# подходы к диагностике согласованности исходных данных в модели социально-значимого поведения[[1]](#footnote-1)

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

**Аннотация**

В работе рассматриваются два возможных подхода к диагностике согласованности данных, полученных от респондентов, для модели социально-значимого поведения.

Такая диагностика важна, так как при ответе о последних эпизодах поведения респонденты могут ошибиться по причинам работы их памяти или же специально дать заведомо ложные сведения, чтобы их поведение выглядело лучше, чем есть на самом деле.

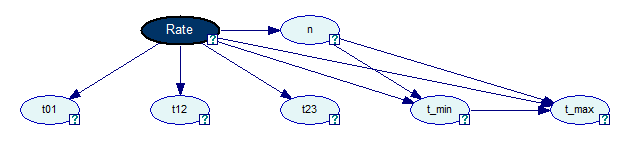
**Введение**

В модели социально-значимого поведения, предложенной в [1–4], используются данные респондентов о последних и минимальном и максимальном эпизодах их поведения. В связи с тем, что исследуемое поведение может быть социально-неодобряемым, респонденты могут давать заведомо ложные данные, чтобы интенсивность их поведения казалась меньшей (или же наоборот большей в случае социально-одобряемого поведения). Также из-за работы памяти респонденты могут несознательно ошибиться в своих ответах.

Для того чтобы оценить степень согласованности ответов респондентов с реальностью и требуется аппарат диагностики. При использовании такого аппарата работа модели социально-значимого поведения сможет давать лучшие результаты и предсказания.

**Модель рискованного поведения**

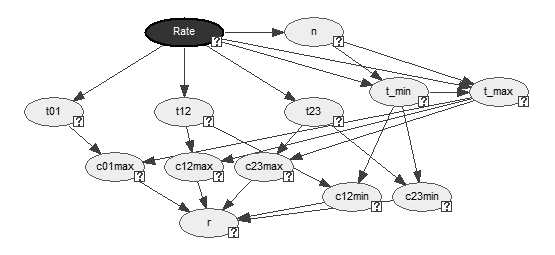
На рис. 1 (здесь и далее изображения созданы в редакторе GeNIe Modeler [7]) представлена модель социально-значимого поведения в виде байесовской сети доверия [5, 6]. Структура модели представлена графом где  – множество вершин,  – множество направленных связей между вершинами. Другими словами, на рис. 1 представлены случайные элементы, входящие в модель, и связи между ними.

  
Рис. 1. Модель социально-значимого поведения

Rate – случайная величина, характеризующая интенсивность поведения, – 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 ( и  соответственно).

**Первый подход к диагностике согласованности**

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. 2).

  
Рис. 2. Расширенная узлами диагностики модель социально-значимого поведения

Вершины  и  показывают степень согласованности эпизода с минимальным интервалом , вершины ,  и  – эпизода  с максимальным интервалом . Мы не рассматриваем , так как она представляет собой интервал между моментом интервью, который не является эпизодом исследуемого поведения, и последним эпизодом поведения. Оценка согласованности  может принимать значения:  ( и  согласованы),  ( и  не согласованы) и  ( и  находятся в одном и том же интервале).

Тензоры условной вероятности, характеризующие переходы к добавленному узлу в общем случае определяются следующим образом:



где , , , , .

Также аналогичным образом рассматриваются оценки согласованности .

Вершина  характеризует оценку надежности респондента в целом. Чтобы упростить формулы для условных вероятностей, обозначим . Тогда ,  и .

**Второй подход к диагностике согласованности**

К описанной модели социально-значимого поведения были добавлены вершины , , ,  и  (см. рис. 3), представляющие интервалы поведения, полученные из ответов респондентов (то же самое, что и , , ,  и  в исходной модели). А , , ,  и теперь скрытые переменные, которые характеризуют действительные последние интервалы поведения. Дело в том, что в ответах респондентов может содержаться неточная или даже заведомо неправильная информация (такое может произойти, например, из-за желания одобрения поведения респондента), то есть реальные интервалы неизвестны, даны только ответы респондентов.

