive Bayes могут быть чрезвычайно быстрыми по сравнению с более сложными методами. Развязка распределений условных признаков класса означает, что каждое распределение может быть независимо оценено как одномерное распределение. Это, в свою очередь, помогает облегчить проблемы, связанные с проклятием размерности. С другой стороны, хотя наивный Байес известен как достойный классификатор, он, как известно, является плохим оценщиком, поэтому вероятность выхода из predict_proba не воспринимается слишком серьезно.

Пример. Допустим, у нас есть три документа для которых известны их классы (HAM означает – не спам): `[SPAM]` предоставляю услуги бухгалтера; `[SPAM]` спешите купить диван; `[HAM]` надо купить молоко. Модель классификатора будет выглядеть следующим образом:

	spam	**ham**
частоты классов	2	1
суммарное количество слов	6	3
предоставляю	1	0
услуги	1	0
бухгалтера	1	0
спешите	1	0
купить	1	1
диван	1	0
надо	0	1
молоко	0	1

Теперь классифицируем фразу «надо купить молоко». Рассчитаем значение выражения для класса SPAM:

$$\log\frac{2}{3}+\log\frac{1}{8+6}+\log\frac{2}{8+6}+\log\frac{1}{8+6}\approx -7.629$$

Теперь сделаем то же самое для класса HAM:

$$\log\frac{1}{3}+\log\frac{2}{8+3}+\log\frac{2}{8+3}+\log\frac{1}{8+3}\approx -6.906$$

В данном случае класс HAM выиграл и сообщение не будет помечено как спам.

1) Формирование вероятностного пространства

- *Все вероятностные оценки должны удовлетворять двум свойствам: они все обязаны быть в диапазоне [0;1]; их сумма должна быть равна единице.*

Для этого нужно из логарифмических оценок сформировать вероятностное пространство. То есть избавиться от логарифмов и нормировать сумму по единице.

$$P(c|d)=\frac{e^{q_c}}{\sum_{c'\in C} e^{q_{c'}}}$$

Здесь q_c – это логарифмическая оценка алгоритма для класса c, а возведение e (основание натурального логарифма) в степень оценки нужно для того, чтобы избавиться от логарифма ($a^{\log_a x}=x$). В баейсовском подходе к анализу возможно применение алгебраической оптимизации, которая позволяет избежать арифметического переполнения при делении экспонент друг на друга. Если сократить экспоненту оценки текущего класса в числителе и знаменателе, то в общем виде получим:

$$P(c|d)=\frac{1}{1+\sum_{c'\in C\backslash\{c\}} e^{q_{c'}-q_c}}$$

Нужно обращать внимание, что сумма в знаменателе выполняется только по классам отличным от того для которого мы считаем вероятность. Но в каждом из слагаемых присутствует логарифмическая оценка оцениваемого класса. В современных экспериментах часто возникает ситуации, когда “классические” методы анализа погрешностей и доверительных интервалов дают неправильный результат. Обычно это связано с малой статистикой или близостью измеряемых величин к физически возможной границе. В подобных случаях байесовские методы оценки вероятностей приводят к более осмысленным результатам. В настоящее время байесовские методы применяются в широком классе задач, связанных с анализом данных, принятием решения, построением экспертных систем.

Научный руководитель статьи от Финансового университета при Правительстве РФ доц. каф. «Системный анализ в экономике» Звягин Л.С.

Список литературы

[1] Айвазян С.А. Байесовский подход в эконометрическом анализе / С.А.Айвазян // Прикладная эконометрика. 2008. №1(9). с. 93-130

[2] Звягин Л.С. Итерационные и неитеративные методы Монте-Карло как актуальные вычислительные методы байесовского анализа// Международная конференция по мягким вычислениям и измерениям. 2017. Т. 1. С. 39-44.

[3] Звягин Л.С. Применение байесовских интеллектуальных технологий при моделировании управлением операционными рисками// Экономика. Управление. Право. 2011. № 11-1. С. 33-42.

[4] Байесовский подход к теории вероятностей. Примеры байесовских рассуждений. [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru дата обращения 05.04.2018

Информационно-измерительная техника и технологии: учебник / В. И. Калашников, С.Ф.Нефедов, А.Б.Путилин и др.; под ред. Г.Г. Раннева. М.: Высш. шк., 2002. 264 с.