

МОДЕЛЬ ПОЯВЛЕНИЯ СОБЫТИЯ В ВЕРОЯТНОСТНОМ МИРЕ

А.И. Богомолов,

Финансовый университет при Правительстве Российской Федерации, г. Москва
bogomol@list.ru

В.П. Невежин

Финансовый университет при Правительстве Российской Федерации, г. Москва
notpolite@ya.ru

Аннотация. Предлагается модель появления события (например, дефолт мировой экономики), которая представляет собой сочетание байесовской сети доверия и линейной вероятностной модели (модели двоичного выбора).

Статья по материалам монографии: “Инновационные и информационные технологии в развитии национальной экономики: теория и практика: Монография / Под ред. Т. С. Клебановой, В. П. Невежина, Е.И. Шохина. – М.: Научные технологии, 2013. – 528 с.

MODEL EVENT APPEARANCE IN THE WORLD OF PROBABILITY

A.I. Bogomolov, V.P. Nevezhin

Financial University Under the Government of Russian Federation,
Moscow

Summary. The model of occurrence of event, (for example, a default of economic) which represents a combination belief Bayesian networks and linear likelihood model (model of a binary choice) is offered.

This article wrote on the basis of monograph: “The innovation and information technologies in the development of national economy: theory and practice”

События, которые происходят в нашем мире, обусловлены причинно-следственными связями с другими событиями, конфигурация которых и вероятность реализации, меняются в зависимости от сообщений, которыми они обмениваются. События, происходящие в нашем мире, каковы бы они не были, можно считать случайными, так как их появление можно предсказать лишь с некоторой вероятностью. Случайное событие определено как событие, которое при осуществлении совокупности условий эксперимента может произойти или не произойти. Если при вычислении вероятности события никаких других ограничений, кроме условий эксперимента, не налагается, то такую вероятность называют **безусловной**; если же налагаются и другие дополнительные условия, то вероятность события называют **условной**. Например, часто вычисляют вероятность события B при дополнительном условии, что произошло событие A . **Условной вероятностью**

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

называют вероятность события B , вычисленную в предположении, что событие A уже наступило. Вероятность **совместного появления двух зависимых событий** равна произведению вероятности одного из них на условную вероятность второго, вычисленную при условии, что первое событие произошло, т.е.

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

В частности, отсюда получаем [1]

$$P(A | B) = \frac{P(AB)}{P(B)}$$

Безусловная вероятность существует только в теории. В реальном мире все события в той или иной степени взаимосвязаны и зависят от вероятности появления влияющих на них других событий.

Талантливый математик Фрэнк Пламптон Рамсей [2] доказал, что полная неупорядоченность невозможна. Каждое достаточно большое множество чисел, точек или объектов обязательно содержит высоко упорядоченную структуру. Отсюда мы делаем предположение, что все события образуют некий бесконечный хаос событий, в котором существуют упорядоченные структуры взаимосвязанных событий и вероятность появления отдельного события можно представить в виде некоторого графа, что и сделал Томас Байес [3].

Байесовская сеть (или Байесова сеть, Байесовская сеть доверия) – это графическая вероятностная модель, представляющая собой множество переменных и их вероятностных зависимостей [4]. Например, байесовская сеть может быть использована для вычисления вероятности того, чем болен пациент по наличию или отсутствию ряда симптомов, основываясь на данных о зависимости между симптомами и болезнями. Математический аппарат Байесовых сетей создан американским ученым Джудой Перлом, лауреатом Премии Тьюринга (2011) [5]. Перлу принадлежит авторство термина «байесовская сеть», который обозначает статистическую модель, имитирующую деятельность человеческого мозга: в такой сети непрерывно происходит обмен сообщениями без участия какого-либо главного контрольного органа.

Формально, байесовская сеть – это направленный ациклический граф, каждой вершине которого соответствует случайная переменная, а дуги графа кодируют отношения условной независимости между этими переменными. Пример простейшей байесовской сети доверия приведен на рис. 1.