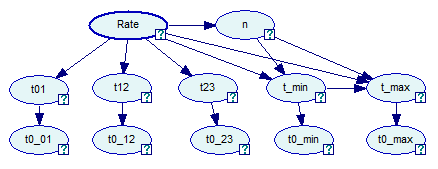


Рис. 3. Расширенная узлами со скрытыми переменными модель социально-значимого поведения

Чтобы рассчитать тензоры условной вероятности было проведено обучение этой байесовской сети доверия. Для этого была использована программа Genie Modeler [7] (рис. 4). На языке R была написана программа, генерирующая 4000 строк сведений о последних эпизодах поведения (данные для вершин , , ,  и ). После удаления строк, в которых оказалось меньше двух эпизодов, осталось 3963 строки. Также для этих значений были сгенерированы оценки интенсивности (для вершины Rate). После этого к данным о последних эпизодах был добавлен нормальный шум, таким образом были смоделированы значения получаемые от респондентов (для вершин , , ,  и ). При обучении был использован алгоритм EM, все вершины, кроме , , ,  и были фиксированы.

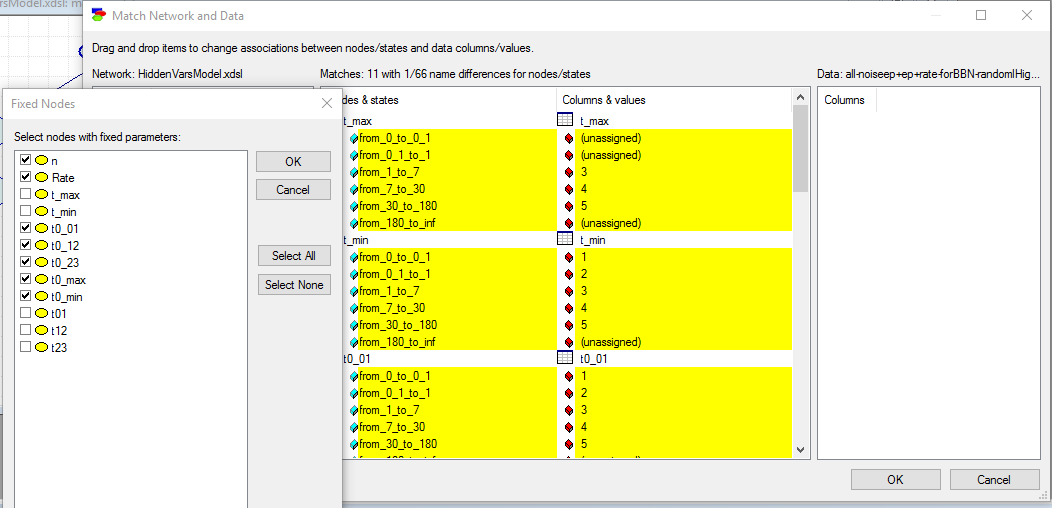


Рис. 4. . Обучение модели в Genie Modeler [7]

Для оценки интенсивности мы взяли дискретизацию на следующие интервалы: [0,0.01), [0.01,0.05), [0.05,0.1), [0.1,1), [1,10) и [10,∞). Для последних эпизодов была использована дискретизация на интервалы [0,0.01), [0.1,1), [1,7), [7,30), [30,180) и [180,∞).

**Заключение**

В работе были рассмотрены два возможных подхода к диагностике согласованности данных, полученных от респондентов, для модели социально-значимого поведения.

Использование одного из предложенных аппаратов для диагностики согласованности данных может улучшить работу и предсказания модели социально-значимого поведения.

**Список литературы**

1. Суворова А.В. Модели и алгоритмы анализа сверхкоротких гранулярных временных рядов на основе байесовских сетей доверия. Диссертация на соискание ученой степени кандидата физико-математических наук. 2013.
2. Суворова А.В. Моделирование социально-значимого поведения по сверхмалой неполной совокупности наблюдений // Информационно-измерительные и управляющие системы. 2013. №9, т. 11. С. 34—38.
3. 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.
4. Суворова А.В., Тулупьева Т.В., Тулупьев А.Л., Сироткин А.В., Пащенко А.Е. Вероятностные графические модели социально-значимого поведения индивида, учитывающие неполноту информации // Труды СПИИРАН. 2012. Вып. 3 (22). С. 101—112.
5. Тулупьев А. Л., Сироткин А. В., Николенко С. И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петерб. ун-та, 2009. 400 с.
6. Perl J. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2000. 400 p.
7. GeNIe& SMILE // URL: http:// http://download.bayesfusion.com/files.html?category=Academia (дата обращения 10.03.2016).

1. В работе представлены результаты исследований, поддержанные грантом РФФИ 16-31-00373 [↑](#footnote-ref-1)