УДК 004
ББК 32.81

Роман Николаевич Гордеев,
кандидат физико-математических наук,
доцент кафедры информационных технологий,
Тверской государственный университет,
(170100, Россия, Тверь, ул. Желябова, 33)
e-mail: roman.gordeev@mail.ru

Применение метода анализа однородности для визуализации и анализа гетерогенных данных

Проблемы классификации и ранжирования весьма часто возникают в современном информационном обществе. Будь то потребность ранжировать пользователей информационного ресурса относительно их интересов и предпочтений, анализ потребительских предпочтений посетителей интернет-магазинов, анализ и сопоставление потребительских свойств однотипных товаров и многое другое.

Для решения большинства подобных проблем зачастую применяются эмпирические методы, в частности случайный лес, который доказал свою состоятельностьдля составления очень точных прогнозов при решении задач регрессии и классификации.

В работе рассматриваются проблемы ранжирования и классификации, а также предложена адаптация метода анализа однородности для эффективной визуализации комитетов решающих деревьев, в том числе и для визуализации новых наблюдений, не вошедших в обучающую выборку.

Ключевые слова: классификация, анализ однородности, визуализация графов.

Roman Nikolaevich Gordeev,
Candidate of Physical and Mathematical Sciences, Associate Professor,
Tver State University
(33 Zhelyabov St., Tver, Russia, 170100)
e-mail: roman.gordeev@mail.ru

Application of the Homogeneity Analysis for the Visualization and Analysis of Heterogeneous Data

Problems of classification and ranking arise in today's information society. Whether it's the need to rank users of information resources according to their interests and preferences, or analysis of consumer preferences of the visitors of internet shops, or may be an analysis and comparison of consumer properties of the some goods and more.

And an empirical methods suit well for such problems solving, in particular, a random forest, which has proved its worth for making very accurate predictions for solving regression and classification.

In this paper we consider the problem of ranking and classification. The adaptation of the analysis of homogeneity method had been proposed for effective visualization committees of decision trees, including visualization and new observations that were not included in the training set.

Keywords: classification, homogeneity analysis, visualization of graphs.

1. Введение. В работе мы рассмотрим проблему классификации данных и их визуального представления, а также предложим эффективный алгоритм, позволяющий значительно сократить количество вычислений, требуемых при малых изменениях анализируемых данных, возникающих в связи с добавлением или удалением какой-либо информации.

Для классификации по K классам положим, что $(X_1, Y_1), ..., (X_n, Y_n)$ — n наблюдений прогнозирующей переменной $X \in \Xi$, $Y \in \{1, ..., K\}$ – переменная ответа класса. Прогнозирующая переменная X является вектором размерности p и может содержать непрерывные или факторные переменные.

Визуализация малоразмерных вложений данных может принимать различные формы, от неконтролируемого обучения до контролируемого уменьшения размерности или малоразмерных вложений данных, таких как анализ соседних компонентов [1] и схожих методов [2]. Однако у нас несколько другие цели: мы хотим рассмотреть существующий алгоритм, точнее, класс алгоритмов, включающих случайный лес и вложенные деревья классификации, и разработать эффективные методы визуализации для членов этого класса.

Ансамбли деревьев показали, что могут делать достаточно точные прогнозы, а случайный лес, возможно, один из лучших самообучающихся машинных алгоритмов в том смысле, что его точность предсказаний очень близка к истинному значению даже без особой настройки параметров.

В нашей работе мы будем рассматривать случайный лес [3], вложенные деревья решений [4], и некоторые новые методы, например, ансамбли правил [5].

Некоторые из существующих работ по визуализации деревьев [8] и ансамблей деревьев были всесторонне рассмотрены в работе [12] и включают рассмотрение рельефных графиков и графиков слежения, используемых, например, для определения устойчивости деревьев и определения, какие переменные были выбраны в качестве критерия классификации.

Так называемые матрицы близости часто используются для маломерных вложений наблюдений в классификации [3; 9; 10]. Вопросы неконтролируемого обучения с применением случайного леса рассмотрены в работе [11]. Каждая запись в матрице близости отражает долю деревьев в ансамбле, для которого пара наблюдений находится в одном и том же узле. Недостатком указанных методов визуализации является то, что исходные узлы дерева не отображаются, поскольку происходит потеря информации за счёт агрегации в матрице близости. Кроме того, это затрудняет добавление новых наблюдений.

В нашей работе мы будем применять методы анализа однородности для визуализации ансамблей деревьев. В данном подходе наблюдения и правила (узлы) формируют двудольный граф. Минимизация квадратов длин рёбер в этом графе приводит к очень интересным малоразмерным проекциям данных.

В размещении правил (узлов) и наблюдений на одном графике есть некоторые преимущества, это позволяет лучше интерпретировать проекции. Кроме того, это позволяет точнее определять границы классов и отражать точность прогнозирования используемого ансамбля деревьев. Если количество классов классификации мало, то можно показать, что обычное правило классификации «ближайшего соседа» для малоразмерных вложений данных позволяет производить прогнозирование с точностью, аналогичной точности прогнозирования при использовании случайного леса.

