ЦИКЛЫ В БАЙЕСОВСКИХ СЕТЯХ: ВЕРОЯТНОСТНАЯ СЕМАНТИКА И ОТНОШЕНИЯ С СОСЕДНИМИ УЗЛАМИ

А. Л. Тулупьев[1], С. И. Николенко[2], А. В. Сироткин[3]

[1,3]Санкт-Петербургский институт информатики и автоматизации РАН, [2]Санкт-Петербургское отделение Математического института им. В. А. Стеклова РАН

[1,3]СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178: [2]ПОМИ РАН, Санкт-Петербург, наб. р. Фонтанки, д. 27, 191023

[1]<alt@iias.spb.su>, [2]<sergey@logic.pdmi.ras.ru>, [3]<sav_ptts@inbox.ru>

УДК 681.3

Тулупьев А. Л., Николенко С. И., Сироткин А. В. **Циклы в байесовских сетях: вероятностная семантика и отношения с соседними узлами** // Труды СПИИРАН. Вып. 3, т. 1. — СПб.: Наука, 2006.

Аннотация. *Мы рассматриваем направленные циклы (циклы обратной связи) в байесовских сетях доверия, учитывая их взаимодействие с предками и потомками цикла в БСД, а также приводим результаты в случае циклов с многозначными переменными. Учет предков цикла приводит к тому, что цикл в БСД может сам по себе повлиять на своих предков, сделав некоторые их означивания противоречивыми. Мы также рассматриваем достаточное условие противоречивости цикла в АБС, при выполнении которого цикл можно преобразовать в линейную цепь ФЗ константного порядка. Результаты снабжены численными примерами, в том числе примером цикла, некоторые означивания предков которого оказываются противоречивыми.* — Библ. 11 назв.

UDC 681.3

Tulupyev A. L., Nikolenko S. I., Sirotkin A. V. **Cycles in Bayesian networks: probabilistic semantics and relations with neighboring nodes** // SPIIRAS Proceedings. Issue 3, vol. 1. — SPb.: Nauka, 2006.

Abstract. We consider directed (feedback) cycles in Bayesian belief networks, taking into consideration their interaction with successors and predecessors in the network. We also consider the case of multivalued variables. Our discussion leads to the conclusion that a cycle in a BBN may by itself influence its parents, making some of their assignments inconsistent. We also consider a sufficient condition of a cycle's consistency in an algebraic Bayesian network. If it holds, the cycle may be transformed into a linear chain of knowledge patterns of constant size. We also make numerical examples, including an example of a cycle, some assignments of whose parents turn out to be inconsistent. — Bibl. 11 items.

1. Введение

Проблема учета направленных циклов (или циклов обратной связи) в байесовских сетях доверия (БСД) возникает как на практике (направленные циклы вполне могут возникнуть в процессе экспертного построения БСД или автоматического построения сетей зависимостей [1]), так и при теоретическом анализе естественных обобщений аппарата БСД.

В наших предыдущих работах [2–6], как правило, шла речь исключительно об *изолированных* циклах, т.е. фактически о сети, целиком состоящей из направленного цикла. В настоящей работе мы не только изложим вероятностную семантику изолированных циклов, в том числе и над многозначными переменными, но и рассмотрим, какое влияние оказывает цикл в БСД в некоторых частных случаях на своих соседей: родителей и сыновей.

Сразу отметим основные особенности нашего подхода: мы работаем в рамках существующих теорий и определений. Вероятностный подход гарантирует выполнение аксиом теории вероятностей, что очень важно в искусствен-

ном интеллекте (к сожалению, некоторые аппараты выходят за рамки аксиоматики, но объявляют, что полученные результаты могут быть интерпретированы как вероятности). И к теории БСД мы подходим столь же прямо и «наивно»: понятие d-разделимости принимается в том виде, в котором оно существует в теории БСД. Однако вполне возможно, что для более адекватной работы с циклическими БСД это понятие нужно будет скорректировать; это остается интересной проблемой, которая в настоящей статье не рассматривается.

2. Основные понятия

Байесовская сеть доверия — это ациклический направленный граф, вершины которого представляют собой переменные, а связи между ними указывают на непосредственную зависимость между этими переменными. Основная идея байесовской сети доверия заключается в том, что слишком большое для прямого анализа множество утверждений (переменных) структурируется при помощи графа, в котором связаны друг с другом вершины, отвечающие зависимым друг от друга утверждениям, а получающаяся структура поддается достаточно простому анализу.

В байесовской сети доверия связь (говоря языком теории вероятностей и вероятностной логики, зависимость) на самом деле имеется отнюдь не только между узлами, непосредственно связанными ребром. Определение БСД включает в себя понятие так называемой *d-разделимости,* которая говорит, когда узлы БСД условно независимы. Так как d-разделимость окажется весьма важной для рассмотрения соседей циклов в БСД, мы должны рассмотреть ее достаточно подробно. Поскольку это понятие — одно из самых базовых в теории БСД, мы могли бы порекомендовать читателю обратиться к уже имеющимся источникам на английском языке [7,8]. В русской литературе байесовским сетям доверия внимания до сих пор практически не уделялось; мы постарались в какой-то мере исправить это досадное упущение в [9].

