

ДИАГНОСТИКА СОГЛАСОВАННОСТИ ДАННЫХ В МОДЕЛИ РИСКОВАННОГО ПОВЕДЕНИЯ¹

Торопова А. В., м.н.с., СПИИРАН alexandra.toropova@gmail.com

Аннотация

В работе рассматривается модель рискованного поведения, а также аппарат диагностики согласованности данных, полученных от респондентов, для этой модели. Приводятся примеры с различными данными и при изменении параметров аппарата диагностики. Описывается программное обеспечение, позволяющее удобную работу с диагностикой согласованности данных в модели.

Введение

Диагностики согласованности данных встречается во многих задачах искусственного интеллекта. Многие системы принятия решений и модели строятся с помощью экспертных методов, которые представляют собой комплекс математически-статистических и логических методов и процедур, направленных на получение необходимой для подготовки и выбора рациональных решений информации от специалистов в исследуемой области. Использование информации полученной от экспертов возможно при том, что она представлена в форме, удобной для дальнейших исследований. Следовательно, необходимо не только формализовать поступающие от экспертов сведения, но также оценить их согласованность и достоверность [13, 1, 7]. Подобные задачи возникают и при диагностике неисправностей систем питания [14, 15, 11], диагностике аномалий, связанных с проблемами безопасности распределенных систем [9, 10], а также в других областях. Во всех этих случаях согласованность данных является крайне важной.

Задача

В работах [2, 3, 4, 5] предложена модель оценки интенсивности рискованного поведения. К рискованному поведению можно отнести например вступление в незащищенные половые связи, посещение каких-либо определенных заведений, потребление алкоголя и наркотиков и много другое.

¹Статья содержит материалы исследований, частично поддержанных грантами РФФИ 16-31-00373, 14-01-00580, 15-01-09001

Однако предложенная модель не учитывает степень согласованности исходных данных, полученных от исследуемых, поэтому задача данной работы предложить подход для решения этой проблемы.

Модель рискованного поведения

На рис. 2 представлена модель рискованного поведения $M = (G(V, L), \mathbf{P})$ в виде байесовской сети доверия [6, 12]. Структура модели представлена графом $G(V, L)$, где $V = \{t_{01}, t_{12}, t_{23}, t_{\min}, t_{\max}, \lambda, n\}$ – множество вершин, $L = \{(u, v) : u, v \in V\}$ – множество направленных связей между вершинами. Другими словами, на рис. 2 представлены случайные элементы, входящие в модель, и связи между ними.

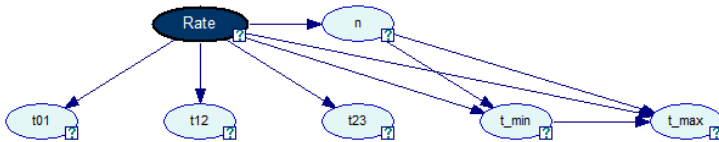


Рис. 1: Модель рискованного поведения

$Rate$ – случайная величина, характеризующая интенсивность поведения, t_{ij} – случайная величина, характеризующая длину интервала между i -ым и j -ым с конца эпизодами, распределена экспоненциально (в предположении, что поведение представляет собой пуассоновский процесс). Кроме того, дополнительную информацию можно получить при включении в модель минимального и максимального интервалов между эпизодами (t_{\min} и t_{\max} соответственно).

Расширение модели рискованного поведения

Предлагается расширить модель, описанную выше, вершинами, отвечающими за отображение оценки согласованности ответов, полученных от респондента, а также вершиной, объединяющей эти и отвечающей за оценку достоверности сведений, полученных от респондента, в целом (см. рис. ??).

Вершины $c_{t_{12}, \min}$ и $c_{t_{23}, \min}$ показывают степень согласованности эпизода t_{ij} с минимальным интервалом t_{\min} , вершины $c_{t_{01}, \max}$, $c_{t_{12}, \max}$ и

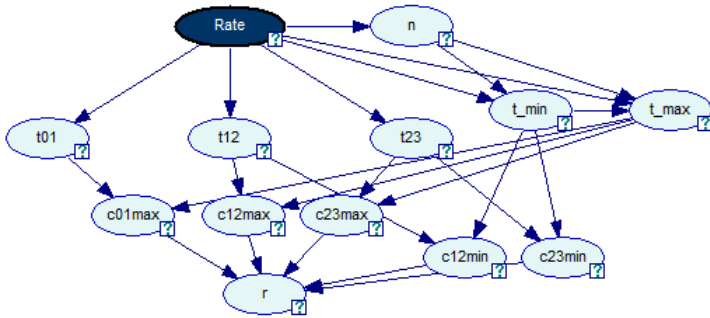


Рис. 2: Расширенная модель рискованного поведения

$c_{t_{23}, \max}$ – эпизода t_{ij} с максимальным интервалом t_{\max} . Оценка согласованности $c_{t_{ij}, \min / \max}$ может принимать значения: $c^+_{t_{ij}, \min / \max}$ (t_{ij} и $t_{\min / \max}$ согласованы), $c^-_{t_{ij}, \min / \max}$ (t_{ij} и $t_{\min / \max}$ не согласованы) и $c^?_{t_{ij}, \min / \max}$ (t_{ij} и $t_{\min / \max}$ находятся в одном и том же интервале). Тензоры условной вероятности, характеризующие переходы к добавленному узлу в общем случае определяются следующим образом:

