

УДК 28.17.19

doi: 10.17586/2226-1494-2020-20-6-835-840

## ВЫВОД ГЕННЫХ РЕГУЛЯТОРНЫХ СЕТЕЙ ПО ДАННЫМ ЭКСПРЕССИИ ГЕНОВ ПРИ ПОМОЩИ БАЙЕСОВСКИХ СЕТЕЙ

А.А. Лобода, А.А. Сергушичев

Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация  
Адрес для переписки: alserg@itmo.ru

### Информация о статье

Поступила в редакцию 02.10.20, принята к печати 31.10.20  
Язык статьи — русский

**Ссылка для цитирования:** Лобода А.А., Сергушичев А.А. Вывод генных регуляторных сетей по данным экспрессии генов при помощи байесовских сетей // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 6. С. 835–840. doi: 10.17586/2226-1494-2020-20-6-835-840

### Аннотация

**Предмет исследования.** Рассмотрена задача вывода генных регуляторных сетей в форме байесовских сетей из данных экспрессии генов как задача оценки частной вероятности вхождения каждого ребра в истинную байесовскую сеть при известных уровнях экспрессии генов. Для решения задачи предложен метод, использующий подход Монте-Карло на основе марковских цепей. **Метод.** Предлагаемый метод состоит в сэмплировании пар из байесовской сети и дискретизационной политики, позволяющей применять сеть к данным экспрессии генов согласно апостериорному распределению. Для сэмплирования используется подход Монте-Карло на основе марковских цепей с реализацией с помощью алгоритма Метрополиса–Гастингса. На основе полученной выборки выполняется оценка искомых вероятностей. **Результаты.** Предложенный метод протестирован на симулированных данных из соревнования DREAM4 Challenges. Сравнение с лидерами показало, что качество разработанного метода на некоторых тестах превосходит лидера из существующих методов – метод regularized gradient boosting machines (RGBM) — и сравнимо на остальных тестах. В то же время метод является довольно гибким и позволяет адаптировать его к другим видам экспериментальных данных. **Практическая значимость.** Метод может быть использован в вычислительной биологии для изучения механизмов регуляции генов в различных процессах, в том числе в развитии опухолей или работе иммунной системы.

### Ключевые слова

генные регуляторные сети, байесовские сети, дискретизация, методы Монте-Карло, марковские цепи

doi: 10.17586/2226-1494-2020-20-6-835-840

## INFERRING OF REGULATORY NETWORKS FROM EXPRESSION DATA USING BAYESIAN NETWORKS

A.A. Loboda, A.A. Sergushichev

ITMO University, Saint Petersburg, 197101, Russian Federation  
Corresponding author: alserg@itmo.ru

### Article info

Received 02.10.20, accepted 31.10.20  
Article in Russian

**For citation:** Loboda A.A., Sergushichev A.A. Inferring of regulatory networks from expression data using Bayesian networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 6, pp. 835–840 (in Russian). doi: 10.17586/2226-1494-2020-20-6-835-840

### Abstract

**Subject of Research.** The paper considers the inferring of gene regulatory networks in the form of Bayesian networks from gene expression data. We present this problem as the problem of the marginal probability estimation for each edge appearance in the true Bayesian network under the known gene expression levels. Monte Carlo approach based on the Markov chains is proposed. **Method.** The proposed method involved the sampling of Bayesian network pairs and a discretization policy, providing a way for the network to be applied to continuous gene expression data according to a posteriori distribution. The Markov chain Monte Carlo approach was used for sampling with implementation via the Metropolis-Hastings algorithm. Then, the desired probabilities were estimated based on the obtained sample. **Main**

**Results.** The proposed method is tested on simulated data from the DREAM4 Challenges. Comparison with the leaders shows that the developed method quality surpasses the leader among the existing methods, the regularized gradient boosting machines method (RGBM), on some tests and is comparable to the others in view of the results. At the same time, the proposed method is flexible enough and can be adapted to the other types of experimental data. **Practical Relevance.** The method is applicable in computational biology for research of the gene regulation mechanisms in various processes, including the tumor growth or the immune system operation.