В данном исследовании мы будем применять метод анализа однородности для визуализации больших массивов наблюдений и правил (узлов) на одном графике.

2. Анализ однородности.

2.1. Матрица индикаторов. Анализ однородности был разработан в прикладных науках для исследования и прогнозирования социальных процессов и визуального представления данных с факторными переменными. Положим, что у нас есть f факторных переменных $h = 1, ..., f$, каждая с уровнем фактора l_h. Данные могут быть представлены в виде бинарной матрицы, если закодировать каждую из $h = 1, ..., f$ переменных виде $n \times l_n$ матрицы бинарного индикаторы $G^{(h)}$, в которой k-я колонка содержит 1 для всех наблюдений, имеющих уровень фактора k для переменной h. Эти матрицы могут быть объединены в матрицу $n \times m$ $G = (G^{(1)}, ..., G^{(f)})$, ult $m = \sum_h l_h$ – общее количество фиктивных переменных.

Каждый из листьев дерева может быть представлен бинарной переменной индикатором, в которой 1 означает, что наблюдение попадает в лист дерева, 0 не попадает. Рассматривая листы как обобщённые фиктивные переменные, можно аналогично построить матрицу индикаторов G для ансамблей деревьев. Для данного ансамбля деревьев с общим количеством злов m, пусть $P_j \subset \Xi, j = 1, ..., m$, будут гиперплоскостями пространства Ξ соответствующие листу j. Наблюдение попадает в лист P_j тогда и только тогда, когда $X_i \in P_j$. Результаты ансамбля деревьев с m листами во всех деревьях можно объединить в матрицу индикатор G размерности $n \times m$, где $G_{ij} = 1$, если i-е наблюдение попадает в j-й лист, и 0, иначе:

$$G_{ij} = \begin{cases} 1, X_i \in P_j, \\ 0, X_i \notin P_j. \end{cases}$$

Эта матрица очень похожа на матрицу индикаторов анализа однородности. Построковая сумма матрицы G идентична числу F факторных переменных анализа однородности, более того, построко-

вая сумма G для ансамблей деревьев равна числу деревьев, поскольку каждое наблюдение попадает только один раз в лист в каждом дереве. Далее мы не будем считать построковую сумму постоянной, это позволит нам сделать некоторые обобщения без чрезмерного усложнения обозначений и вычислений. Единственным предположением, которое мы введём, заключается в том,что суммы по строкам и колонкам матрицы G строго положительны, т. е. каждое наблюдение попадает по крайней мере в один узел, а каждый узел содержит по крайней мере одно наблюдение. Кроме того, мы будем полагать, что корневой узел, содержащий все наблюдения, входит в набор правил. Это гарантирует, что двудольный граф, соответствующий G, является связанным.

2.2. Двудольный граф и анализ однородности. Анализ однородности можно рассматривать как формирование двудольного графа, в котором каждое из n наблюдений и каждое из m правил или фиктивных переменных представлены узлом графа. Между наблюдением и правилом (узлом) существует ребро тогда и только тогда, когда наблюдение удовлетворяет правилу. Другими словами, между наблюдением i и правилом j существует ребро тогда и только тогда, когда $G_{ij} = 1$.

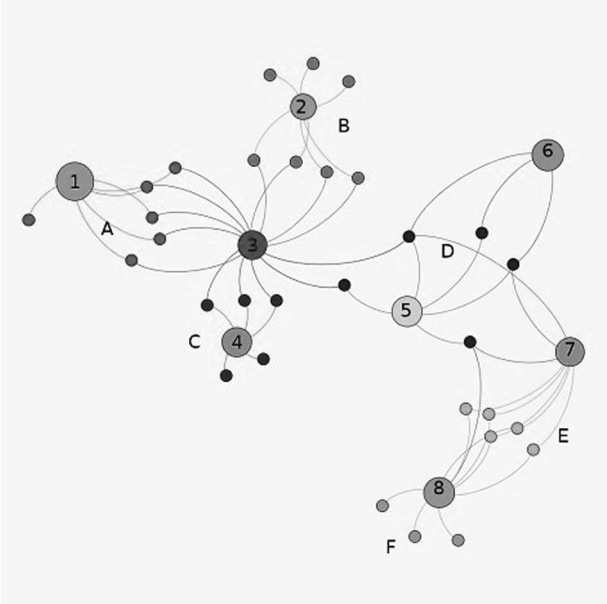

Рис. 1. Двудольный граф, сформированный из авторов, принадлежащих различным областям знаний, и их статей

Таким образом, целью является расположить наблюдение как можно ближе ко всем правилам, которые его содержат. И, наоборот, правило должно находиться как можно ближе к наблюдению, которое оно содержит.

