УДК 51: 101.8
DOI:
10.15372/PS20170404

В.М. Резников

# К ВОПРОСУ ОБ УНИВЕРСАЛЬНОСТИ БАЙЕСОВСКИХ СЕТЕЙ В КОНТЕКСТЕ НЕЗАВИСИМОСТИ ДАННЫХ

Известно, что эффективность приложений теории вероятностей, математической статистики, байесовских сетей связана с использованием свойства независимости. Показано, что не существует универсальных подходов к определению независимости. Полагаем справедливым мнение А.Н. Колмогорова и Ю.В. Сачкова, что созданию обобщенного определения независимости в математике препятствует отсутствие философских исследований по категории независимости.

*Ключевые слова*: независимость, теория вероятностей, байесовские сети, теорема Байеса, экспертные системы.

V.M. Reznikov

# On the universality of Bayesian networks in the context of data independence

It is known that the efficiency of application of the probability theory, mathematical statistics and Bayesian networks depends on using the independence property. The article shows that there are no universal approaches to definition of independence. We believe that A.N. Kolmogorov's and Yu.V. Shachkov's opinion that the lack of philosophical study of the category of independence impedes providing a general definition of independence in mathematics is correct.

*Keywords*: independence; probability theory; Bayesian networks; Bayesian theorem; expert systems

Методологическую основу экспертных систем составляет, как известно, концепция байесовизма. Одной из целей байесовизма является определение правдоподобия гипотез. Правдоподобие определяется на основе теоремы Байеса. Применение теоремы предполагает задание первоначальных вероятностей изучаемых гипотез. Затем на основе наблюдений над реализацией событий, связанных с этими гипотезами, осуществляется изменение правдоподобия гипотез. Изменение вероятностей гипотез интерпретируется как прогресс в научной области. Байесовизм является вполне универсальной концепцией. Она включает в себя такие различные направления, как эмпирический байесовизм, объективный байесовизм, субъективный байесовизм. Представители эмпирического байесовизма определяют первоначальные вероятности гипотез на основе реально наблюдаемых частот, объективные байесовисты – на основе синтеза частот и субъективных предпочтений, субъективные байесовисты – на основе индивидуальных предпочтений. Можно ли считать экспертные системы, базирующиеся на байесовских сетях, универсальными? На первый взгляд, да. Вычисления в них осуществляются на основе теоремы Байеса, причем первоначальные вероятности назначаются на основе самых различных подходов, что косвенно свидетельствует об универсальности байесовских экспертных систем.

Отметим, что современные экспертные системы предназначены для анализа больших объемов данных, множества событий и т.д. на основе формального аппарата, который является синтезом теории графов и теории вероятностей. В теории вероятностей эффективные вычисления вероятностей для множества событий базируются на предположении, что эти события оказываются независимыми. В настоящее время не существует универсальных подходов к определению независимости. Напомним определение независимости в теории вероятностей: события $A_1$, $A_2$, …, $A_n$ считаются независимыми в совокупности, если для любого подмножества этих событий последние оказываются независимыми.

Приведем некоторые способы определения независимости. Во-первых, введение независимости основано на верификации независимости для множества событий. Однако такой подход предполагает знание совместной вероятности для любого подмножества этих событий. Получается, что для того чтобы верифицировать независимость событий, необходимо знать $2^n$ различных вероятностей. В прикладных исследованиях вероятности событий сложно определить. В этих исследованиях анализ независимости осуществляется на основе корреляции данных.

Во-вторых, например, в математической статистике, к модели независимых наблюдений приходят на основе малосвязанных данных. Так, если данные оказываются слабо коррелированными, то их считают независимыми. Этот подход не универсален, так как понятие корреляции не определено для произвольных распределений.

Например, понятие является обоснованным для нормального распределения, но не имеет смысла для распределения Пуассона. Насколько оправдан переход от модели практически некоррелированных, мало зависимых наблюдений к модели независимых экспериментов? В работе П.Е. Эльясберга приведен пример, в котором переход от модели практически независимых наблюдений с коэффициентом корреляции 0,01 к модели независимых экспериментов полагали вполне оправданным. В этом примере вычислялась дисперсия для полностью независимых результатов наблюдений и мало зависимых результатов, полученных на основе большого числа наблюдений, равного 1000. Оказалось, что дисперсия независимых данных более чем на порядок меньше дисперсии малосвязанных данных [5].