**Keywords**

gene regulatory networks, Bayesian networks, discretization, Monte-Carlo methods, Markov chains

**Введение**

Вывод генных регуляторных сетей из данных экспрессий генов — это одна из фундаментальных задач современной биоинформатики. Ее решение позволит лучше понимать процессы в живых клетках, определять потенциальные точки воздействия для лекарств от различных заболеваний [1] и идентифицировать новые метаболические пути.

Одной из наиболее совершенных моделей динамики уровней экспрессии генов является та, в которой уровни экспрессии представлены решением системы дифференциальных уравнений [2]. Однако в силу высокого шума в данных и большого числа параметров для настройки модели обычно проблематично построение точной модели этого процесса.

Другим подходом является построение генных регуляторных сетей (ГРС). Обычно в таких методах не происходит попытка полностью спрогнозировать развитие системы. При этом первостепенное внимание уделяется графической модели, в которой вершинами графа являются гены, а направленное ребро между ними означает прямое влияние экспрессии одного гена на другой. Для решения этой задачи предложено множество методов [3]. Последние исследования показывают, что методы на основе отбора признаков дают хорошие результаты и превосходят методы, основывающиеся на других подходах.

Несмотря на то, что байесовские сети были изобретены для поиска причинных отношений между случайными переменными, что по своей природе схоже с задачей вывода ГРС, методы [4], основанные на этом подходе, не лишены недостатков. Наиболее существенным из них остается вычислительная сложность поиска оптимальной байесовской сети. Для решения это проблемы предлагаются следующие решения: использование эвристик [5], ограничение максимального числа входящих в вершину ребер в искомой сети [6], решение задачи для небольших подмножеств с последующим их объединением [7].

В настоящей работе описан способ реконструкции генной регуляторной сети, которая использует модель байесовских сетей. Показано, что нет необходимости в поиске наиболее оптимальной из них; скорее, важно иметь хороший набор байесовских сетей для оценки вероятностей причинно-следственных связей между экспрессиями каждой пары генов. Предложено решение для задачи, которое использует подход Монте-Карло на основе марковских цепей, и выполнено сравнение его с другими известными методами по выводу ГРС.

**Формальные определения**

Дадим определения понятиям, используемым на протяжении всей работы.

**Определение 1.** Пусть  $G = (V, E)$  — направленный граф. *Реберное ранжирование*  $G$  — перестановка множества всех ребер  $E$ . Для ранжирования  $r = (r_1, r_2, \dots, r_{|E|})$  будем говорить, что ребра в начале списка  $r$  (например,  $r_1, r_2, \dots$ ) более важные и отранжированы выше, чем ребра в его конце (например,  $r_{|E|-1}, r_{|E|-2}, \dots$ ).

**Определение 2.** *Байесовская сеть* — набор  $B = (X, G, \Theta)$ , где  $G = (V, E)$  — направленный ациклический граф с множеством вершин  $V$ ;  $X = (X_v)$  — случайный вектор, проиндексированный значениями из  $V$ ;  $E$  — набор упорядоченных пар  $(v, u) \in V \times V$ , таких, что каждая случайная переменная  $X_u, u \in V$  зависит только от случайных величин с индексами из множества родителей  $\Pi_G(u) = \{v | (v, u) \in E\}$  и условно не зависит от переменных с индексами вне множества  $V \setminus \Pi_G(u)$ .

**Определение 3.** Пусть  $X$  — вещественная случайная величина и  $x$  — возможный исход  $X$ . *Дискретизационной политикой*  $\Lambda_X = \{e_1, e_2, \dots, e_K\}$ , как определено в [8], называется отображение из  $\mathbb{R}$  в  $\{1, 2, \dots, K + 1\}$  такое, что:

$$\Lambda(x) = \begin{cases} 1, & \text{если } x < e_1 \\ i, & \text{если } e_{i-1} < x < e_i \\ K + 1, & \text{иначе.} \end{cases}$$

**Определение 4.** *Конфигурацией* случайного вектора  $X = (X_1, \dots, X_n)$  будем называть элемент его области определения  $w \in D(X)$ .

