

БАЙЕСОВСКИЕ СЕТИ ДОВЕРИЯ: ИНСТРУМЕНТЫ И ИСПОЛЬЗОВАНИЕ В УЧЕБНОМ ПРОЦЕССЕ

Торопова А.В.¹

¹СПИИРАН, Санкт-Петербург, Россия

Аннотация

Байесовские сети доверия благодаря своей наглядности и популярности в различных областях науки являются перспективным математическим и программным инструментом для использования и изучения в высших учебных заведениях. В работе дается небольшое введение в байесовские сети доверия, описываются программные системы для работы с ними. Далее некоторые, наиболее популярные и известные, программные системы более подробно описываются на примере модели оценки интенсивности социально-значимого поведения, являющейся составной частью реальных научно-исследовательских проектов как студентов, так и аспирантов.

Ключевые слова: байесовские сети доверия, БСД.

Цитирование: Торопова А.В. Байесовские сети доверия: инструменты и использование в учебном процессе // Компьютерные инструменты в образовании, 2016. № 4. С. 43–53.

1. ВВЕДЕНИЕ

За последние несколько десятков лет довольно популярным инструментом во многих областях исследования стали байесовские сети доверия. Это обусловлено их возможностями объединения нескольких видов информации, например полученной от экспертов или статистически, а также возможностями работы с неполной или неточной информацией и другими полезными свойствами. Байесовские сети доверия находят применения в медицине, экономике, социологии, психологии, экологии и многих других отраслях [3]. Благодаря своей наглядности, байесовские сети доверия являются хорошим объектом для изучения в высших учебных заведениях, удобным инструментом для реализации творческих научных проектов от уровня лабораторных работ и курсовых до всех видов итоговых и выпускных квалификационных работ.

Байесовская сеть доверия — это вероятностно-графическая модель, представляющая собой ациклический направленный граф (то есть граф, в котором отсутствуют направленные циклы, ненаправленные циклы допускаются), вершины которого представляют собой случайные элементы, а ребра между вершинами — условные зависимости между элементами. Каждый случайный элемент описывается функцией распределения вероят-

ности, случайные элементы могут быть бинарными, многозначными и непрерывными [4, 5, 30].

В данной статье мы рассмотрим программные системы для работы с байесовскими сетями доверия. Наиболее обширный список программного обеспечения, позволяющего работу с байесовскими сетями доверия и другими графическими моделями, составлен Кевином Мерфи [26]. В списке около 70 пунктов, однако многие упомянутые там проекты уже не поддерживаются (например BayesBuilder [8], Elvira [19], XBAIES 2.0 [34]). В последующих разделах мы расскажем о наиболее популярных программных системах, которые поддерживаются и сейчас.

2. ИНСТРУМЕНТЫ С ГРАФИЧЕСКИМ ПОЛЬЗОВАТЕЛЬСКИМ ИНТЕРФЕЙСОМ

Отметим, что одним из достоинств аппарата байесовских сетей доверия является возможность визуализации взаимосвязей между элементами, входящими в модель. Такая особенность приводит к возникновению ряда требований. С одной стороны, необходимы средства визуализации результатов, что в той или иной степени реализовано во многих программных системах. С другой стороны в частности, при обучении работе с байесовскими сетями, являются полезными такие системы, которые позволят создавать и исследовать модель постепенно, шаг за шагом, не требуя (по крайней мере, на первоначальных этапах) знаний о технических деталях. Последняя задача хорошо решается в программных системах с графическим пользовательским интерфейсом, обеспечивающих возможность составления структуры модели «вручную» из блоков определенного вида и указателей связи между блоками. Рассмотрим примеры таких систем.

AgenaRisk [6]

Компания Agena появилась в 1998 году как следствие исследования крупномасштабных байесовских сетей в различных областях. В 2003 году опыт и знания компании были воплощены в AgenaRisk. AgenaRisk — коммерческий проект, но есть и бесплатная версия. Отличие бесплатной версии заключается в отсутствии полной технической поддержки, ограничении функциональности и невозможности запуска на 64-битном процессоре.

BayesiaLab [9]

Очень мощная система, результат почти 20-летней работы. На сегодняшний день компания Bayesia, выпускающая её, является лидирующим на рынке поставщиком ПО для работы с байесовскими сетями. Есть возможность работать через Bayesia Engine API (интерфейс прикладного программирования), библиотеку, реализованную на Java. Недостаток данного комплекса состоит в высокой стоимости, однако можно скачать 30-дневную бесплатную версию (сохранять сети в ней нельзя).

Bayes Server [7]

Bayes Server использует байесовские и темпоральные байесовские сети для таких задач, как классификация, регрессионный анализ, прогнозирование временных рядов, сегментация / кластеризация, оценка плотности, обнаружение аномалий, поддержка принятия решений, вывод, анализ многомерных данных и др. Позволяет работать как с пользовательским интерфейсом, так и с кроссплатформенным API. Первая версия вышла в 2008 году, но до этого разработки велись в течение десяти лет. Это также платное программное обеспечение, но есть два варианта бесплатного использования: Bayes Server Express edition (ограниченное число узлов, случаев, потоков) и Bayes Server Trial edition (каждая сессия до 120 минут, не позволяет сохранять сети). Еще один минус за-

ключается в том, что он работает только с файлами, созданными в нем, и нет возможности работать с моделями, созданными в других программах.

GeNIe Modeler [21]

GeNIe Modeler был разработан в университете Питтсбурга в 1998 году, с 2015 года лицензия на этот продукт принадлежит компании BayesFusion.

GeNIe Modeler («Graphical Network Interface» — графический интерфейс сети) — это пакет программного обеспечения, который может быть использован для создания теоретических моделей. GeNIe Modeler является графическим интерфейсом для SMILE.

SMILE («Structural Modeling, Inference, and Learning Engine» — движок для структурного моделирования, вывода и обучения) является полностью независимой от платформы библиотекой классов C++, реализующей вероятностно-графические модели и модели решений, такие как Байесовские сети, диаграммы влияния и модели структурных уравнений. Его отдельные классы, определенные в SMILE API, позволяют создавать, редактировать, сохранять и загружать графические модели и использовать их для вероятностного вывода и принятия решений в условиях неопределенности.

SMILE поддерживает методологию объектно-ориентированного программирования. Отдельные классы SMILE доступны из C++ или (как функции) из C языков программирования. Благодаря реализации на C, SMILE доступен для практически любых языков и систем. SMILE может быть встроен в программы, использующие вероятностно-графические модели в качестве движка для вывода. SMILE выпущен в виде динамической библиотеки (DLL). Также имеются несколько оберток (SMILE.NET, SMILEX, jSMILE и др.). Большим достоинством GeNIe and SMILE является то, что это бесплатное ПО для научно-исследовательских и обучающих целей. Еще одним плюсом является работа с форматами, поддерживаемыми в других программах (например Hugin Expert и Netica).

Hugin Expert [23]

Hugin был первым программным продуктом для работы с байесовскими сетями Nicholson. Он был разработан и выпущен в Дании в Университете Аальборга в 1989 году. Hugin Expert — это платный продукт. Бесплатная пробная версия имеет ограничения по количеству состояний и максимальному количеству случаев при обучении моделей.

Netica [27]

Netica — это универсальная, быстрая и удобная программа, которую можно использовать, чтобы находить закономерности в данных, создавать диаграммы, кодировать знания, создавать вероятностные экспертные системы и для многого другого. Она подходит для применения в таких областях, как диагностика, прогнозирование, анализ принятия решений, вероятностное моделирование, управление рисками, экспертное построение систем, анализа устойчивости, некоторых других видов статистического анализ и получения данных.

Разработка Netica началась в 1992 году исполнительным директором компании Norsys Б. Борлэйджем, в 1995 году стала доступной для покупки и сейчас широко используется.

Netica — платный продукт, однако у нее есть бесплатная версия с ограничением на размер модели.

3. ПРИМЕР РАБОТЫ С МОДЕЛЬЮ

Из перечисленных программ Agenarisk, Bayes Server, Hugin Expert созданы для использования в коммерческих проектах, поэтому далее рассмотрим подробнее работу с BayesiaLab, GeNIe Modeler и Netica.

В качестве иллюстрации приведем пример, который является составной частью научных исследований как студентов, так и аспирантов, а именно рассмотрим, как будет выглядеть модель для решения задачи оценивания интенсивности социально-значимого поведения по совокупности наблюдений, включающей сведения о длине интервалов между тремя последними эпизодами такого поведения, а также сведения о длине минимального и максимального интервалов за определенный промежуток времени (более подробное описание модели приведено в [1]).

На рис. 1–3 показано, как указанная модель выглядит в различных программах.

В данной модели *Rate* — интенсивность социально-значимого поведения; *t01* — длина интервала между последним эпизодом поведения и моментом интервью; *t12* — длина интервала между последним и предпоследним эпизодами поведения; *t23* — длина интервала между предпоследним и третьим с конца эпизодами поведения; *t_max* — длина максимального интервала между последовательными эпизодами за заданный промежуток времени; *t_min* — длина минимального интервала между последовательными эпизодами за заданный промежуток времени; *n* — число эпизодов поведения за заданный промежуток времени.

Значения случайной величины *Rate* разбиты на следующие дизъюнктные промежутки: [0;0.01), [0.01;0.02), [0.02;0.03), [0.03;0.05), [0.05;0.1), [0.1;0.5), [0.5;1), [1;+∞); значения *t01*, *t12*, *t23*, *t_max* и *t_min* — на следующие: [0;0.1), [0.1;1), [1;7), [7;30), [30;180), [180;+∞). Формулы для вычисления значений тензоров условных вероятностей приведены в [1].

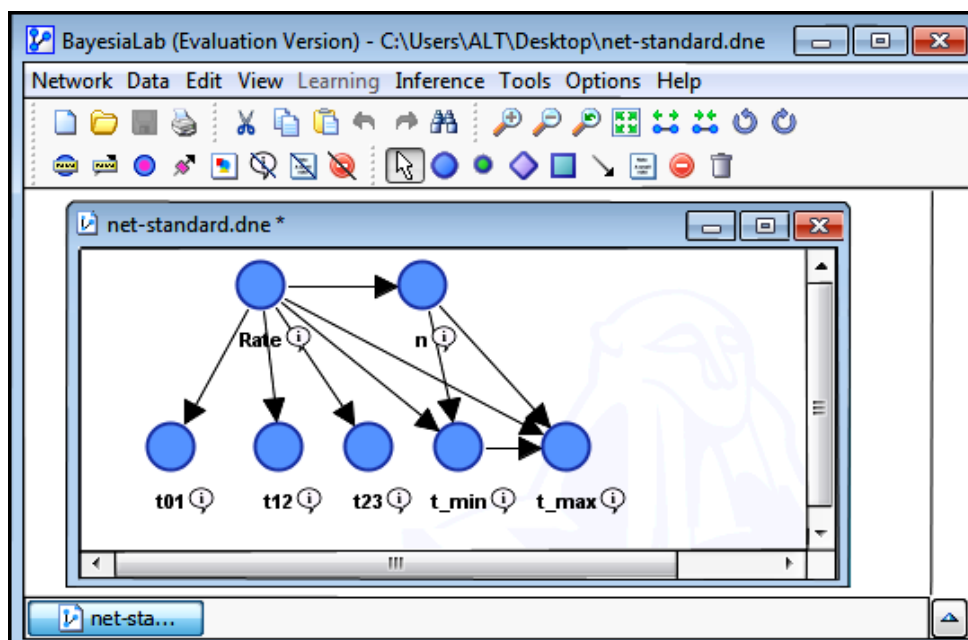


Рис. 1. Модель социально-значимого поведения. Редактор BayesiaLab

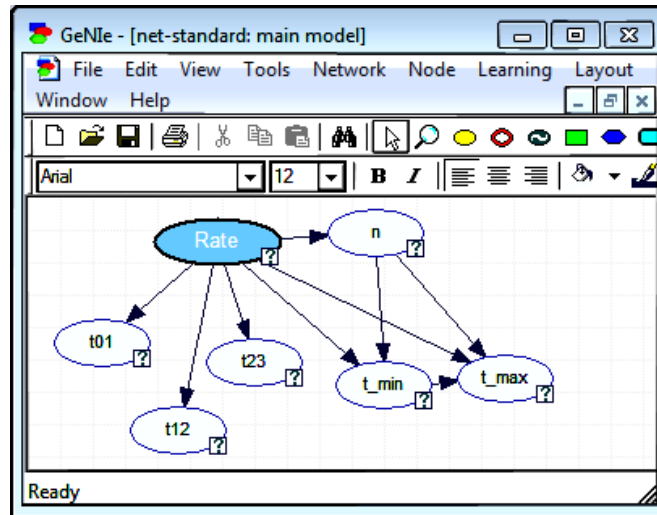


Рис. 2. Модель социально-значимого поведения. Редактор GeNIe Modeler

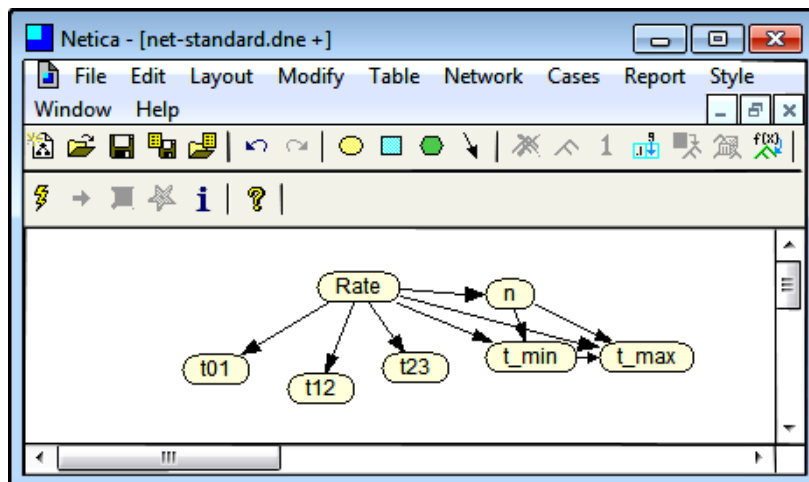


Рис. 3. Модель социально-значимого поведения. Редактор Netica

Внесение этих значений осуществляется в таблицу условных вероятностей, существенной разницы между инструментами нет, однако в BayesiaLab значения, кроме того, можно задать формулой, что является довольно удобной функцией. На рис. 4 показано, как выглядит таблица условных вероятностей в BayesiaLab.

Следующий этап — пересчет вероятностных распределений при поступлении данных. После полного определения тензоров условных вероятностей для всех вершин численные расчеты апостериорного распределения интенсивности социально-значимого поведения при поступлении статистических данных будут выполнены программным обеспечением.

Пусть поступившая информация о последних эпизодах поведения выглядит следующим образом: последний эпизод произошел вчера, предпоследний 15 дней назад, а третий с конца — больше месяца назад. Таким образом, $t_{01} \in [1; 7)$, $t_{12} \in [7; 30)$, $t_{23} \in [30; 180)$.

Node Selection: t01 Rename

States

Probability Distribution

Properties

Classes

Values

State Names

Reference State

Filtered State

Comment

Probabilistic

Deterministic

Tree

Equation

Rate	t01_from_...	t01_from_...	t01_from_...	t01_from_...	t01_from_...	t01_from_...
Lambda_fr...	0.100	0.895	5.766	19.158	57.552	16.530
Lambda_fr...	0.200	1.780	11.084	32.055	52.149	2.732
Lambda_fr...	0.300	2.656	15.986	40.401	40.205	0.452
Lambda_fr...	0.499	4.378	24.654	48.156	22.301	0.012
Lambda_fr...	0.995	8.521	40.825	44.680	4.979	0.000
Lambda_fr...	4.877	34.470	57.633	3.020	0.000	0.000
Lambda_fr...	9.516	53.696	36.697	0.091	0.000	0.000
Lambda_fr...	100.000	0.000	0.000	0.000	0.000	0.000

CompleteNormalizeRandomize

Accept

Cancel

Рис. 4. Таблица условных вероятностей в BayesiaLab

Посмотрим, что получится при введении этих данных (рис. 5–7).

На рисунках видно, что итоговое распределение случайной величины соответствующей интенсивности социально-значимого поведения одинаково во всех рассматриваемых инструментах, отличия только в оформлении (например в отображаемом количестве знаков после запятой). Отметим, что в BayesiaLab и в Netica вершины, которые соответствуют поступившим данным, выделяются цветом, а в GeNIe Modeler — символом рядом с вершиной. Такое выделение позволяет делать выводы о том, на основании каких данных производился перерасчет, имея только итоговое изображение.

Как было отмечено ранее, одним из достоинств инструментов с графическим пользовательским интерфейсом является возможность постепенного изменения модели, например добавления новых вершин в готовую модель. Проиллюстрируем на примере GeNIe Modeler, в остальных рассматриваемых инструментах этот шаг осуществляется сходным образом.

Попробуем добавить к модели вершину, отвечающую за оценку согласованности t_{12} и t_{min} , то есть мы хотим, чтобы при поступлении новых данных, наряду с пересчетом распределения интенсивности, вычислялась и оценка того, насколько поступившие данные согласуются друг с другом (подробности можно посмотреть в [3]). Назовём эту вершину $c12min$. Оценка согласованности $c12min$ может принимать следующие три состояния:

- 1) t_{12} и t_{min} согласованы ($t_{12} > t_{min}$),
- 2) t_{12} и t_{min} не согласованы ($t_{12} < t_{min}$),
- 3) $c12min$ неопределена ($t_{12} = t_{min}$).

Добавим к нашему примеру информацию о том, что минимальный интервал между эпизодами 1 день, то есть $t_{min} \in [1; 7]$. Апостериорное распределение оценки согласованности в GeNIe Modeler представлено на рис. 8.

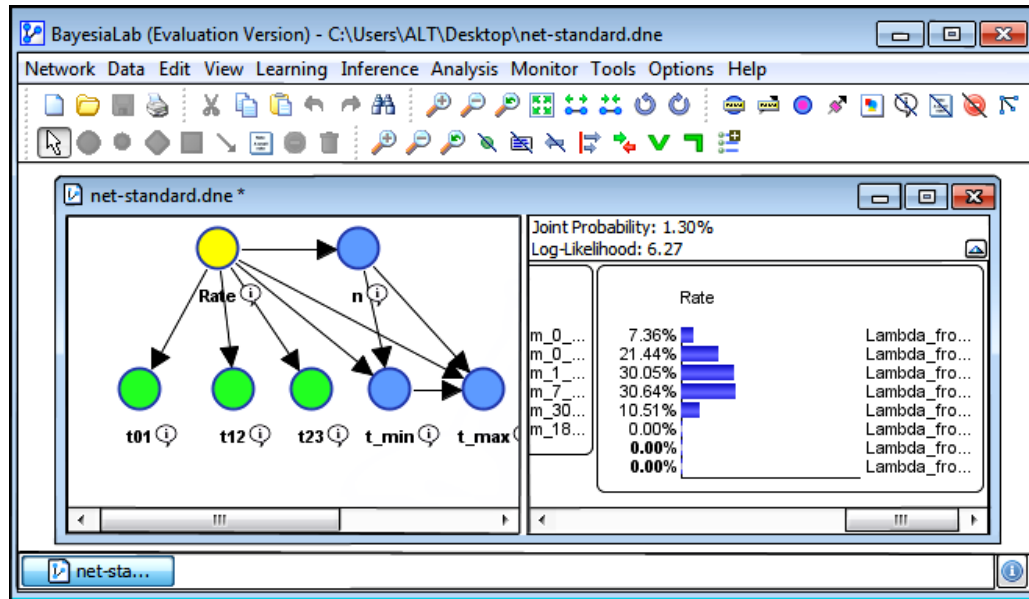


Рис. 5. Апостериорное распределение. Редактор BayesiaLab

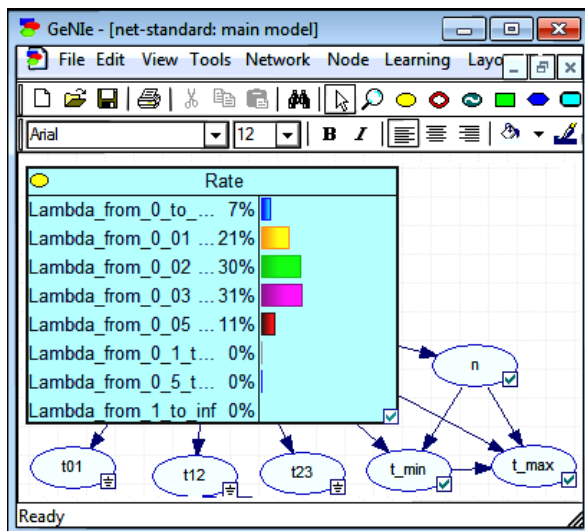


Рис. 6. Апостериорное распределение. Редактор GeNIe Modeler

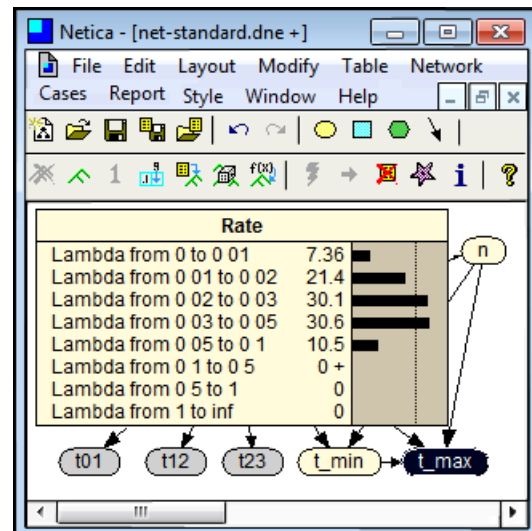


Рис. 7. Апостериорное распределение. Редактор Netica

4. РАСШИРЕННЫЕ ВОЗМОЖНОСТИ

Отметим, что рассмотренные базовые этапы работы с моделью не описывают всю функциональность данных программных систем, их возможности гораздо шире. Так, GeNIe Modeler позволяет осуществлять предварительную обработку данных (например дискретизировать значения), обучать (строить) структуру модели по имеющимся данным, вычислять различные метрики модели. Для автоматизации работы с более сложными сетями имеется также хорошо задокументированная библиотека SMILE, графиче-

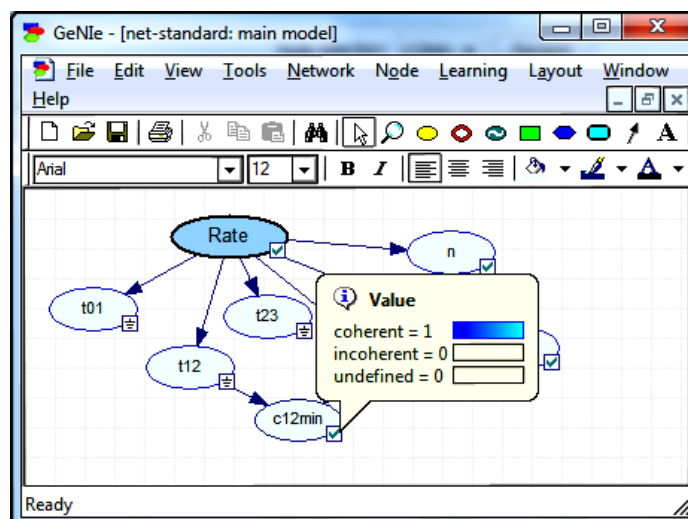


Рис. 8. Апостериорное распределение оценок согласованности. Редактор GeNIe Modeler

ским интерфейсом которой, по сути, и является GeNIe Modeler. Это очень удобно, так как в наличии есть и библиотека для различных языков программирования, и графический интерфейс, позволяющий просто и удобно просмотреть результат работы с библиотекой.

В Netica также есть возможность работать через API, реализованные для Java, C, C#, C++, Visual Basic, Matlab и CLisp. Кроме того у Netica есть следующие функции: генерирование веб-сайтов на основе байесовских сетей; Geo-Netica, добавляющая совместимость с геоинформационной системой; создание пользовательских отчетов по различным аспектам байесовской сети и др.

BayesiaLab поддерживает исследовательский процесс от создания модели до анализа, моделирования и оптимизации. Весь процесс полностью находится в единой среде разработки, которая обеспечивает удобство перемещения между различными элементами исследовательской задачи. Уже было отмечено, что ввод условных вероятностей в BayesiaLab может оказаться намного удобнее, чем в других программах. Кроме того, этот продукт доступен на 5 языках (английском, французском, испанском, китайском и японском).

5. ДРУГИЕ ВОЗМОЖНОСТИ РАБОТЫ С БАЙЕСОВСКИМИ СЕТЯМИ

Кроме рассмотренных программных систем, средства для работы с байесовскими сетями можно найти в рамках более крупных проектов: это и пакеты для R (bnlearn [12], gRain [22], catnet [15], deal [17]), пакеты для работы в Matlab (BNT [13], BNL [11]), библиотеки и фреймворки для Java (Weka [33], Dimple [18]), C++ (OpenGM2 [29], libDAI [25]), Scala (Factory [20], Bayes-Scala [10]), Python (PyMC [31]), C# (Infer.NET [24]) и др.

Специально для реализации полного байесовского статистического вывода был разработан язык вероятностного программирования Stan [32]. Первая версия была выпущена в 2012 году. Stan написан на языке C++, совместим с Unix-подобными ОС, MS Windows и Mac OS X. Это свободное программное обеспечение (лицензия BSD). Есть возможность работать с API, математической библиотекой на C++, использованной в основе языка, содержащей все специальные функции, вероятностные функции, линейную алгебру и

производные, также разработаны интерфейсы RStan (для работы в R), PyStan (для работы в Python), CmdStan (для работы командной в строке) и ShinyStan (пакет для R).

Также стоит упомянуть о проекте BUGS (Bayesian inference Using Gibbs Sampling) [14]. Этот проект начался в 1989 году в Кембридже, в отделении биостатистики совета медицинских исследований, и привел к созданию классической программы BUGS, а затем WinBUGS — ПО, разработанного совместно с Лондонским Имперским колледжем медицины в Сент-Мэри. Сейчас разработка сфокусирована на проекте OpenBUGS.

6. ЗАКЛЮЧЕНИЕ

Как уже отмечалось, программных систем для работы с байесовскими сетями на данный момент существует очень много, у каждой из них есть свои достоинства и недостатки. Приложения с графическим пользовательским интерфейсом очень полезны для иллюстрации различных теоретических аспектов, для демонстрации примеров, для начала работы с аппаратом байесовских сетей доверия. С другой стороны, библиотеки для разных языков программирования позволяют автоматизировать часть работы, осуществить детальную настройку модели, чуть подробнее разобраться в особенностях технической реализации. На наш взгляд, комбинация «графический интерфейс + библиотека» (например GeNIe и SMILE) является полезной именно с точки зрения обучения, в частности, обучения программистов: сначала введение в теорию и построение сетей «вручную», а затем углубление знаний и реализация моделей на одном из изучаемых языков программирования.

Таким образом, обучение будет проходить наиболее наглядно и продуктивно, а обучающие получают возможность довольно быстро включаться в уже существующие проекты, связанные с байесовскими сетями доверия, а также создавать свои.

Список литературы

1. Суворова А.В., Тулупьев А.Л., Сироткин А.В. Байесовские сети доверия в задачах оценивания интенсивности рискованного поведения // Нечеткие системы и мягкие вычисления, 2014. № 2. С. 115–129.
2. Торопова А.В. Подходы к диагностике согласованности данных в байесовских сетях доверия // Труды СПИИРАН, 2015. № 6(43). С. 156–178.
3. Торопова А.В., Суворова А.В., Тулупьев А.Л. Диагностика согласованности в модели для оценивания интенсивности социально-значимого поведения // Нечеткие системы и мягкие вычисления, 2015. Т. 10. № 1. С. 93–107.
4. Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006.
5. 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.
6. Agenarisk URL: <http://www.agenarisk.com/> (дата обращения 20.05.2016).
7. Bayes Server URL: <http://www.bayesserver.com/> (дата обращения 20.05.2016).
8. BayesBuilder URL: http://www.snn.ru.nl/nijmegen/index.php?option=com_content&view=article&id=89&Itemid=212 (дата обращения 20.05.2016).
9. BayesiaLab URL: <http://www.bayesia.com/> (дата обращения 20.05.2016).
10. Bayes-Scala URL: <https://github.com/danielkorzekwa/bayes-scala> (дата обращения 20.05.2016).
11. BNL URL: http://www.downscripts.com/bnl_matlab-script.html (дата обращения 20.05.2016).
12. bnlearn URL: <https://cran.r-project.org/web/packages/bnlearn/bnlearn.pdf> (дата обращения 20.05.2016).
13. BNT URL: <https://code.google.com/p/bnt/> (дата обращения 20.05.2016).

14. BUGS// MRC Biostatistics Unit URL: <http://www.mrc-bsu.cam.ac.uk/software/bugs/> (дата обращения 20.05.2016).
15. catnet URL: <https://cran.r-project.org/web/packages/catnet/catnet.pdf> (дата обращения 20.05.2016).
16. CRAN Task View: Bayesian Inference URL: <https://cran.r-project.org/web/views/Bayesian.html> (дата обращения 20.05.2016).
17. deal URL: <https://cran.r-project.org/web/packages/deal/deal.pdf> (дата обращения 20.05.2016).
18. Dimple URL: <https://github.com/AnalogDevicesLyricLabs/dimple> (дата обращения 20.05.2016).
19. Elvira URL: <http://leo.ugr.es/elvira/> (дата обращения 20.05.2016).
20. Factory URL: <http://factorie.cs.umass.edu/> (дата обращения 20.05.2016).
21. GeNIe Modeler // BayesFusion, LLC. Data Analytics, Mathematical Modeling, Design Support. URL: <http://www.bayesfusion.com/> (дата обращения 20.05.2016).
22. gRain URL: <https://cran.r-project.org/web/packages/gRain/vignettes/gRain-intro.pdf> (дата обращения 20.05.2016).
23. Hugin Expert URL: <http://www.hugin.com/> (дата обращения 20.05.2016).
24. Infer.NET URL: <http://research.microsoft.com/en-us/um/cambridge/projects/infernet/> (дата обращения 20.05.2016).
25. libDAI URL: <https://staff.fnwi.uva.nl/j.m.mooij/libdai/> (дата обращения 20.05.2016).
26. Murphy K. Software Packages for Graphical Models URL: <http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html> (дата обращения 20.05.2016).
27. Netica URL: <http://www.norsys.com/> (дата обращения 20.05.2016).
28. Nicholson A., Korb K. URL: http://www.csse.monash.edu.au/bai/book1e/appendix_b.pdf (дата обращения 20.05.2016).
29. OpenGM2 URL: <http://hci.iwr.uni-heidelberg.de/opengm2/> (дата обращения 20.05.2016).
30. *Perl J.* Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2000.
31. PyMC URL: <http://pymc-devs.github.io/pymc/> (дата обращения 20.05.2016).
32. Stan URL: <http://mc-stan.org/> (дата обращения 20.05.2016).
33. Weka URL: <http://www.cs.waikato.ac.nz/ml/weka/> (дата обращения 20.05.2016).
34. XBAIES 2.0 // Robert Cowell's Personal Home Pages: Software URL: <http://www.staff.city.ac.uk/~rgc/software.html> (дата обращения 20.05.2016).

Поступила в редакцию 24.04.2016, окончательный вариант — 06.06.2016.

Computer tools in education, 2016

№ 4: 43–53

<http://ipo.spb.ru/journal>

BAYESIAN BELIEF NETWORKS: A TOOL FOR EDUCATION USE

Toropova A. V.¹

¹SPIIRAS, Saint-Petersburg, Russia

Abstract

Bayesian belief networks science are a perspective mathematical and software tool for use and study in higher educational institutions because of its clarity and popularity in different scientific research fields. A brief introduction to Bayesian belief network is given, the software systems for working with them are described. Some software systems more popular and famous are described in the details by the example of socially significant behavior model, which is a part of the actual research projects of both students and postgraduates.

Keywords: *bayesian belief networks, BBN.*

Citation: Toropova, A., 2016. "Baiesovskie seti doveriya: instrumenty i ispol'zovanie v uchebnom protsesse" ["Bayesian Belief Networks: a Tool for Education Use"], *Computer tools in education*, no. 4, pp. 43–53.

Received 24.04.2016, the final version — 06.06.2016.

Aleksandra V. Toropova, junior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, 199178, Russia, St. Petersburg, 14-th line V.O., 39, SPIIRAS, alexandra.toropova@gmail.com

Торопова Александра Витальевна,
младший научный сотрудник СПИИРАН,
199178, Россия, Санкт-Петербург,
14-я линия В.О., д. 39, СПИИРАН,
alexandra.toropova@gmail.com

© Наши авторы, 2016.
Our authors, 2016.