В-третьих, часто к модели независимых наблюдений приходят на основе содержательных соображений и контроля над фоновыми условиями проведения экспериментов. Так, полагают, что если эксперименты содержательно независимы, то имеются основания принять модель независимых наблюдений. Дополнительным аргументом является контроль фоновых условий проведения экспериментов. Однако даже в физике не всегда удается осуществить их полный контроль. Из истории физики известно, что когда в лаборатории Резерфорда был открыт радон, все приборы зашкаливали [1]. Долгое время никто не понимал, что происходит, потом догадались, что новое вещество является газообразным. Тогда рафинировали условия проведения экспериментов и приборы стали показывать осмысленные результаты.

Отметим, что в экспертных системах используют собственные подходы к определению независимости. Прежде чем описывать эти подходы, следует напомнить основные идеи относительно применяемого в экспертных системах математического аппарата. В современных экспертных системах используется синтез двух математических дисциплин: теории графов и теории вероятностей.

Приведем минимальную информацию о теории графов применительно к исследованиям в области экспертных систем. Граф – это совокупность двух множеств: множества вершин и множества связей. Множество вершин обычно обозначается символом $V$, а множество ребер символом $E$, соответственно граф символически изображается так: $G = (V, E)$. Обычно вершины – это интерпретации некоторых событий или значений переменных, описывающих исследуемые положения дел. С помощью множества ребер определяются отношения между вершинами или связи описывающих вершины переменных. Если переменные связаны, то они соединяются ребром. Важным для дальнейшего изложения является понятие пути. Последовательность ребер $X_1$, $X_2$, …, $X_N$ называется путем, если любые два соседних ребра $X_i$, $X_{i+1}$ оказываются связанными. Если известно, что одна вершина не просто связана с другой, но оказывает влияние на нее, то эти вершины соединяются ребром со стрелкой. В этом случае графы называются направленными. В приложениях наиболее популярны направленные графы без петель (Directed Acyclic Graph, DAG).

В теории графов для описания отношений используются отношения родственности. Пусть дан ориентированный граф $G = (V, E)$, $X$ и $Y$ вершины графа, если ребро направлено от $X$ к $Y$, то $X$ называется родителем, а $Y$ – ребенком. Если существует путь от $X$ к $Y$, то $X$ называется предком, а $Y$ – потомком, а если не существует пути от $X$ к $Y$, то $Y$ не является потомком $X$ и называется непотомком.

Теперь укажем на необходимые понятия из теории вероятностей. Нам понадобится только понятие распределения вероятностей для переменных, описывающих вершины графа, символически – $P(x_1, x_2, \ldots, x_N)$. В простейшем случае для дискретной переменной распределение вероятностей означает задание значений, которые принимает эта переменная, и соответствующие этим значениям вероятности, с которыми переменная принимает эти значения. Предположим, что задан граф $G = (V, E)$, у которого вершины имеют распределение $P$. Для указания на граф и распределение вероятностей его вершин вводится обозначение $(G, P)$. Теперь перейдем к анализу независимости в экспертных системах.

Независимость в экспертных системах определяется на основании марковского условия на графах. Пусть дан граф DAG $G$ с совместным распределением $P$ для вершин графа. Будем говорить, что $(G, P)$ удовлетворяет марковскому условию, если каждая переменная $X \in V$ в вероятностном плане является условно независимой от множества непотомков, где в качестве условия взяты родители переменной $X$. Если граф удовлетворяет марковскому условию, то вычисление совместного распределения вероятностей $P$ для переменных, описывающих вершины графа, многократно упрощается на основании следующей теоремы [7].

*Теорема* 1. Если ($G$, $P$) удовлетворяет марковскому условию, то совместное распределение $P$ равняется произведению условных сингулярных распределений переменных, где в качестве условия берутся родители для каждой переменной. Пусть Par ($X_i$) – это обозначение множества родителей для вершины $X_i$, тогда символическое представление результата теоремы имеет вид

$$P\,(x_1, x_2, \ldots, x_N) = P\,(x_1/\mathrm{Par}(X_1))\;P\,(x_2/\mathrm{Par}(X_2))\;\ldots\;P\,(x_N/\mathrm{Par}(X_N) \quad (1)$$

Получили, что совместное распределение вероятностей от $N$ переменных сводится к произведению $N$ сингулярных условных распределений. Однако более сильным в прагматическом плане является обратный результат, показывающий, что на основании знания условных сингулярных распределений вероятностей для $N$ переменных определяется совместное распределение для этих переменных. Этот результат формулируется в виде следующей теоремы [7].