**Определение 5.** *BDeu оценка* — совместная вероятность  $P(B, T)$  байесовской сети  $B = (X, G, \Theta)$  и набора данных  $T$  при использовании допущений, описанных в [9], равное

$$P(B, T) = P(B) \cdot \prod_{i=1}^n \prod_{j=1}^q \frac{N'}{q_i} \cdot \prod_{k=1}^r \frac{\Gamma\left(N_{ijk} + \frac{N'_{ri}}{q_i}\right)}{\Gamma\left(N_{ij} + \frac{N'}{q_i}\right) \Gamma\left(\frac{N'}{r_i q_i}\right)},$$

где  $\Gamma$  — гамма-функция;  $n = |\{X_i\}|$  — число переменных байесовской сети  $B$ ;  $r_i = |D(X_i)|$ ;  $q_i = |W_i|$ ;  $P(B)$  — априорная вероятность на структуру сети.

Пусть  $A^{(i)} = \{k | X_k \in \Pi_G(X_i)\}$  — набор индексов, соответствующий родительским переменным переменной  $X_i$  в графе  $G$  байесовской сети  $B$ , тогда  $N_{ijk}$  — число экземпляров  $y = (y_1, \dots, y_n)$  множества  $T$  таких, что  $y^{(i)} \dots y^{(i)} = w_{ij}$  и  $y_i = x_{ik}$ . Тогда  $N_{ij} = \sum_{k=1k} N_{ijk} \cdot N$  — гиперпараметр оценочной функции (мнимый размер выборки).

### Теоретическое обоснование

Приведем теоретическое обоснование описываемого метода, использующего модель байесовских сетей для вывода значимости каждого регуляторного воздействия.

Для начала введем набор случайных переменных:  
 —  $M$  — переменная, представляющая модель;  
 —  $D$  — переменная, представляющая данные;  
 —  $\Lambda$  — переменная, представляющая дискретизационную политику;  
 — набор переменных  $\{X_e | e \in A\}$ , обозначающий присутствие ребра  $e$  в модели.

В данной работе выберем байесовские сети в качестве модели для регуляторной сети. Таким образом,  $M$  имеет структуру направленного ациклического графа (НАГ), что в результате сводит основную задачу к поиску значения вероятностей для каждого ребра присутствия в модели при известных данных, или  $P(X_e|D)$ . Исходя из вышесказанного:

$$P(X_e|M) = \begin{cases} 1, & \text{если } e \in M, \\ 0, & \text{иначе.} \end{cases} \quad (1)$$

Пусть  $M$  — множество возможных НАГ. Применяя правило полной вероятности ко всем возможным моделям, имеем:

$$P(X_e|D) = \sum_{M \in M} P(X_e|M)P(M|D) = E_{M \in M}[(e \in M)|D]. \quad (2)$$

Однако обычно данные экспрессии представлены в виде набора наблюдений вещественных переменных. Исходя из определения дискретизационной политики  $\Lambda$  как случайной переменной, можно применить правило полной вероятности еще раз:

$$P(M|D) = \sum_{\Lambda} P(M|D, \Lambda)P(\Lambda).$$

Подставив выражение (2) в формулу (1) при выносе суммы по дискретизациям получим:

$$P(X_e|D) = \sum_{M \in M} \sum_{\Lambda} P(X_e|M)P(M|D, \Lambda)P(\Lambda) = E_{M, \Lambda}[(e \in M)|D].$$

Таким образом, если провести семплирование переменных  $M$  и  $\Lambda$  согласно их совместному распределению, то можно получить несмещенную оценку  $P(X_e|D)$ .

### Семплирование

Покажем, как эффективно семплировать случайные переменные  $M$  и  $\Lambda$ . Заметим, что данные переменные могут быть семплированы независимо согласно

$$P(M, \Lambda|D) = P(M|\Lambda, D)P(\Lambda|D). \quad (3)$$

Сначала может быть семплирована дискретизационная политика  $\Lambda$ , а затем — сгенерирована байесовская сеть  $M$  с априорно заданной политикой. Предположим, что семплирование дискретизационной политики может быть реализовано эффективно для широкого класса априорных распределений, и оставим этот вопрос за рамками работы. Семплирование байесовских сетей пропорционально их вероятностям, является более

сложным вопросом и будет разобрано в следующем разделе.