Вершины могут представлять переменные любых типов, быть взвешенными параметрами, скрытыми переменными или гипотезами. Существуют эффективные методы, которые используются для вычислений и обучения байесовских сетей.

Формула совместного распределения вероятностей в узлах сети записывается как результат локальных распределений в узле и его предках (2).

$$P(X_1, \dots, X_N) = \prod_{i=1}^N P(X_i | \text{parents}(X_i))$$

Байесовские сети доверия позволяют решать две важные задачи: прогноз и диагноз [4]. Существует множество алгоритмов, представляющих приближенный вероятностный вывод: на основе посылки

сообщений, преобразовании Байесовской сети в дерево, стохастических выборок: вероятностного взвешивания, выборки по собственной важности, выборки по адаптивной важности, случайной выборки и др. [6]. Некоторые из этих алгоритмов реализованы программно в приложениях с открытым кодом (см. Приложение А).

Сложность применения сетей доверия к исследованию реальных экономических и социальных процессов заключается в громоздкости вычислений при больших размерах сетей и в приближенности оценок вероятности появления событий, влияющих на интересующие нас процесс или событие. Задачу можно попытаться облегчить, если рассматривать случайные события как бинарные переменные и определять их значение на основе бинарной модели выбора.

Модель бинарного выбора — применяемая в эконометрике модель зависимости бинарной переменной (принимающей всего два значения – 0 и 1) от совокупности факторов.

$$y_i = \begin{cases} 1, & y_i^* \geq 0 \\ 0, & y_i^* < 0 \end{cases}$$

Построение обычной линейной регрессии для таких переменных теоретически некорректно, так как условное математическое ожидание таких переменных равно вероятности того, что зависимая переменная примет значение 1, а линейная регрессия допускает и отрицательные значения и значения выше 1 [7]. Поэтому обычно используются некоторые интегральные функции распределения. Чаще всего используются нормальное распределение (пробит), логистическое распределение (логит),

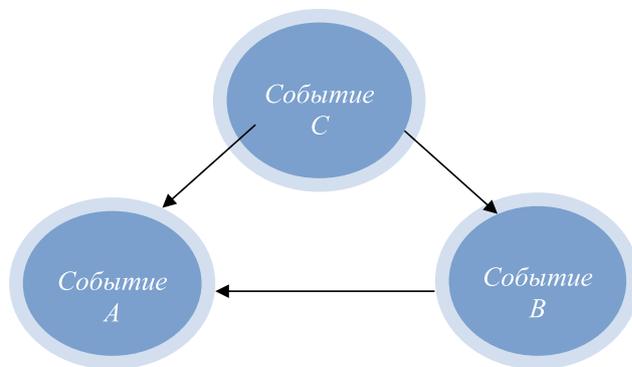


Рисунок 1. Простейшая сеть доверия Байеса

распределение Гомперца (гомпит). В логистическом распределении вероятность события определяется функцией

$$p_i = F(Z_i) = \frac{1}{1 + e^{-Z_i}},$$

где Z является линейной функцией объясняющих переменных:

$$Z = b_0 + b_1x_{i1} + \dots + b_jx_{ij} + \dots + b_kx_{ik} + e_i$$

Переменная Z является внутренней, объясняемой переменной, а X_i – внешние, объясняющие переменные (рис. 2):