$$p\left(c_{t_{ij}, \min / \max}^{(s)} | t_{ij}, t_{\min / \max}\right) = \begin{cases} \alpha^{(s)}, & t_{ij} > t_{\min / \max}; \\ \beta^{(s)}, & t_{ij} < t_{\min / \max}; \\ 1 - \alpha^{(s)} - \beta^{(s)}, & t_{ij} = t_{\min / \max}; \end{cases}$$

где $s \in \{+, -, ?\}$, $\alpha^{(s)}, \beta^{(s)} \in [0; 1]$, $\alpha^{(s)} + \beta^{(s)} \leq 1$, $\sum \alpha = 1$, $\sum \beta = 1$. Вершина r характеризует оценку надежности респондента в целом. Чтобы упростить формулы для условных вероятностей, обозначим $c = (c_{t_{12}, \min}, c_{t_{23}, \min}, c_{t_{01}, \max}, c_{t_{12}, \max}, c_{t_{23}, \max})$. Тогда $p(r^+ | c) = \frac{\sum c^+}{\sum c}$, $p(r^- | c) = \frac{\sum c^-}{\sum c}$ и $p(r^? | c) = \frac{\sum c^?}{\sum c}$.

ПО для диагностики согласованности данных в модели рискованного поведения

Для более удобной работы с представленным аппаратом диагностики согласованности данных для модели рискованного поведения было разработано программное обеспечение. При разработке использовались C# и библиотека Smile [8].

В начале работы пользователю предлагается определить интервалы

t_{ij} , t_{\min} и t_{\max} (см. рис. 3).

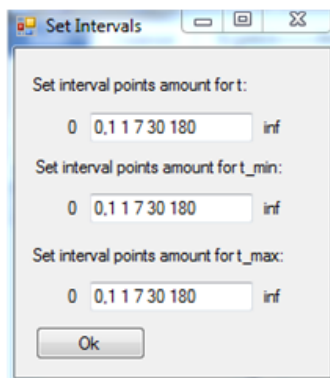


Рис. 3: Окно определения интервалов и добавление к модели аппарата диагностики согласованности данных

После этого можно задать значения параметрам $\alpha^{(s)}$ и $\beta^{(s)}$, $s \in \{+, -, ?\}$ и дополнить модель аппаратом диагностики согласованности данных (см. рис. 4).

На следующем шаге можно вводить данные, полученные от респондентов.

Данные можно ввести вручную (см. рис. 5), с помощью файла электронной таблицы или взять из БД MySQL. В первом случае результаты выводятся в предусмотренное для этого поле (см. рис. 5), во втором – отдельный файл.

Примеры

Для простоты разобьем t_{ij} , t_{\min} и t_{\max} на интервалы, предложенные по умолчанию (см. рис. 3), то есть $(0; 0, 1)$, $[0, 1; 1)$, $[1; 7)$, $[7; 30)$, $[30; 180)$ и $[180; +\infty)$. Задав параметры по умолчанию, то есть $\alpha^+ = 1$, $\alpha^- = 0$, $\alpha^? = 0$, $\beta^+ = 0$, $\beta^- = 1$ и $\beta^? = 0$, мы будем считать, что вероятность согласованности информации, полученной от респондента равна нулю, если данные, в которых она содержится, противоречат друг другу, и единице, если противоречий нет. Данные респондентов возьмем следующие (см. табл. 1):

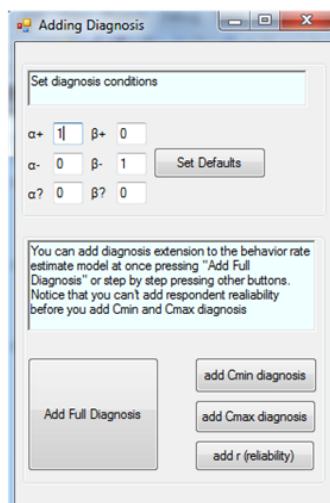


Рис. 4: Окно добавления к модели аппарата диагностики согласованности данных

Результаты применения аппарата диагностики при данных условиях представлены в табл. 2.

Далее с полученными результатами можно действовать по-разному в зависимости от условий конкретной задачи. Кажется логичным исключать из выборки ответы респондентов в тех случаях, когда оценка надежности r^- больше суммы r^+ и $r^?$, в нашем случае это ответы респондента 10. В каких-то случаях могут понадобиться сведения максимально надежных респондентов (3, 6, 7).

Заключение

В данной работе была рассмотрена модель рискованного поведения, было указано, что одним из недостатков данной модели являлось отсутствие аппарата для диагностики согласованности данных, полученных от респондентов, для этой модели. Предложено расширение модели, решающее эту проблему. Для более удобной работы с аппаратом диагностики согласованности данных было разработано программное обеспечение. Приводятся примеры его использования при изменении параметров аппарата и с различными данными.