Применим *BDeu* оценку для исходной задачи. Из определения оценки ее значение — это совместная вероятность дискретных данных и байесовской сети. Дискретные данные выводятся из вещественных путем их дискретизации согласно выбранной политике. Пусть  $L$  — дискретный набор данных, полученный применением политики к набору данных  $D$ . Ясно, что:

$$P(M, L|D, \Lambda) = P(M|D, \Lambda),$$

поскольку  $L$  является лишь результатом применения  $\Lambda$  к  $D$ , т. е. существует только один возможный набор данных при условии фиксированных  $\Lambda$  и  $D$ .

Правая часть выражения (3) может быть посчитана непосредственно исходя из определения *BDeu* оценки, и ее эффективное вычисление описано далее.

### Алгоритм Метрополиса–Гастингса

Для семплирования экземпляров из распределения  $P(M|D)$  применим подход Монте-Карло на основе марковских цепей (Monte Carlo Markov Chain, MCMC). Для этого сначала необходимо определить состояния соответствующей марковской цепи и переходы в ней. В качестве состояний рассмотрим все возможные байесовские сети. В качестве переходов будем использовать следующие операции над сетями:

- добавление направленного ребра в граф;
- удаление существующего направленного ребра;
- смена направления присутствующего в графе ребра.

Для MCMC семплирования воспользуемся методом Метрополиса–Гастингса [10, 11]. Стартовым состоянием в рамках этого метода может быть либо пустой граф, либо случайный направленный ациклический граф. Выполним попытку применить операции, описанные выше. Для этого равновероятно выберем упорядоченную пару вершин  $(v, u)$ , и в случае наличия соответствующего ребра, выполним попытку его удаления или смены направления. В противоположной ситуации ребро может быть добавлено. Путем подстановки предлагаемых переходных вероятностей в метод Метрополиса–Гастингса получим вероятности принятия новых состояний:

$$A(S'|S) = \min \left( 1, \frac{P(S') g(S|S')}{P(S) g(S'|S)} \right), \quad (4)$$

где  $g$  — переходная вероятность;  $P$  — вероятность состояния  $S$ . В соответствии с переходной вероятностью легко видеть, что:

$$\frac{g(S')}{g(S)} = \begin{cases} 1, & \text{при попытке сменить направление ребра,} \\ 0,5 & \text{при попытке добавить ребро,} \\ 2,0 & \text{при попытке удалить ребро.} \end{cases}$$

В (4) используем  $P(S') = P(M'|D)$  для состояний в случае, если  $M'$  представляет собой направленный ациклический граф. В противном случае приравняем вероятность полученной модели к нулю, так как полученный граф не соответствует корректной байесовской сети.

Вследствие применения этих операций марковская цепь постепенно сходится к желаемому распределению.

Теперь можно получить несколько экземпляров  $\mathbb{S} = \{(M_i, \Lambda_i)\}$  и оценить значение вероятности  $P(X_e|D)$ :

$$P(X_e|D) \approx \frac{|\{M \in \mathbb{S} | e \in M\}|}{|\mathbb{S}|}$$

### Детали реализации

Поскольку алгоритм сходится после большого числа итераций, увеличение производительности в его критических секциях будет иметь огромный эффект на производительность всего алгоритма.

Таких критических мест в алгоритме два: вычислительные *BDeu* оценки и проверка графа на ацикличность.

1. Для каждой переменной используем кеш фиксированного размера для хранения вычисленных значений *BDeu* оценки для ограниченного набора множеств родителей вершины.
2. Используем подход *встреча-в-середине* [12], чтобы уменьшить время проверки на ацикличность. Это требует поддержки множества обратных ребер в графе. Хотя динамические графы могут быть применены в этой задаче для уменьшения алгоритмической сложности алгоритма, на практике это не приносит выгоды из-за небольших размеров типичных для этой задачи графов.

В предлагаемом методе есть ряд параметров, которые должны быть установлены перед запуском программы, реализующей алгоритм:

- так называемый мнимый размер выборки  $N'$  в оценке *BDeu*;
- априорное распределение байесовских сетей;
- априорное распределение дискретизационных политик.

Определено, что на практике изменение мнимого размера выборки не приводит к существенно отличающимся результатам. В связи с этим приравняем данный параметр к единице, что является общепринятым стандартным значением для этого параметра.

