

А.В. СУВОРОВА, Т.В. ТУЛУПЬЕВА, А.Л. ТУЛУПЬЕВ,  
А.В. СИРОТКИН, А.Е. ПАЩЕНКО  
**ВЕРОЯТНОСТНЫЕ ГРАФИЧЕСКИЕ МОДЕЛИ  
СОЦИАЛЬНО-ЗНАЧИМОГО ПОВЕДЕНИЯ ИНДИВИДА,  
УЧИТЫВАЮЩИЕ НЕПОЛНОТУ ИНФОРМАЦИИ**

---

*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.**

**Аннотация.** Представлен подход, позволяющий формализовать задачу оценки интенсивности социально-значимого поведения в терминах вероятностных графических моделей. Сведение этой задачи к разработке особой вероятностной графической модели класса байесовских сетей доверия позволяет воспользоваться уже существующим мощным алгоритмическим аппаратом теории байесовских сетей доверия и свободно распространяемым программным инструментарием для проведения вычислительных экспериментов и для использования построенной модели в практических целях. Описана простейшая модель, основанная на данных об интервалах между эпизодами поведения, предложены варианты ее дальнейшего развития.

**Ключевые слова:** вероятностные графические модели, рискованное поведение, последние эпизоды, байесовские сети доверия.

*Suvorova A.V., Tulupyeva T.V., Tulupyev A.L., Sirotkin A.V., Paschenko A.E.* **Probabilistic graphical models of individual socially significant behavior on the base of incomplete data.**

**Abstract.** We consider a problem of socially significant behavior rate estimates in terms of probabilistic graphical models. Such formal description of the problem allows applying powerful methods and developed algorithms of the theory of Bayesian belief networks. We can use existing software to make computational simulations and apply the model to solve practical tasks. We describe a simple model based on the incomplete data about time intervals between behavior episodes and propose ways of its development.

**Keywords:** probabilistic graphical models, risky behavior, last episodes, Bayesian belief networks.

---

**1. Введение.** Задачи моделирования социально-значимого поведения респондентов и разработки методов оценивания параметров такого поведения возникают во многих отраслях социологических, психологических, маркетинговых исследований [8]. Это такие области, как лечение хронических заболеваний, эпидемиология ВИЧ/СПИД, оценка потребления тех или иных товаров или продуктов. В частности, можно выделить группы потребителей, существенно различающиеся интенсивностью потребления продуктов, товаров или услуг. Аналогичные задачи возникают в системе управления и принятия решений, например для оценки степени удовлетворенности той или иной предоставляемой государственной услугой. В целом, можно сказать, что поставленные вопросы относятся к более широкой, комплексной области

исследований, посвященной анализу и моделированию поведения в контексте интеллектуальных систем [1, 4, 5, 7].

Под интенсивностью поведения — одной из важнейших (но не единственной важной) характеристик поведения — понимается число его эпизодов в определенный промежуток времени. Существует несколько подходов, которые позволяют регистрировать эпизоды поведения различного вида. Мы исследуем подход к обработке результатов интервьюирования / анкетирования респондента об одном или нескольких последних эпизодах его поведения, а также о "рекордных" интервалах между соседними эпизодами. Опросы такого рода позволяют судить о непрерывных количественных величинах — интервалах между эпизодами, а также об интервале между временем опроса и последним эпизодом, однако требуют развития соответствующих математических методов и моделей.

Одним из возможных решений является анализ рассмотренной задачи в рамках подхода, связанного с вероятностно-графическими моделями. Сведение исходной задачи к построению байесовской сети доверия позволит воспользоваться уже существующим мощным алгоритмическим аппаратом теории байесовских сетей доверия.

Цель данной работы — формальная постановка задачи оценки интенсивности и производных характеристик социально-значимого поведения респондентов в терминах теории вероятностных графических моделей и, более узко, в терминах теории байесовских сетей доверия.

**2. Вероятностные графические модели.** Исследования вероятностных графических моделей (ВГМ) [2, 9–14, 19, 20, 22–29] носят междисциплинарный характер — их изучают и используют специалисты ряда научных дисциплин: искусственного интеллекта (ИИ), математической статистики, статистической физики, теоретической информатики, когнитивных исследований. Теория (более точно можно было бы сказать "теории") вероятностных графических моделей является одной из составных частей мягких вычислений [2]. Об особой актуальности и признании особой важности исследований и применения ВГМ говорит тот факт, что Дж. Перл (J.Pearl) — основоположник теории байесовских сетей доверия, одного из видов ВГМ — стал лауреатом 2011 года престижной премии Тьюринга [16].

Вероятностные графические модели, с одной стороны (с точки зрения ИИ), могут рассматриваться как строго формальные математические модели баз фрагментов знаний с неопределенностью (в данном случае важно то, что система знания допускает декомпозицию на небольшие фрагменты, связи между фрагментами достаточно разреженные, а степени доверия представлены оценками вероятности) [9–14], а с другой стороны (с точки зрения чистой математики), как системы

взаимосвязанных случайных элементов [16], причем в зависимости от вида модели связь может быть как направленной, и тогда она воспринимается как причинно-следственная (causal) [29], так и ненаправленной, и тогда она свидетельствует лишь о наблюдаемой взаимной зависимости случайных элементов [9, 12, 23]. Заметим, что вопросами использования вероятности истинности для представления степени доверия к суждению (высказыванию) успешно занимался еще Дж. Буль [18], однако он не ставил, разумеется, вопросы автоматизации вероятностного и логико-вероятностного вывода и не рассматривал сложные модели на основе случайных бинарных либо многозначных элементов.

Различные вероятностные графические модели тесно связаны друг с другом; в ряде случаев одну модель можно трансформировать в другую с сохранением вероятностной семантики [12–14].

С точки зрения локальной цели настоящей работы, задачу об оценке интенсивности социально-значимого поведения [3, 6, 8, 15] удобно свести через ряд промежуточных шагов к разработке особой вероятностной графической модели класса байесовских сетей доверия [20, 22, 24, 28]. При этом следует отметить, что исходная задача может не иметь аналитического решения, либо такое решение будет весьма громоздким, либо потребует долгих поисков представления в "удобной и красивой" форме. Сведение же исходной задачи к построению байесовской сети доверия позволяет воспользоваться уже существующим мощным алгоритмическим аппаратом теории байесовских сетей доверия и свободно распространяемым программным инструментарием (например, [20]) для проведения вычислительных экспериментов, а затем, и для использования построенной модели в практических целях.

**3. Описание модели.** Сначала рассмотрим более общую, чем байесовские сети доверия, модель [19], представленную на рис. 1. В данной модели  $\hat{t}_{12}$  — случайная величина, характеризующая длину интервала между последним и предпоследним эпизодами, она имеет некоторое вероятностное распределение, одним из параметров которого является интенсивность  $\lambda$  поведения каждого респондента. Например, в случае рассмотрения поведения как пуассоновского процесса длины интервалов распределены по экспоненциальному закону, то есть  $\hat{t}_{12} \sim \lambda e^{-\lambda t_{12}}$ . Случайная величина  $\hat{\lambda}$  в свою очередь распределена по некоторому вероятностному закону, например, может быть иметь гамма-распределение  $\hat{\lambda} \sim \text{Gamma}(\alpha, \sigma)$  [3]. Такое распределение отражает априорные предположения об интенсивности социально-значимого поведения респондента. Кроме того, распределение может основыв-

ваться на собранной статистике. Таким образом, случайная величина  $\hat{\lambda}$  с некоторой вероятностью принимает значение, которое затем определяет распределение длины интервала между эпизодами.



Рис. 1. Простейшая модель наблюдаемого проявления рискованного поведения — интервала между двумя его последовательными эпизодами, ближайшими к моменту интервью.

Описанная выше модель основывается на данных только об одном интервале — интервале между последним и предпоследним эпизодами поведения. На рис. 2 представлены более сложные модели, основывающиеся на большем числе величин. Так,  $\hat{t}_{23}$  (рис. 2а) — случайная величина, характеризующая длину интервала между предпоследним и третьим с конца эпизодами, распределена экспоненциально, как и  $\hat{t}_{12}$  (в предположении, что поведение представляет собой пуассоновский процесс). Интервал между моментом интервью и последним эпизодом не является интервалом между эпизодами, поэтому нет оснований считать, что соответствующая случайная величина  $\hat{t}_{01}$  (рис. 2б) распределена также экспоненциально; одним из альтернативных вариантов, как было показано в [3], является бета-простое распределение. Кроме того, возможно использование не только данных о последних эпизодах, но и других величин, в частности, дополнительную информацию можно получить при включении в модель минимального интервала между эпизодами  $\hat{t}_{\min}$  за определенный период времени  $T$  (рис. 2в).

Однако все представленные модели, хотя и относятся к известному классу "причинно-следственных" ВГМ, являются непрерывными. Для каждого сочетания классов распределений в узлах таких моделей пришлось бы особым образом развивать или адаптировать систему алгоритмов вероятностного вывода.

Пожертвовав точностью, чтобы выполнить "дискретизацию" исходных данных о распределении случайных элементов и перейти к тензорам условной вероятности для характеристики связи между узлами, мы окажемся в классе байесовских сетей доверия.

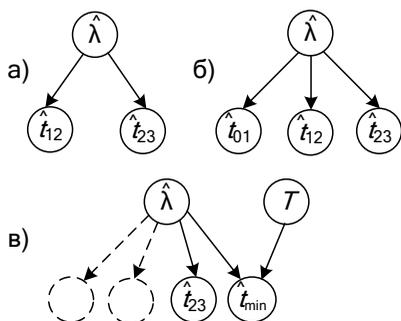


Рис. 2. Более сложные модели, учитывающие скрытые и наблюдаемые переменные, связанные с рискованным поведением индивида

Отметим, что из-за особенностей исходных данных получение точных оценок затруднено даже при использовании непрерывных распределений. Данные об эпизодах поведения респондентов поступают на естественном языке, т.е. являются в значительной степени неточными и неполными. Например, высказывание "некоторое событие произошло месяц назад" в очень редких случаях означает именно "30 дней назад"; чаще всего при таком ответе респондент подразумевает "30 дней плюс-минус неделя".

Чтобы проиллюстрировать результат дискретизации кратко, рассмотрим, что будет происходить с моделью на рис. 1. Соответствующая дискретная модель будет представлена на рис. 3. Узлу со случайной величиной  $\hat{\lambda}$  соответствует вес  $p(\hat{\lambda}) = p(\hat{\lambda} = \hat{\lambda}^{(j)})$  — априорная вероятность попадания интенсивности в один из полуоткрытых промежутков  $[\hat{\lambda}^{(j-1)}; \hat{\lambda}^{(j)})$ , то есть  $p(\hat{\lambda}) = p(\hat{\lambda} \in [\hat{\lambda}^{(j-1)}; \hat{\lambda}^{(j)}))$ ,  $j = 1, \dots, m$ , причем  $\hat{\lambda}^{(0)} = 0$ ,  $\hat{\lambda}^{(m)} = +\infty$ ,  $\hat{\lambda}^{(0)} < \dots < \hat{\lambda}^{(j)} < \dots < \hat{\lambda}^{(m)}$ .

Например, пусть интенсивность имеет следующее априорное распределение:

$$p(\hat{\lambda}) = \begin{cases} 0,1; & \hat{\lambda} = \hat{\lambda}^{(1)} = 0,1, \text{ м.е. } \lambda \in [0; 0,1) \\ 0,2; & \hat{\lambda} = \hat{\lambda}^{(2)} = 1, \text{ м.е. } \lambda \in [0,1; 1) \\ 0,6; & \hat{\lambda} = \hat{\lambda}^{(3)} = 10, \text{ м.е. } \lambda \in [1; 10) \\ 0,1; & \hat{\lambda} = \hat{\lambda}^{(4)} = +\infty, \text{ м.е. } \lambda \in [10; +\infty) \end{cases}$$

Распределение подобного вида может быть получено по статистическим данным либо по предположениям о типе непрерывного распределения случайной величины, характеризующей интенсивность.