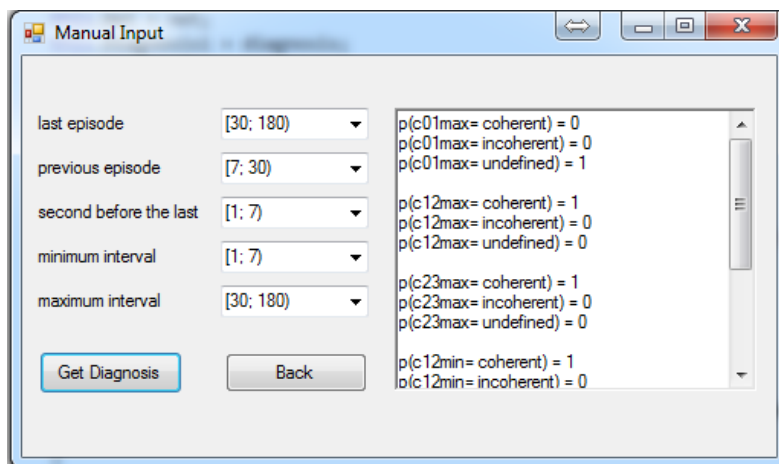


Рис. 5: Окно ввода данных об эпизодах вручную

респондент	t_{01}	t_{12}	t_{23}	t_{\min}	t_{\max}
1	5	5	5	5	5
2	3	0,2	4	0,1	10
3	0,5	4	0,5	0,05	20
4	10	0,5	10	5	60
5	7	1	0,05	0,05	7
6	0,5	0,5	0,5	0,07	180
7	30	10	2	0,05	180

Таблица 1: Данные респондентов

Список литературы

- [1] Каид В.А. Оценка согласованности экспертной информации для моделей принятия решений // Известия Южного федерального университета. Технические науки. 2014. N5 (154). С. 194–202.
- [2] Суворова А.В. Модели и алгоритмы анализа сверхкоротких графулярных временных рядов на основе байесовских сетей доверия. Диссертация на соискание ученой степени кандидата физико-математических наук. 2013

респондент	r^+	r^-	$r^?$
1	0	0	1
2	0,8	0	0,2
3	1	0	0
4	0,8	0,2	0
5	0,6	0	0,4
6	1	0	0
7	1	0	0

Таблица 2: Результаты диагностики согласованности данных

- [3] Суворова А.В. Моделирование социально-значимого поведения по сверхмалой неполной совокупности наблюдений // Информационно-измерительные и управляющие системы. 2013. №9, т. 11. С. 34–38.
- [4] Суворова А.В., Тулупьев А.Л., Сироткин А.В. Байесовские сети доверия в задачах оценивания интенсивности рискованного поведения // Нечеткие системы и мягкие вычисления. 2014. Т. 9, № 2. С. 115–129.
- [5] Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В., Пашенко А.Е. Вероятностные графические модели социально-значимого поведения индивида, учитывающие неполноту информации // Труды СПИИРАН. 2012. Вып. 3 (22). С. 101–112.
- [6] Тулупьев А. Л., Сироткин А. В., Николенко С. И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петербург. ун-та, 2009. 400 с.
- [7] Alonso S., Chiclana F., Herrera F., Herrera-Viedma E., Alcalá-Fdez J., Porcel C. A consistency-based procedure to estimate missing pairwise preference values // International Journal of Intelligent Systems. 2008. No. 23. Pp. 155–175.
- [8] GeNIe& SMILE [дата просмотра 10.03.2016].
- [9] Helouet L., Marchand H., Genest B., Gazagnaire T. Diagnosis from scenarios // Discrete event dynamic systems-theory and applications. 2014. Vol. 24, issue 4. Pp. 353–415.

- [10] Kabiri P., Ghorbani A.A. Research on intrusion detection and response: a survey // International Journal of Network Security. 2005. No. 1(2). Pp. 84–102.
- [11] Mansour M.M., Wahab M.A.A., Soliman W.M. Bayesian networks for fault diagnosis of a large power station and its transmission lines // Electric Power Components and Systems. 2012. Vol. 40, no. 8. Pp. 845–863.
- [12] Perl J. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2000. 400 p.
- [13] Tsyganok V.V., Kadenko S.V. On Sufficiency of the Consistency Level of Group Ordinal Estimates // Journal of Automation and Information sciences. 2010. Vol. 42, issue 8. Pp. 42–47.
- [14] Xiong G., Shi D., Zhu L., Duan X. A New Approach to Fault Diagnosis of Power Systems Using Fuzzy Reasoning Spiking Neural P Systems // Mathematical Problems in Engineering. 2013. Vol. 2013. Article ID 815352.
- [15] Zhu Y., Huo L., Lu J. Bayesian networks-based approach for power systems fault diagnosis // IEEE Transactions on Power Delivery. 2006. Vol. 21, no. 2. Pp. 634–639.