Для сетей, имеющих отношение к биологии, зачастую используются так называемые безмасштабные сети [13]. Исходя из этого, априорная вероятность байесовских сетей вычисляется в соответствии с мультиномиальным тестом на исходящие степени вершин

```
function get_ranking(D, n_samples, mcmc_iterations):
    S = {}
    for i in [0, n_samples):
        L ← random_discretization_policy(D)
        L ← Λ(D)
        B ← empty_graph(|D|)
        for j in [0, mcmc_iterations):
            (v, u) ← random_edge(B)
            op ← random_operation((v, u), B)
            B̄ ← apply(B, op, (v, u))
            A ← acceptance_probability(B, B̄, L)
            if random_uniform() ≤ A:
                B ← B̄
        S = {S, B}
    for e in E(G):
        p(e) ← |\{e ∈ s for s in S\}|/|S|
    return p
```

Листинг. Предлагаемый метод

в графе, использующимся при проверке сети на безмасштабность.

В работе [4] отмечено, что разделение вещественной переменной на три части дает лучшие результаты при применении байесовских сетей. Таким образом, для дискретизационной политики выберем равномерно случайно два наблюдения переменной, по которым и будем разделять вещественную переменную.

Псевдокод алгоритма приведен в листинге.

### Результаты экспериментов

Для оценки качества предлагаемого метода применим результаты экспериментов на экземплярах задачи из DREAM4 Challenges [3]. Поскольку в настоящей работе не предлагается способа интерпретации экспериментов нокаута генов, метод был протестирован только на наборе данных 100\_multifactorial. Выполнено сравнение предлагаемого метода с наиболее результативным после анализа литературы методом RGBM (Regularized Gradient Boosting Machines), описанным в [14], и победителем соревнований DREAM4 Challenges GENIE 3 [15].

Результаты анализа точности и чувствительности результатов рассмотренных методов представлены на рисунке. Они наглядно показывают, что вероятностный метод дает ранжирование по крайней мере не хуже, чем алгоритмы, основанные на выборе признаков. Более того, из рисунка можно увидеть, что предложенный метод дает лучшие результаты в начале ранжирования, что является важным при экспериментальной проверке гипотез в лаборатории. При этом площади под кривой остаются близкими по значению.

### Заключение

В работе представлен полностью вероятностный метод для решения задачи вывода генных регуляторных сетей из данных экспрессии генов путем вычисления оценки на частные вероятности вхождения каждого ребра в искомую генную регуляторную сеть, представленную в виде байесовской сети. Этот подход не требует вычисления крайне трудных задач оптимизации, как в случае с нахождением оптимальной байесовской сети или лучшей дискретизационной политики. Показано, что метод может быть реализован посредством применения метода Монте-Карло на основе марковских цепей. Предложенный метод протестирован на симулированных экземплярах задачи из соревнования DREAM4 Challenges и может быть использован как сам по себе, так и в ансамбле с другими методами.

Предложенный подход имеет важное значение – он предоставляет вероятности для каждого ребра и легко может быть расширен для любых априорных распределений генных регуляторных сетей, что может быть использовано при адаптации метода к экспериментам другого вида. Так, будущая работа может быть направлена на получения таких априорных вероятностей из данных экспериментов по нокауту генов и по подавлению активности некоторых генов.

Исходный код метода доступен на веб-странице <https://github.com/alexloboda>

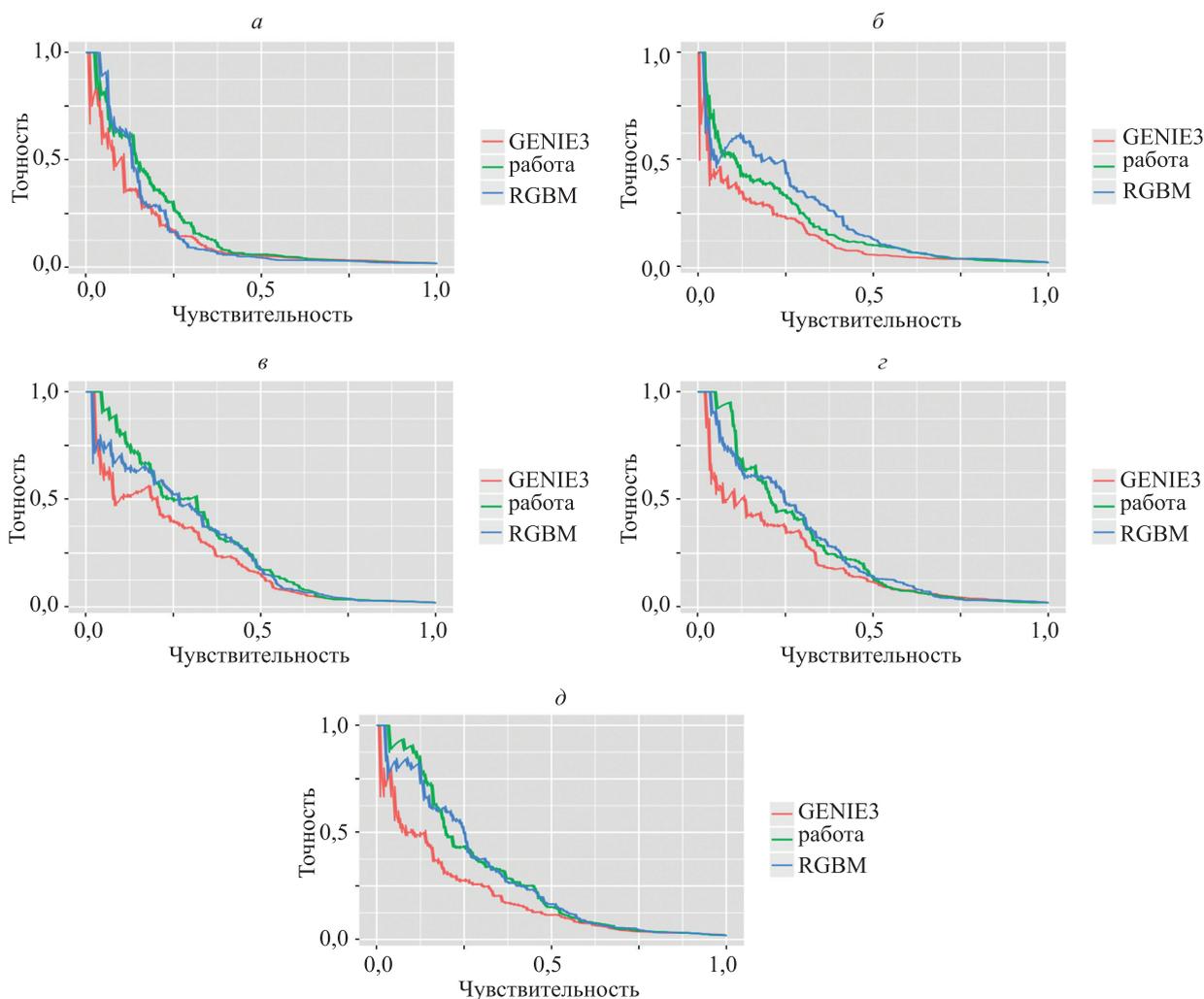


Рисунок. Графики кривых точности и чувствительности для методов GENIE3, RGBM и предложенного метода. Графики а–д соответствуют пяти симулированным экземплярам 1–5 задачи из набора insilico\_100\_multifactorial

### Литература

1. Csermely P., Ágoston V., Pongor S. The efficiency of multi-target drugs: the network approach might help drug design // *Trends in Pharmacological Sciences*. 2005. V. 26. N 4. P. 178–182. doi: 10.1016/j.tips.2005.02.007
2. Schaffter T., Marbach D., Floreano D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods // *Bioinformatics*. 2011. V. 27. N 16. P. 2263–2270. doi: 10.1093/bioinformatics/btr373
3. Marbach D., Costello J.C., Küffner R., Vega N.M., Prill R.J., Camacho D.M., Allison K.R., Kellis M., Collins J.J., Aderhold A., Stolovitzky G., Bonneau R., Chen Y., Cordero F., Crane M., Dondelinger F., Drton M., Esposito R., Foygel R., De La Fuente A., Gertheiss J., Geurts P., Greenfield A., Grzegorzczak M., Haury A.-C., Holmes B., Hothorn T., Husmeier D., Huynh-Thu V.A., Irrthum A., Karlebach G., Lèbre S., De Leo V., Madar A., Mani S., Mordelet F., Ostrer H., Ouyang Z., Pandya R., Petri T., Pinna A., Poultney C.S., Rezny S., Ruskin H.J., Saeys Y., Shamir R., Sirbu A., Song M., Soranzo N., Statnikov A., Vega N., Vera-Licona P., Vert J.-P., Visconti A., Wang H., Wehenkel L., Windhager L., Zhang Y., Zimmer R. Wisdom of crowds for robust gene network inference // *Nature Methods*. 2012. V. 9. N 8. P. 796–804. doi: 10.1038/nmeth.2016
4. Friedman N., Linial M., Nachman I., Pe'er D. Using Bayesian networks to analyze expression data // *Journal of Computational Biology*. 2000. V. 7. N 3-4. P. 601–620. doi: 10.1089/106652700750050961
5. Aghdam R., Ganjali M., Zhang X., Eslahchi C. CN: a consensus algorithm for inferring gene regulatory networks using the SORDER

### References

1. Csermely P., Ágoston V., Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. *Trends in Pharmacological Sciences*, 2005, vol. 26, no. 4, pp. 178–182. doi: 10.1016/j.tips.2005.02.007
2. Schaffter T., Marbach D., Floreano D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 2011, vol. 27, no. 16, pp. 2263–2270. doi: 10.1093/bioinformatics/btr373
3. Marbach D., Costello J.C., Küffner R., Vega N.M., Prill R.J., Camacho D.M., Allison K.R., Kellis M., Collins J.J., Aderhold A., Stolovitzky G., Bonneau R., Chen Y., Cordero F., Crane M., Dondelinger F., Drton M., Esposito R., Foygel R., De La Fuente A., Gertheiss J., Geurts P., Greenfield A., Grzegorzczak M., Haury A.-C., Holmes B., Hothorn T., Husmeier D., Huynh-Thu V.A., Irrthum A., Karlebach G., Lèbre S., De Leo V., Madar A., Mani S., Mordelet F., Ostrer H., Ouyang Z., Pandya R., Petri T., Pinna A., Poultney C.S., Rezny S., Ruskin H.J., Saeys Y., Shamir R., Sirbu A., Song M., Soranzo N., Statnikov A., Vega N., Vera-Licona P., Vert J.-P., Visconti A., Wang H., Wehenkel L., Windhager L., Zhang Y., Zimmer R. Wisdom of crowds for robust gene network inference. *Nature Methods*, 2012, vol. 9, no. 8, pp. 796–804. doi: 10.1038/nmeth.2016
4. Friedman N., Linial M., Nachman I., Pe'er D. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 2000, vol. 7, no. 3-4, pp. 601–620. doi: 10.1089/106652700750050961
5. Aghdam R., Ganjali M., Zhang X., Eslahchi C. CN: a consensus algorithm for inferring gene regulatory networks using the SORDER