Для того чтобы определить d-разделимость, сначала определяют три способа связи между узлами. *Последовательная связь* (s*erial connection*) — это прямая связь между узлами сети (рис. 1а). В этом случае *1* влияет на *2*, а *2*, в свою очередь, влияет на *3*, и узлы *1* и *3* получаются связанными. Однако, если в *2* поступило свидетельство, связь между *1* и *3* нарушается: если мы твердо знаем, что промежуточное событие произошло, связи между его причиной и следствием уже нет (ее обеспечивал тот факт, что вероятность следствия зависит от вероятности причины, но если следствие уже произошло, этот механизм больше не работает).

В случае *расходящейся связи* (*diverging connection*) у одного узла сети имеется несколько потомков (рис. 1б). Информация об одном из потомков может повлиять на вероятность другого потомка одного и того же узла. Это связано с тем, что не только информация о произошедшей причине повышает вероятность (правильнее, конечно, было бы сказать «оценку вероятности», а еще правильнее и более нейтрально — «степень нашей уверенности в», но для краткости мы будем говорить просто «вероятность») следствия, но и случившееся следствие повышает вероятность причины. Но если общий предок уже получил означивание, то связь нарушается. Действительно, если мы уже твердо знаем, что причина двух следствий имеет место, эти следствия становятся независимыми.

ᴜᴛᴇᴄʜ ʀ ᴇ ʟ ʀ ʀ [ϕ] . 2006. J ʳ ᶠ. 3. ᴜ. 1. ISSN 2078-9181 (ᶠᶠ.ᶠ.ᶣ.), ISSN 2078-9599 (ᵗᵉᶜʰᴏᴄᵃᵗ)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

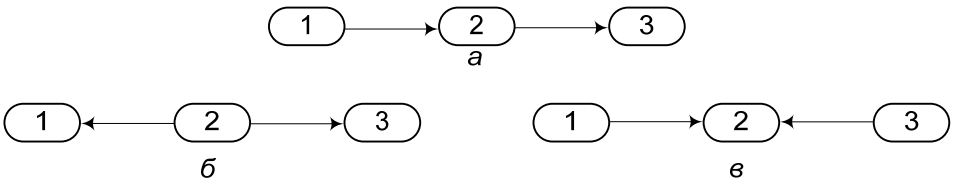

Рис. 1. Типы связи в БСД.

а — последовательная связь; *б* — расходящаяся связь; *в* — сходящаяся связь.

Сходящаяся связь (*converging connection*) возникает, когда два узла имеют общего потомка (рис. 1в). Эта ситуация — самая интересная. Казалось бы, связи между узлами *1* и *3* нет: если произошло *1*, то это повлияет на вероятность события *2*, но вероятность *3* от этого измениться не должна. Однако ситуация меняется, если свидетельство *2* уже получено. Действительно, после получения свидетельства *2* вероятности обеих его причин повышаются: раз уж что-то произошло, значит, какая-то причина у этого была — неизвестно, какая именно, но обе причины становятся более вероятными. Но если теперь мы вдруг узнаем, что одна из причин произошла, это должно понизить вероятность другой причины: ведь следствие уже объяснено, и его влияние на вторую причину должно уменьшиться. Итак, в случае сходящейся связи зависимость должна появляться, только если общий потомок (или любой из его потомков — там ситуация будет аналогичной) уже получил какое-то означивание.

Теперь можно дать формальное определение d-разделимости.

Определение 1. Два узла направленного графа x и y называются d-разделенными, если для всякого пути из x в y (путь здесь не учитывает направление ребер) существует такой промежуточный узел z (не совпадающий ни с x, ни с y), что либо связь в пути в этом узле последовательная или расходящаяся, и узел z получил означивание, либо связь сходящаяся, и ни узел z, ни какой-либо из его потомков означивания не получил. В противном случае узлы называются d-связанными.

Для удобства нам потребуется обозначение $\tilde{x}$ — аргументное место, т. е. позиция в формуле, куда может быть поставлено одно из двух значений — либо x, либо $\bar{x}$. Мы будем использовать его для обозначения произвольного означивания.

3. Вероятностная семантика изолированного цикла в БСД

Рассмотрим БСД, представляющую собой изолированный цикл из конечного числа вершин. Это означает, что в качестве начальных данных заданы тензоры условных вероятностей

$$p(x_2 \mid x_1),\ p(x_2 \mid \bar{x}_1),\ p(x_3 \mid x_2),\ p(x_3 \mid \bar{x}_2),\ \ldots,\ p(x_n \mid x_{n-1}),\ p(x_n \mid \bar{x}_{n-1}).$$

Попытаемся выразить тензоры совместных вероятностей через начальные данные. Следуя основным свойствам условных вероятностей, мы можем записать систему линейных уравнений:

$$\begin{cases} p(x_1) = p(x_1 \mid x_n)p(x_n) + p(x_1 \mid \bar{x}_n)(1 - p(x_n)), \\ p(x_2) = p(x_2 \mid x_1)p(x_1) + p(x_2 \mid \bar{x}_1)(1 - p(x_1)), \\ \quad \vdots \\ p(x_n) = p(x_n \mid x_{n-1})p(x_{n-1}) + p(x_n \mid \bar{x}_{n-1})(1 - p(x_{n-1})). \end{cases}$$

ՍԵՀՐԱԳԻՐ. 2006. Հատ. 3. Ս. 1. ISSN 2078-9181 (էջանց.), ISSN 2078-9599 (տպագրված)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru