

# БАЙЕСОВСКИЕ СЕТИ ДОВЕРИЯ В R <sup>1</sup>

Зернов А. В., студент 3 курса математико-механического факультета СПбГУ по направлению подготовки «Фундаментальная информатика и информационные технологии», стажер лаб. ТиМПИ СПИИРАН, [gzernov@gmail.com](mailto:gzernov@gmail.com);

## Аннотация

Байесовские сети доверия — очень удобный и мощный инструмент, имеющий в данный момент, учитывая накапливаемые объемы знаний, как никогда высокую актуальность. Предложенный проект представляет собой интерактивный справочник по данной теме, позволяющий, в том числе, познакомиться с библиотекой `bnlearn` в языке программирования R. Проект реализован на языке R с использованием библиотек `bnlearn` и `shiny`. Последняя предоставляет возможность сделать справочник наиболее интерактивным и наглядным.

## Введение

Умение пользоваться современными и актуальными инструментами и технологиями является одним из необходимых параметров успешного специалиста. Однако часто возникают трудности в поиске или усвоении необходимой информации, касаемой этих инструментов и технологий.

Байесовские сети доверия — очень удобный и мощный инструмент, имеющий в данный момент, учитывая накапливаемые объемы знаний, как никогда высокую актуальность. Они используются для моделирования в биоинформатике (генетические сети [1], полногеномный поиск ассоциаций [2]), медицине [3], системах поддержки принятия решений [4] и во многом другом.

В работе пойдет речь о приложении, являющимся интерактивным справочником, содержащим базовые понятия теории байесовских сетей и введение в программирование на языке R с возможностью интерактивного взаимодействия с графами. А также будет описан процесс разработки приложения и используемые технологии.

---

<sup>1</sup>Статья содержит материалы исследований, частично поддержанных грантом РФФИ 15-01-09001 — «Комбинированный логико-вероятностный графический подход к представлению и обработке систем знаний с неопределенностью: алгебраические байесовские сети и родственные модели».

## Структура проекта

Проект представляет из себя веб-сайт [5] с возможностью интерактивного взаимодействия. В интерфейсе пользователя для осуществления навигации используется боковая панель, включающая в себя следующие пункты:

- **Overview.** Главная страница, в которой содержится основная информация о назначении и содержании данного сайта.
- **Basics.** Пункт с подразделами, содержащими теоретические основы по графам, программированию на R и упражнения для закрепления материала или просто тренировки.
- **Bayesian Networks Theory.** Пункт, содержащий базовую информацию, связанную с байесовскими сетями доверия.
- **Inference.** Пункт, в котором наглядно можно провести вероятностный вывод или ввести любую свою команду, интерпретируемую R.

После выбора каждого из пунктов в основной части сайта отображается соответствующее содержание.

## Инструменты

В проекте использовались следующие инструменты и технологии:

- **Язык программирования R [6].** Как показывает опрос, проведенный KDnuggets [7] среди занимающихся анализом данных, R лидирует с большим отрывом (рис. 1). В конечном итоге было принято решение остановиться именно на нем.
- **Shiny [8].** Библиотека, доступная через CRAN<sup>2</sup> и позволяющая очень просто визуализировать код, написанный на языке R, в виде интерактивного веб-приложения. Shiny поддерживает множество удобных функций, включая также возможность развертывания в одной из облачных платформ, что делает приложение еще более доступным.

---

<sup>2</sup>CRAN (The Comprehensive R Archive Network) — система для распространения R и дополнительных пакетов.

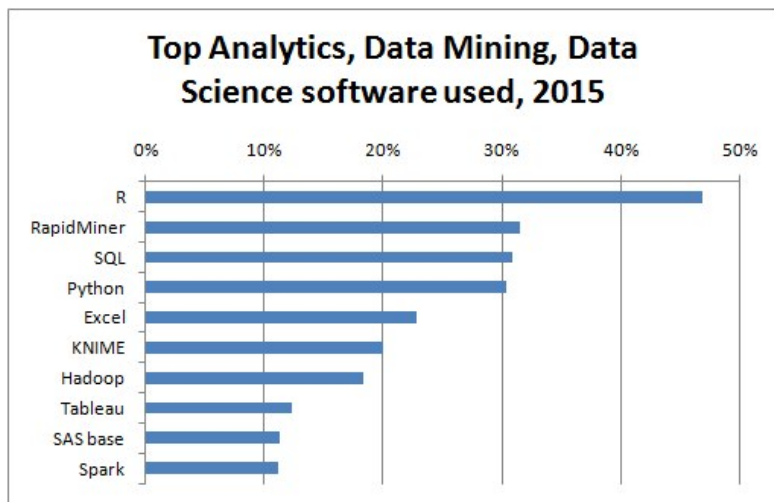


Рис. 1: Результаты опроса KDnuggets.

- **bnlearn [9].** Среди множества библиотек была выбрана bnlearn, обладающая несравненным преимуществом в виде удобной документации. Для более наглядного и широкого сравнения можно воспользоваться таблицей 1 [10, Popular R Packages for Bayesian Network Modeling, p. 25].

## Разработка

Приложение на Shiny состоит из двух основных компонент (однако их можно объединить в один файл в случае небольшого размера приложения):

- **Скрипт пользовательского интерфейса («ui.R»).** В нем описывается расположение элементов управления, доступных пользователю.
- **Серверный скрипт («server.R»).** Содержит функции, связанные с определенными элементами управления. Большинство функций являются «реактивными», то есть вызываются каждый раз, когда изменяются входные данные. Причем не происходит обновления страницы, а меняется лишь конкретный элемент.

	bnlearn	catnet	deal	pcalg	gRbase	gRain
Дискретные данные	+	+	+	+	+	+
Непрерывные данные	+	–	+	+	+	–
Смешанные данные	–	–	+	+	+	–
Обучение (ограничения)	+	–	–	+	–	–
Обучение (оценки)	+	+	+	–	–	–
Обучение (смешанное)	+	–	–	–	–	–
Изменение структуры	+	+	–	–	+	–
Оценка параметров	+	+	+	+	–	–
Предсказания	+	+	–	–	–	+
Вывод	+	–	–	–	–	+

Таблица 1: Сравнение возможностей различных библиотек.

После написания скрипта пользовательского интерфейса и серверного скрипта, было необходимо развернуть приложение на облачной платформе. Такой платформой выступил shiniapps.io [11]. После регистрации на сервисе у каждого пользователя появляется так называемые токен и секрет, необходимые для развертывания приложения. Перейдя в пункт Tools → Global Options → Publishing (для RStudio), можно настроить аккаунт для развертывания, после чего отметить необходимые файлы для загрузки и указать имя приложения.

Можно было выделить два главных направления разработки, в которых приходилось работать. Первое связано с теоретическим материалом, который нужно было предоставить пользователю в доступном виде. Второе — с интерактивным взаимодействием при использовании байесовских сетей.

### *Теоретический материал*

В качестве основного теоретического источника использовалась книга «Bayesian Networks in R with Applications in Systems Biology» [10]. Материал в приложении представлен на английском языке, в связи с тем, что он покрывает наибольшую аудиторию [12].

В самом начале я столкнулся с проблемой набора математических формул, очень хотелось иметь возможность набора такого рода вещей на  $\text{\TeX}$ -подобном языке. С этим справился MathJax [13], после чего для набора математических формул достаточно взять их в знаки « $\$$ », ровно как и в  $\text{\TeX}$ , однако для набора команд, необходимо два знака « $\backslash$ » вместо одного.

Библиотека Shiny предоставляет возможность использования наи-

более используемых HTML-тегов, которые значительно упростили такие задачи, как вывод нумерованного списка или создание гиперссылок.

После краткого введения в теорию графов и язык программирования R пользователю предоставляются на самостоятельную работу несколько упражнений на отработку навыков.

### *Интерактивная работа*

Первая проблема, с которой я столкнулся, заключалась в том, что необходимо динамически генерировать определенные элементы управления в зависимости от каких-то параметров, выбранных пользователем. Нельзя было просто перечислить их в файле «ui.R». Решением этой проблемы стал элемент под названием `uiOutput`. Он связывается с функцией `renderUI`, которая возвращает необходимый элемент управления.

Отдельно хочется отметить, что в разделе **Inference** приложения после создания пользовательского интерфейса с нужными элементами управления для выбранной сети, пользователь имеет возможность выбрать, с какими параметрами выполнить запрос условной вероятности. Сама сеть берется из загруженного файла или готового примера, а событие и свидетельства — вводятся пользователем. Далее генерируется строка кода, где нужные параметры подставлены в нужные места, и выполняется с помощью функции `eval`.

Функция `eval` позволяет реализовать и следующую функцию — своеобразную «командную строку». После импорта или выбора примера пользователь может попробовать некоторые возможности R, введя в ней любую команду.

### Заключение

В результате разработано приложение, которое позволяет освоить основы теории графов, познакомиться с языком программирования R, а также испробовать его возможности при работе с байесовскими сетями.

## Список литературы

- [1] Friedman N., Linial M., Nachman I., Pe'er D. Journal of Computational Biology. July 2004, 7(3-4): 601-620. doi:10.1089/106652700750050961.
- [2] Learning genetic epistasis using Bayesian network scoring criteria URL: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-89>
- [3] Genetic Counseling and Cancer Risk Modeling: An Application of Bayes Nets. URL: [http://www.john-uebersax.com/stat/bayes\\_net\\_breast\\_cancer.doc](http://www.john-uebersax.com/stat/bayes_net_breast_cancer.doc)
- [4] DIAVAL, a Bayesian expert system for echocardiography. URL: <http://www.sciencedirect.com/science/article/pii/S0933365797003849>
- [5] BNetworks. URL: <https://zernov.shinyapps.io/bnetworks/>
- [6] The R Project for Statistical Computing. URL: <https://www.r-project.org/>
- [7] R leads RapidMiner, Python catches up, Big Data tools grow, Spark ignites. URL: <http://www.kdnuggets.com/2015/05/poll-r-rapidminer-python-big-data-spark.html>
- [8] A web application framework for R. URL: <https://cran.r-project.org/web/packages/shiny/>
- [9] R package for Bayesian network learning and inference . URL: <http://www.bnlearn.com/>
- [10] Nagarajan R., Scutari M., Lèbre S. Bayesian Networks in R. — United States: Springer, 2013. 157 c.
- [11] Share your Shiny Applications Online. URL: <http://www.shinyapps.io/>
- [12] Usage of content languages for websites. URL: [http://w3techs.com/technologies/overview/content\\_language/all](http://w3techs.com/technologies/overview/content_language/all)
- [13] A JavaScript display engine for mathematics that works in all browsers. URL: <https://www.mathjax.org/>