Возможные значения  $\hat{t}_{12}$  также разбиваются на дизъюнктивные промежутки  $t_{12}^{(1)}, \dots, t_{12}^{(k)}$ . Например,  $t_{12}^{(1)} = (0; 0, 1)$ ,  $t_{12}^{(2)} = [0, 1; 1)$ ,  $t_{12}^{(3)} = [1; 10)$ ,  $t_{12}^{(4)} = [10; +\infty)$ . Отметим, что дизъюнктивные разбиения  $\hat{\lambda}$  и  $\hat{t}_{12}$  необязательно должны быть одинаковыми, как в данном примере.

Для полной спецификации модели осталось посчитать

$$\begin{aligned}
 p\left(t_{12}^{(i)} \mid \hat{\lambda}^{(j)}\right) &= p\left(t_{12} \in \left[t_{12}^{(i-1)}; t_{12}^{(i)}\right) \mid \lambda \in \left[\hat{\lambda}^{(j-1)}; \hat{\lambda}^{(j)}\right)\right) = \\
 &= E_{\lambda} \int_{t_{12}^{(i-1)}}^{t_{12}^{(i)}} \lambda e^{-\tau \lambda} d\tau = \int_{\lambda^{(i-1)}}^{\lambda^{(i)}} \int_{t_{12}^{(i-1)}}^{t_{12}^{(i)}} \lambda e^{-\tau \lambda} d\tau dF(\lambda),
 \end{aligned}
 \tag{1}$$

где  $F(\lambda)$  зависит от источника наших знаний:

- 1) если мы используем статистические данные, то можно предположить, что внутри каждого промежутка  $\left[\hat{\lambda}^{(j-1)}; \hat{\lambda}^{(j)}\right)$  случайная величина  $\hat{\lambda}$  распределена равномерно;
- 2) если  $\hat{\lambda} \sim \text{Gamma}(\alpha, \sigma)$ , то  $F(\lambda) = \text{Gamma}(\alpha, \sigma)$ ;
- 3) использование другой функции распределения вероятностей случайной величины  $\hat{\lambda}$  в зависимости от предположений о распределении значений интенсивности.

Пусть в результате интервью получен ответ, что интервал между последним и предпоследним эпизодами поведения равен одной неделе. Другими словами, мы получили свидетельство  $t_{12} = 7$ ; в этом случае согласно описанному выше правилу дискретная случайная величина принимает значение  $\hat{t}_{12} = t_{12}^{(3)}$ . По этому свидетельству мы получим апостериорное распределение случайных величин модели.

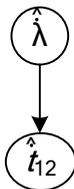


Рис. 3. Дискретная модель

*Примечание: в отличие от рис. 1 здесь случайные величины распределены дискретно, а не непрерывно*

После того, как модель на рис. 3 специфицирована, то есть заданы  $p(\hat{\lambda})$  и  $p(\hat{t}_{12} | \hat{\lambda})$ , алгоритмический аппарат байесовских сетей доверия и его программные реализации обеспечивают возможность учесть поступающее свидетельство.

**4. Заключение.** Даже в описанном простом случае получается интересная модель, хотя при небольших усилиях можно было бы в явном виде применить теорему Байеса. Однако при появлении большего числа узлов за счет большего числа наблюдений, скрытых переменных, необходимости отобразить более сложные связи, сведение исходной задачи к построению модели в виде байесовской сети доверия, а затем — к пропагации поступившего свидетельства или свидетельств, обладает очевидным преимуществом: достаточно выполнить первый этап, а второй будет выполнен существующим программным обеспечением.

## Литература

1. *Виноградов А.Н., Жиликова Л.Ю., Осипов Г.С.* Динамические интеллектуальные системы. Моделирование целенаправленного поведения // Известия Российской академии наук. Теория и системы управления. 2003. № 1. С. 87-94.
2. *Заде Л.* Роль мягких вычислений и нечеткой логики в понимании, конструировании и развитии информационных/интеллектуальных систем // Новости искусственного интеллекта. 2001. № 2–3(44–45). С. 7–11.
3. *Зельтерман Д., Тулупьев А.Л., Суворова А.В., Пащенко А.Е., Мусина В.Ф., Тулупьева Т.В., Красносельских Т.В., Гро Л., Хаймер Р.* Обработка систематической ошибки, связанной с длиной временных интервалов между интервью и последним эпизодом в гамма-пуассоновской модели поведения // Труды СПИИРАН. 2011. Вып. 16. С. 160–185.
4. *Кузнецов О.П.* Интеллектуализация поддержки управляющих решений и создание интеллектуальных систем // Проблемы управления. 2009. № 3. С. 64-72.
5. *Осипов Г.С.* Планирование и моделирование целенаправленного поведения в интеллектуальных системах // Вестник Российского университета дружбы народов. Серия: Прикладная и компьютерная математика. 2002. № 1. С. 51-61
6. *Пащенко А.Е., Суворова А.В., Тулупьева Т.В., Тулупьев А.Л.* Вероятностные распределения порядковых статистик в анализе сверхкоротких нечетких и неполных временных рядов // Труды СПИИРАН. 2009. Вып. 10. СПб.: Наука, 2009. С. 184–207.
7. *Стефанюк В.Л., Жожикашвили А.В.* Наследование свойств категории при переходе от статических систем, использующих знания, к динамическим // Искусственный интеллект и принятие решений. 2008. № 1. С. 4-14.
8. *Суворова А.В., Тулупьев А.Л., Пащенко А.Е., Тулупьева Т.В., Красносельских Т.В.* Анализ гранулярных данных и знаний в задачах исследования социально значимых видов поведения // Компьютерные инструменты в образовании. №4. 2010. С. 30–38.
9. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.

10. *Тудупьев А.Л.* Алгебраические байесовские сети: локальный логико-вероятностный вывод: Учеб. пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 80 с. (Сер. Элементы мягких вычислений).
11. *Тудупьев А.Л.* Алгебраические байесовские сети: глобальный логико-вероятностный вывод в деревьях смежности: Учеб. пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 40 с. (Сер. Элементы мягких вычислений).
12. *Тудупьев 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.
13. *Тудупьев А. Л.* Алгебраические байесовские сети: система операций локального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2009. №4. С. 41–44.
14. *Тудупьев А.Л.* Алгебраические байесовские сети: система операций глобального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2010. №11. С. 65–72.
15. *Тудупьева Т.В., Пащенко А.Е., Тудупьев А.Л., Красносельских Т.В., Казакова О.С.* Модели ВИЧ-рискованного поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.
16. *Ширяев А.Н.* Вероятность: Учеб.пособ. для вузов. 2-е изд. М.: Наука. Гл. ред. физ.-мат. лит., 1989. 640 с.
17. ACM Turing Award Winner, Judea Pearl, United States–2011, For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning // Association for Computer Machinery (web site) URL: [http://amturing.acm.org/award\\_winners/pearl\\_2658896.cfm](http://amturing.acm.org/award_winners/pearl_2658896.cfm) (доступ 05.05.2012).
18. *Boole G.* An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities. Cambridge: Macmillan / London: Walton & Maberly, 1854. (Reprinted in 1951, Dover Publications, New York.)
19. *Cowell R.G., Dawid A.Ph., Lauritzen S.L., Spiegelhalter D.J.* Probabilistic Networks and Expert Systems. Berlin: Springer, 2003. 321 p.
20. *Darwiche F.* Modeling and reasoning with Bayesian networks. Cambridge: Cambridge University Press, 2009. 548 с.
21. GeNIe & SMILE // Decisions systems laboratory. School of Information Sciences. University of Pittsburg. URL: <http://genie.sis.pitt.edu/> (доступ 05.05.2012).
22. *Jensen F.V.* Bayesian Networks and Decision Graphs. NY.: Springer-Verlag, 2001. 268 p.
23. *Kindermann R., Snell J.L.* Markov Random Fields and Their Applications. Providence, RI: Amer. Math. Soc., 1980. 142 p.
24. *Koller D., Friedman N.* Probabilistic graphical models: principles and techniques. Cambridge, MS: The MIT Press, 2009. 1231 p.
25. *Korb K.B., Nicholson A.E.* Bayesian Artificial Intelligence. NY.: Chapman and Hall/CRC, 2004. 364 p.
26. *Neapolitan R.E.* Learning Bayesian Networks. Pearson Prentice Hall, 2003. 674 p.
27. *Nilsson N. J.* Artificial Intelligence: a New Synthesis. San Francisco, CA: Morgan Kaufmann Publ., 1998. 513 p.
28. *Perl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. NY etc.: Morgan Kaufmann Publ., 1994. P. 552.
29. *Perl J.* Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2000. 400 с.

**Суворова Алена Владимировна** — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, аспирант математико-механического факультета Санкт-Петербургского государственного университета (СПбГУ). Область научных интересов: математическая статистика, теория вероятности,

применение методов математического моделирования в эпидемиологии. Число научных публикаций — 21. SuvorovaAV@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

**Suvorova Alena Vladimirovna** — junior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS, PhD student, Faculty of Mathematics and Mechanics of St. Petersburg State University (SPbSU). Research interests: mathematical statistics, probability theory, application of mathematical modeling in epidemiology. The number of publications — 21. SuvorovaAV@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupyev.

**Тулупьева Татьяна Валентиновна** — канд. психол. наук, доцент; старший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ), доцент кафедры психологии управления и педагогики Северо-Западной академии государственной службы (СЗАГС). Область научных интересов: применение методов математики и информатики в гуманитарных исследованиях, информатизация организации и проведения психологических исследований, применение методов биostatистики в эпидемиологии, психология личности, психология управления. Число научных публикаций — 70. TVT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

**Tulupyeva Tatiana Valentinovna** — PhD in Psychology, associate professor; senior researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, SPIIRAS, associate professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU), associate professor, Management Psychology and Pedagogic Department, North-Western Academy of Public Administration (NWAPA). Research interests: application of mathematics and computer science in humanities, informatization of psychological studies, application of biostatistics in epidemiology, psychology of personality, management psychology. Number of publications — 70. TVT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

**Тулупьев Александр Львович** — д-р физ.-мат. наук, доцент; заведующий лабораторией теоретических и междисциплинарных проблем информатики (ТМППИ) СПИИРАН, профессор кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биostatистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 210. ALT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

**Tulupyev Alexander Lvovich** — PhD in Appl. Math. and CS, Dr. Sci. in CS, associate professor; head of laboratory, Theoretical and Interdisciplinary Computer Science Laboratory (TICS Lab), SPIIRAS, professor, Computer Science Department, Faculty of Mathematics and Mechanics, St. Petersburg State University (SPbSU). Research interests: uncertain knowledge and data representation and processing, application of mathematics and computer science in

sociocultural studies, applications of biostatistics and mathematical modeling in modern epidemiology, software technologies and development of information systems with databases. The number of publications — 210. ALT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14<sup>th</sup> Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)3284450.

**Сироткин Александр Владимирович** — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики, С.-Петербургский институт информатики и автоматизации РАН. Область научных интересов: алгебраические байесовские сети; вычислительные аспекты логико-вероятностного вывода в условиях неопределенности, математические методы анализа генома. Число научных публикаций — 64. avs@iias.spb.su; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

**Sirotkin Alexander Vladimirovich** — junior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. Research interests: algebraic Bayesian networks, algorithms of probabilistic-logic inference under uncertainty The number of publications — 64. avs@iias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

**Пашенко Антон Евгеньевич** — научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: математическая статистика, статистическое моделирование, применение методов биostatистики и математического моделирования в эпидемиологии. Число научных публикаций — 45. AEP@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

**Paschenko Anton Evgen'evich** — researcher, Laboratory of Theoretical and Interdisciplinary Computer Science, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: mathematical statistics, statistical modeling, application of biostatistics and mathematical modeling in epidemiology. The number of publications — 45. AEP@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

**Поддержка исследований.** Статья содержит материалы исследований, частично поддержанных грантами РФФИ 09-01-00861-а, 10-01-00640-а, 12-01-00945-а, а также грантом Комитета по науке и высшей школе Правительства Санкт-Петербурга.