- algorithm and conditional mutual information test // *Molecular BioSystems*. 2015. V. 11. N 3. P. 942–949. doi: 10.1039/c4mb00413b
6. Husmeier D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks // *Bioinformatics*. 2003. V. 19. N 17. P. 2271–2282. doi: 10.1093/bioinformatics/btg313
  7. Liu F., Zhang S.-W., Guo W.-F., Wei Z.-G., Chen L. Inference of gene regulatory network based on local bayesian networks // *PLoS Computational Biology*. 2016. V. 12. N 8. P. e1005024. doi: 10.1371/journal.pcbi.1005024
  8. Chen Y.-C., Wheeler T.A., Kochenderfer M.J. Learning discrete Bayesian networks from continuous data // *Journal of Artificial Intelligence Research*. 2017. V. 59. P. 103–132. doi: 10.1613/jair.5371
  9. Buntine W. Theory refinement on Bayesian networks // *Proc. 7<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*. 1991. P. 52–60. doi: 0.1016/B978-1-55860-203-8.50010-3
  10. Hastings W.K. Monte Carlo sampling methods using Markov chains and their applications // *Biometrika*. 1970. V. 57. N 1. P. 97–109. doi: 10.1093/biomet/57.1.97
  11. Chib S., Greenberg E. Understanding the metropolis-hastings algorithm // *American Statistician*. 1995. V. 49. N 4. P. 327–335. doi: 10.1080/00031305.1995.10476177
  12. Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ: [пер. с англ.]. М.: Издательский дом Вильямс, 2009. 1290 с.
  13. Albert R. Scale-free networks in cell biology // *Journal of Cell Science*. 2005. V. 118. N 21. P. 4947–4957. doi: 10.1242/jcs.02714
  14. Mall R., Cerulo L., Garofano L., Frattini V., Kunji K., Bensmail H., Sabedot T.S., Noushmehr H., Lasorella A., Iavarone A., Ceccarelli M. RGBM: regularized gradient boosting machines for identification of the transcriptional regulators of discrete glioma subtypes // *Nucleic Acids Research*. 2018. V. 46. N 7. P. e39. doi: 10.1093/nar/gky015
  15. Petralia F., Wang P., Yang J., Tu Z. Integrative random forest for gene regulatory network inference // *Bioinformatics*. 2015. V. 31. N 12. P. i197–i205. doi: 10.1093/bioinformatics/btv268
- algorithm and conditional mutual information test. *Molecular BioSystems*, 2015, vol. 11, no. 3, pp. 942–949. doi: 10.1039/c4mb00413b
6. Husmeier D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 2003, vol. 19, no. 17, pp. 2271–2282. doi: 10.1093/bioinformatics/btg313
  7. Liu F., Zhang S.-W., Guo W.-F., Wei Z.-G., Chen L. Inference of gene regulatory network based on local bayesian networks. *PLoS Computational Biology*, 2016, vol. 12, no. 8, pp. e1005024. doi: 10.1371/journal.pcbi.1005024
  8. Chen Y.-C., Wheeler T.A., Kochenderfer M.J. Learning discrete Bayesian networks from continuous data. *Journal of Artificial Intelligence Research*, 2017, vol. 59, pp. 103–132. doi: 10.1613/jair.5371
  9. Buntine W. Theory refinement on Bayesian networks. *Proc. 7<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, 1991, pp. 52–60. doi: 0.1016/B978-1-55860-203-8.50010-3
  10. Hastings W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, vol. 57, no. 1, pp. 97–109. doi: 10.1093/biomet/57.1.97
  11. Chib S., Greenberg E. Understanding the metropolis-hastings algorithm. *American Statistician*, 1995, vol. 49, no. 4, pp. 327–335. doi: 10.1080/00031305.1995.10476177
  12. Cormen T.H., Leiserson Ch.E., Rivest R.L., Stein C. *Introduction to Algorithms*. McGraw-Hill, 2003, 1056 p.
  13. Albert R. Scale-free networks in cell biology. *Journal of Cell Science*, 2005, vol. 118, no. 21, pp. 4947–4957. doi: 10.1242/jcs.02714
  14. Mall R., Cerulo L., Garofano L., Frattini V., Kunji K., Bensmail H., Sabedot T.S., Noushmehr H., Lasorella A., Iavarone A., Ceccarelli M. RGBM: regularized gradient boosting machines for identification of the transcriptional regulators of discrete glioma subtypes. *Nucleic Acids Research*, 2018, vol. 46, no. 7, pp. e39. doi: 10.1093/nar/gky015
  15. Petralia F., Wang P., Yang J., Tu Z. Integrative random forest for gene regulatory network inference. *Bioinformatics*, 2015, vol. 31, no. 12, pp. i197–i205. doi: 10.1093/bioinformatics/btv268

**Автор**

**Лобода Александр Александрович** — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 57190982085, ORCID: 0000-0002-0285-5841, aleks.loboda@gmail.com  
**Сергушичев Алексей Александрович** — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 55772694000, ORCID: 0000-0003-1159-7220, alserg@itmo.ru

**Author**

**Alexander A. Loboda** — Postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 57190982085, ORCID: 0000-0002-0285-5841, aleks.loboda@gmail.com  
**Alexey A. Sergushichev** — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 55772694000, ORCID: 0000-0003-1159-7220, alserg@itmo.ru