*Теорема* 2. Пусть дан граф DAG $G$, каждая его вершина описывается с помощью переменной, и для каждой этой переменной известно условное распределение, где в качестве условия взяты переменные, описывающие родительские вершины. Тогда произведение условных распределений обеспечивает получение совместного распределения переменных:

$$P\,(x_1/\mathrm{Par}(X_1))\;\; P(x_2/\mathrm{Par}(X_2))\;\ldots\;P\,(x_N/\mathrm{Par}(X_N)) = P\,(x_1, x_2,\ldots,x_N) \quad (2)$$

Более того, ($G$, $P$) удовлетворяет марковскому условию. Утверждение (2) имеет место для всех дискретных распределений и для некоторых непрерывных, например, для нормального распределения. Показано [6], что для некоторых непрерывных распределений результат (2) не имеет места.

Однако для реальных приложений вычисление совместных распределений с помощью формулы (2) не является вполне операциональным. Формула определяет совместное распределение для независимых переменных, когда заданы сингулярные вероятности этих переменных, однако вероятность является теоретической величиной и, как правило, априори неизвестна. Описание независимости с помощью родственных отношений не является в полной мере конструктивным, так как оно не определяет, в силу каких препятствий (блокировок) части графа оказываются независимыми.

Существуют три вида блокировок. Первая блокировка для пути от вершины $X$ к вершине $Y$ возникает в промежуточной вершине $Z$ (между вершинами $X$ и $Y$), причем эта вершина совпадает с общим концом двух расходящихся от $Z$ смежных ребер. Во втором случае промежуточная вершина $Z$ совпадает с начальной вершиной одного ребра и концевой вершиной другого ребра. Третья блокировка для пути от вершины $X$ к вершине $Y$ возникает, если в некоторой вершине $Z$ (между вершинами $X$ и $Y$) встречаются два ребра, причем пересечение происходит в начальных вершинах этих ребер.

Дадим формальное определение блокирования.

*Определение*. Пусть $G = (V, E)$ DAG, $A \subseteq V$, $X$ и $Y$ – различные вершины, принадлежащие $V$, но не принадлежащие $A$. Пусть $\rho$ – путь между $X$ и $Y$. Тогда $\rho$ блокируется $A$, если существует вершина $Z \in A$, такая, что в ней имеет место первая или вторая блокировка. Кроме того, $\rho$ блокируется $A$, если имеет место третий случай блокирования в вершине $Z$, такой, что ни $Z$, ни вершины, являющиеся потомками $Z$, не принадлежат $A$ [7].

Говорят, что две вершины $X$ и $Y$ отделяются с помощью множества $A$, если они блокируются с помощью этого множества. С помощью определения отделимости вершин определяется понятие отделимости множества вершин. Два множества $A$ и $B$ отделяются с помощью множества $C$, если любые две вершины $a \in A$ и $b \in B$ отделимы с помощью $C$.

Обобщением условия Маркова для родственных отношений является условие Маркова для отделимых непересекающихся множеств. Пусть $G$ – направленный граф, вершины графа интерпретируются с помощью переменных и для этих переменных задано распределение $P$. Тогда $(G, P)$ удовлетворяет условию Маркова, если для любых трех взаимно непересекающихся множеств $A$, $B$, $C$, элементами которых являются вершины графа, оказывается, что как только множества $A$ и $B$ отделяются с помощью множества $C$, то множества $A$ и $B$ оказываются условно вероятностно независимыми, где в качестве условия выступает множество $C$.

Этот результат не является неожиданным. С одной стороны, это обобщение марковости для родственных связей на случай несвязанных частей графа. Не обязательно определять независимость для множеств, связанных родственными отношениями. Вполне естественно определять независимость для несвязанных множеств, и тогда отношение «не быть родственниками» оказывается частным случаем отношения несвязанности для множеств. С другой стороны, в геометрической интерпретации критерий марковости операционализируется, так как указывается конкретный механизм (блокирование), реализующий условия, при которых множества становятся независимыми.

Необходимо отметить бесспорную интуитивность этого результата. Действительно, вполне согласуется с интуицией то, что отделимые, несвязанные множества оказываются условно независимыми в вероятностном плане, где в качестве условия используется множество, осуществляющее отделение множеств, которые при этом становятся вероятностными условно независимыми множествами. Итак, марковость получается, когда геометрическая независимость влечет вероятностную независимость. Символически это записывают следующим образом:

$$I_G(A, B/C) \to I_P(A, B/C). \qquad (3)$$

Здесь $I_G(A, B/C)$ – это независимость или отделимость множеств $A$ и $B$ с помощью множества $C$, $I_P(A, B/C)$ – это условная вероятностная независимость множеств $A$ и $B$, где в качестве условия взято множество $C$. Не только геометрическая независимость влечет вероятностную независимость, при определенных условиях имеет место и обратный результат: из марковского условия выводима независимость графических структур. Пусть дан граф DAG $G = (V, E)$; $P$ – множество всех вероятностных распределений, таких, что $(G, P)$ удовлетворяет условию Маркова. Тогда для любых трех взаимно непересекающихся множеств $A$, $B$, $C \subseteq V$, и для любых $p \in P$ имеет место

$$I_P\,(A, B/C) \to I_G\,(A, B/C) \qquad (4)$$

Отметим, что экспертные системы на основе байесовизма представляют интерес для философов. Во-первых, теорема Байеса значима для философских исследований, так как ее применение обеспечивает метод определения фактора, ответственного за то, что не оправдалось предсказание, и тем самым теорема обеспечивает средства для критического анализа методологического холизма. Во-вторых, направленные графы в экспертных системах часто строят посредством причинных связей между вершинами графа. Построение причинных сетей осуществляется на основе концепции манипуляционной причинности. В рамках этой концепции *A* есть причина *B*, если привлечение *A* изменяет *B* или изменяет условную вероятность реализации *B*. В причинных концепциях отношения причинности и независимые отношения являются противоположными сущностями. Причинная связь – это наиболее сильная возможная связь, а независимость тогда означает отсутствие связи. В известной работе Г. Рейхенбаха предложен способ введения вероятностных причинных отношений на основе принципа общей причины [3]. Согласно этому принципу, если два события *B* и *C* оказываются определенным образом связанными, то или *B* причина *C*, или *C* причина *B*, или существует событие *A*, которое является причиной *B* и *C*, причем в случае учета *A* события *B* и *C* оказываются независимыми.

В приложениях экспертных систем независимость часто вводится на основании принципа максимальной энтропии [8]. Этот принцип вполне обоснован в термодинамике, в теории информации. Однако решающими аргументами в пользу его применения для вычисления вероятностей в экспертных системах считаются доводы, приведенные Дж. Пэрисом и А. Венковской [8]. Они показали совместимость принципа максимальной энтропии с целым рядом рациональных оснований:

1) принцип переименования. Полученное решение для задачи вычисления вероятности выпадения грани для правильной кости не изменится, если будет рассмотрена другая грань;

2) принцип иррелевантной информации. Знание, иррелевантное проблеме, игнорируется;

3) принцип постоянства. Использование данных, поддерживающих решение, не изменяет принятое решение;

4) принцип относительности. Новая информация имеет значимость относительно решаемой проблемы;

5) принцип непрерывности. Небольшие изменения в знании не приводят к большим изменениям в назначенных вероятностях;

6) принцип эквивалентности. Вероятностные оценки на основе двух одинаковых баз знаний должны быть одинаковыми.

Таким образом, в экспертных системах предложены собственные способы введения независимости. Это, во-первых, геометрическая независимость на графах, во-вторых, независимость на основе принципа общей причины, в-третьих, независимость на основе принципа максимальной энтропии. Ни один из этих способов не является универсальным. Геометрический способ ориентирован исключительно на геометрические структуры. Принцип общей причины исключает все виды отношений, кроме причинных, хотя, несомненно, важными являются логические связи, функциональные связи и некоторые другие. Принцип энтропии обоснован для неживой природы.

Помимо специфического способа задания независимости слабым звеном формального аппарата DAG является то, что он не рассчитан на анализ циклов в графах. В качестве альтернативы предлагается логико-вероятностный аппарат на основе метода неподвижных точек [9]. Отметим, что одним из препятствий в разработке формализаций понятия независимости является то, что это понятие фактически не исследуется в философии. Хотя, как справедливо отмечал Ю.В. Сачков, понятие независимости имеет категориальный статус [4]. Оно имеет значение уже в классической философии, например при анализе проблемы воли, в социальной философии, в ряде наук, например в биологии, логике, математике, особенно в теории вероятностей. А.Н. Колмогоров полагал, что определение общих условий, при которых имеет смысл обращение к модели независимых экспериментов, относится к компетенции философии естественных наук [2]. Сложность конструктивного определения независимости связана с тем, что исследования философов традиционно направлены на поиски разного рода зависимостей, а независимость остается вне внимания. Кроме очевидного понимания независимости, состоящего в том, что *A* не зависит от *B*, если *A* при учете *B* не меняется или при учете *B* исследователь не узнает ничего нового об *A*, трудно предложить какой-либо осмысленный и конструктивный способ определения независимости. Нужны новые идеи и новые подходы.

## Литература

1. *Алимов Ю.И.* Альтернатива методу математической статистики. – М.: Знание, 1980.

2. *Колмогоров А.Н.* Основные понятия теории вероятностей. – М.: Наука, 1974.

3. *Рейхенбах Г.* Направление времени. – М.: Едиториал УРСС, 2003.

4. *Сачков Ю.В.* Вероятность как загадка познания // Эпистемология и философия науки. – 2011. – Т. 30, № 4. – С. 133-148.

5. *Эльясберг П.Е.* Измерительная информация: Сколько ее нужно? Как обрабатывать? – М.: Наука, 1983.

6. *Dawid A.P., Studeny M.* Conditional products: An alternative approach to conditional independence. URL: https://pdfs.semanticscholar.org/b2b2/74541cfa902956fc589d6daf00b1afbe65eb.pdf. (Дата обращения 01.03.2017).

7. *Neapolitan E.* Learning Bayesian Networks. – Illinois University Press, 1990.

8. *Paris J., Vencovska A.* In defense of maximum entropy inference process // International Journal of Approximate Reasoning. – 1997. – No. 17. P. 77–103.

9. *Vityaev E., Demin A., Ponomaryov D.* Probabilistic generalization of formal concepts // Lecture Notes in Computer Science. – 2012. – Vol. 7162. – P. 394–410.

## References

1. *Alimov, Yu.I.* (1980). Alternativa metodu matematicheskoy statistiki [An Alternative to the Method of Mathematical Statistics]. Moscow, Znanie Publ.

2. *Kolmogorov, A.N.* (1974). Osnovnye ponyatiya teorii veroyatnostey [Basic Concepts of the Probability Theory]. Moscow, Nauka Publ.

3. *Reichenbach, H.* (2003). Napravlenie vremeni [The Direction of Time]. Moscow, Editorial URSS Publ. (In Russ.).

4. *Sachkov, Yu.V.* (2011). Veroyatnost kak zagadka poznaniya [Probability as a Knowledge Mystery]. Epistemologiya i filosofiya nauki [Epistemology and Philosophy of Science], Vol. 30, No. 4, 133–148.

5. *Elyasberg, P.E.* (1983). Izmeritelnaya informatsiya: Skolko ee nuzhno? Kak obrabatyvat? [Measurement Information: How Much of it Does One Need? How to Process It?]. Moscow, Nauka Publ.

6. *Dawid, A. & M. Studeny.* (1990). Conditional products: An alternative approach to conditional independence. Available at: https://pdfs.semanticscholar.org/b2b2/74541cfa902956fc589d6daf00b1afbe65eb.pdf (date of access: 01.03.2017).

7. *Neapolitan, E.* (1990). Learning Bayesian Networks. Illinois University Press.

8. *Paris, J. & A. Vencovska.* (1997). In defense of maximum entropy inference process. International Journal of Approximate Reasoning, 17, 77–103.

9. *Vityaev, E., A. Demin & D. Ponomaryov*. (2012). Probabilistic generalization of formal concepts. Lecture Notes in Computer Science, 7162, 394 – 410.

### Информация об авторе

*Резников Владимир Моисеевич* – кандидат философских наук, доцент, старший научный сотрудник Института философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8); доцент кафедры логики и методологии науки Новосибирского национального исследовательского государственного университета (630090, Новосибирск, ул. Пирогова 2, e-mail: mathphil1976@gmail.com).

**Information about the author**

*Reznikov, Vladimir Moiseevich* – Candidate of Sciences (Philosophy), Associate Professor, Senior Researcher at the Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090, Russia); Associate Professor at the Department of Logic and Methodology of Science, Novosibirsk National Research State University (2, Pirogov st., Novosibirsk, 630090, Russia, e-mail: mathphil1976@gmail.com).

Дата поступления 17.08.2017