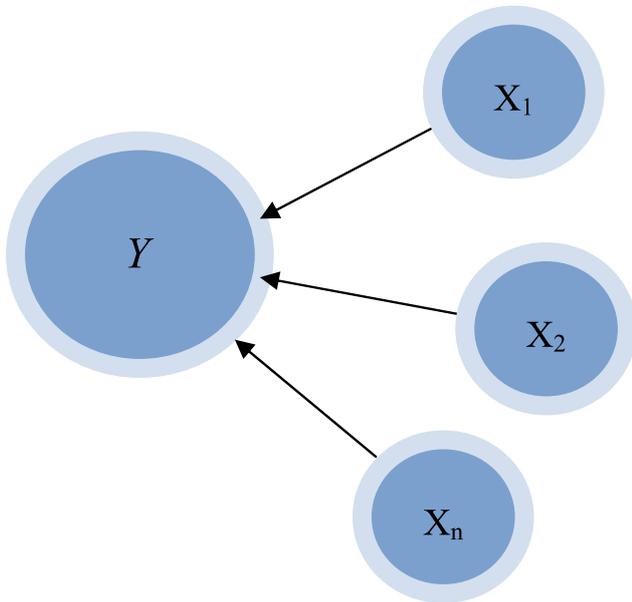


Рисунок 2. Графическая модель линейной множественной регрессии

При построении модели сети доверия для оценки вероятности интересующего нас события, вероятности «объясняющих» событий нижних уровней мы будем определять на основе бинарной модели выбора. Считается, что объясняющие события бинарной модели являются независимыми, в то время как другие события сети доверия могут быть взаимозависимыми (рис. 3).

Рассмотрим модель возможного дефолта российской экономики вследствие развития кризисных

событий в экономиках Европы, США и Китая (события А, Y, В и С). Вероятность развитие негативных событий (дефолта) в экономике Европы – Y, будем определять на основе макроэкономической модели Хоггарта, Соренсена и Зикчино (2005), Алвеса (2005), Тройтлера и Вайлера (2006) [8]. Как правило, в их основе лежит модель векторной авторегрессии

$$Z_{t+1} = C + \sum_{j=1}^p \Phi_j Z_{t+1-j} + \varepsilon_{t+1}$$

где C – вектор постоянных величин,

Φ – матрица коэффициентов,

Z – вектор эндогенных переменных, который включает как вероятность дефолта (или другие прокси масштаба экономической рецессии), так и другие параметры, отражающие состояние экономики.

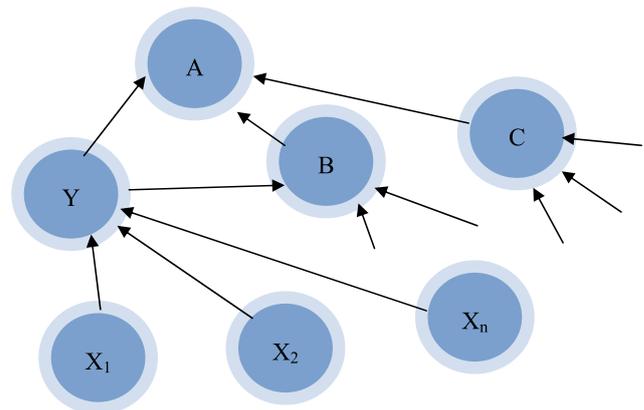


Рисунок 3. Пример сети доверия, где вероятности событий В и С определяются экспертным путём, а вероятность события Y на основе бинарной модели

Вероятности негативных событий В, С и А определяются на основе алгоритмов сети доверия с помощью соответствующей программной системы, например Netica [9].

Комбинированный подход на основе сетей доверия и бинарных моделей расширяет возможности первого и второго методов и имеет большую объяснительную способность, чем каждый из них в отдельности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. http://www.matburo.ru/tvbook_sub.php?p=par15
2. <http://rusnauka.narod.ru/lib/matem/ramsey/rams.htm>
3. http://ru.wikipedia.org/wiki/%C1%E0%E9%E5%F1,_%D2%EE%EC%E0%F1
4. Changyun Wang. Bayesian Belief Network Simulation. Department of Computer Science Florida State University February 24, 73 pages, 2003
5. <http://www.osp.ru/cw/2012/07/13014066/>
6. Дзюбан Юрий, Болдак Людмила. Анализ алгоритмов вывода в Байесовских сетях доверия. <http://www.hpc-ua.org/sites/default/files/proceedings-2012/35.pdf>
7. Пасечник, А.А. Использование эконометрических моделей бинарного выбора для оценки вероятности банкротства российских банков [Текст] / А.А. Пасечник, Д.А. Пасечник, Е.Н. Лукаш // Молодой ученый. — 2011. — №10. Т.1. — С. 137-148.
8. <http://www.beintrend.ru/2011-12-01-20-09-36>
9. <http://www.norsys.com/netica>