Рекомендовано ТИМПИ СПИИРАН, зав. лаб. д-р физ.-мат. наук, доцент А.Л. Тулупьев. Статья поступила в редакцию 23.04.2012.

## РЕФЕРАТ

*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.**

Задачи моделирования социально-значимого поведения респондентов и разработки методов оценивания параметров такого поведения возникают во многих отраслях социологических, психологических, маркетинговых исследований. Это такие области, как лечение хронических заболеваний, эпидемиология ВИЧ/СПИД, оценка потребления тех или иных товаров или продуктов. В частности, можно выделить группы потребителей, существенно различающиеся интенсивностью потребления продуктов, товаров или услуг. Аналогичные задачи возникают в системе управления и принятия решений, например, для оценки степени удовлетворенности той или иной предоставляемой государственной услуги.

Под интенсивностью поведения — одной из важнейших характеристик поведения — понимается число его эпизодов в определенный промежуток времени. Существует несколько подходов, которые позволяют регистрировать эпизоды поведения различного вида. Мы исследуем подход к обработке результатов интервьюирования / анкетирования респондента об одном или нескольких последних эпизодах его поведения, а также о "рекордных" интервалах между соседними эпизодами. Опросы такого рода позволяют судить о непрерывных количественных величинах — интервалах между эпизодами, а также об интервале между временем опроса и последним эпизодом, однако требуют развития соответствующих математических методов и моделей.

Одним из возможных решений является анализ рассмотренной задачи в рамках подхода, связанного с вероятностно-графическими моделями. Сведение этой задачи к разработке особой вероятностной графической модели класса байесовских сетей доверия позволяет воспользоваться уже существующим мощным алгоритмическим аппаратом теории байесовских сетей доверия и свободно распространяемым программным инструментарием для проведения вычислительных экспериментов и для использования построенной модели в практических целях.

В данной работе описана простейшая модель, основанная на данных об интервалах между эпизодами поведения, предложены варианты ее дальнейшего развития. Отмечено, что хотя к изложенной простой модели можно было бы в явном виде применить теорему Байеса, однако при появлении большего числа узлов за счет большего числа наблюдений, скрытых переменных, необходимости отобразить более сложные связи сведение исходной задачи к построению модели в виде байесовской сети доверия, а затем — к пропагации поступившего свидетельства или свидетельств, обладает очевидным преимуществом: достаточно выполнить первый этап, а второй будет выполнен существующим программным обеспечением.

## SUMMARY

*Suvorova A.V., Tulupyeva T.V., Tulupyev A.L., Sirotkin A.V., Paschenko A.E.*  
**Probabilistic graphical models of individual socially significant behavior on the base of incomplete data.**

In many fields of sociological, psychological and marketing research, we face the problem of socially significant behavior rate or frequency estimation on the base of respondents' self-reports about their behavior. The examples of such research include medicine, HIV/AIDS epidemiology, product consumption estimates. For instance, we can distinguish groups of consumers, which have extremely different product consumption rates.

The term behavior rate is determined as a number of behavior episodes happened during the particular period of time. There are several procedures that allow you to get data about different type behavior episodes. We consider a technique based on the data about the last behavior episodes and interval between successive episodes. Interview's results present data about intervals between risky behavior episodes, as well as about whole interval between the moment of interview and the latest episode, and this kind of data requires the development of new mathematical methods and models.

One of possible ways to solve this problem is to consider a problem of socially significant behavior rate estimates in terms of probabilistic graphical models. Such formal description of the problem allows applying powerful methods and developed algorithms of the theory of Bayesian belief networks. We can use existing software to make computational simulations and apply the model to solve practical tasks.

We describe a simple model based on the incomplete data about time intervals between behavior episodes and propose ways of its development. Therefore we can directly obtain the estimates for such simple case, applying the methods of Bayesian belief networks which have special advantages. When we consider more complex relations and as a result include more nodes in the model we just have to rearrange our model and the software does most part of calculations.