На рис. 1 представлен пример отображения выборки авторов, принадлежащих различным областям знаний и их статей. В данном случае статьи (малые узлы графы) являются наблюдениями, а авторы (большие узлы графа) являются правилами. Некоторые авторы могут иметь совместные статьи, т. е. при классификации статей могут формироваться группы, которые отражают не только принадлежность той или иной статьи к определенному роду знаний, но и показывают связи между различными видами знаний. Так, например, фиолетовым (отмечен цифрой 3) и жёлтым (маркер 1) отмечены соответственно автор математик и автор физик, т. о. видно, что виды знаний весьма тесно связаны, а совместные статьи этих авторов образуют отдельный кластер, который содержит ошибочно в т. ч. и отдельную статью по физике. Анализ однородности пытается минимизировать сумму квадратов длин всех рёбер. На картинке представлен граф при фиксированном положении правил. Каждая статья располагается в центре относительно всех правил, которые к ней применяются. Цвет статей соответствует различным классам. Физико-математические статьи, например, рисуются красным (литера A), статьи на стыке информатики и математики зеленым (литера C).

Пусть U – матрица размером $n \times q$, содержащая координаты n наблюдений во вложении размерности q, R – матрица $m \times q$ проецируемых правил. Обозначим через U_i i-ую строку матрицы U, а

через R_j j-ю строку матрицы R. Анализ однородности выбирает проекцию, минимизируя квадраты длин рёбер:

$$\arg\min_{U,R} \sum_{i,j:G_{ij}=1} \|U_i - R_j\|_2^2 . \tag{1}$$

Пусть e_n – n-мерная колонка колонка-вектор, содержащая одни 1, а 1_q – q-мерная единичная матрица. Чтобы избежать тривиальных решений дополнительно накладываются ограничения [6] вида

$$U^T W U = 1_q, \tag{2}$$

$$e_n^T U = 0, \tag{3}$$

где W – положительно определённая весовая матрица.

Таким образом, анализ однородности соответствует нахождению вложения размерности q (где q обычно равно 2) как наблюдений, так и правил, таких, что сумма квадратов длин рёбер является минимальной. Далее мы будем взвешивать выборки количеством правил, с которыми они связаны, таким образом W будет диагональной матрицей с элементами $W_{ii} = \sum_j G_{ij}$, поэтому $W = diag(GG^T)$ является диагональной частью GG^T. В стандартном анализе однородности каждая выборка является частью точно такого же количества правил, поскольку соответствуют уровням фактора, и весовая матрица $W = f1_n$ таким образом будет единичной матрицей, умноженной на число f факторной переменной. Для большинства ансамблей деревьев также справедливо, что $W = T1_n$ является диагональной матрицей, поскольку каждое наблюдение попадает в такое же количество T листьев, где T – количество деревьев в ансамбле.

3. Основные результаты.

3.1. Отображение правил. Если координаты U наблюдений остаются постоянными, то расположения правил R в задаче (1) при ограничениях (2) могут быть легко найдены, поскольку ограничения не зависят от R. Каждая проекция правила R_j, $j = 1, ..., m$ находится в центре относительно всех наблюдений, которые, как правило, содержат:

$$R_j = \frac{\sum_i G_{ij} U_i}{\sum_i G_{ij}}$$

$$R = diag\left(G^T G\right)^{-1} G^T U, \tag{4}$$

где $diag(M)$ – диагональная часть матрицы M, в которой все элементы, не принадлежащие диагонали, равны 0.

Значения матрицы U могут быть найдены методом наименьших квадратов или посредством решения задачи с собственными значениями.

Задача оптимизации (1) при ограничениях (2) может быть решена либо оптимизацией позиций U для n реализаций, либо оптимизацией позиций R для m правил, при этом в каждом случае остальные параметры остаются неизменными. Оптимизация относительно позиций правил R при заданных фиксированных позициях U реализаций решена в (4). Оптимизация относительно U при ограничениях (2) решается методом наименьших квадратов аналогично (4) [7]:

$$U = diag\left(GG^T\right)^{-1} GR, \tag{5}$$

помещая каждую реализацию в центр всех правил, которые её содержат.

В отличие от следующего подхода, основанного на вычислении собственных значений, вычисления состоят только из умножения матриц. А поскольку G в большинстве случаев является разряженной матицей, то это может значительно улучшить эффективность вычислений.

Целевая функция задачи (1) может быть альтернативно записана в виде:

$$\sum_{i,j:G_{ij}=1} \|U_i - R_j\|_2^2 =$$

$$= \sum_i \|U_i\|_2^2 \sum_j G_{ij} + \sum_j \|R_j\|_2^2 \sum_i G_{ij} - 2 \sum_{ij} tr(U_i^T G_{ij} R_j) = \tag{6}$$

$$= tr(U^T diag(GG^T)U) + tr(R^T diag(G^T G)R) - 2tr(U^T GR),$$

где $tr(M)$ – след матрицы M, обозначим:

$$D_u = diag(GG^T),$$

$$D_r = diag(G^T G).$$

Тогда исходная задача может быть сведена к задаче вида:

$$\arg\max_U tr(U^T(GD_r^{-1}G^T)U) \tag{7}$$

при ограничениях

$$U^T D_u U = 1_q$$
$$e_n^T U = 0.$$

Далее, используя разложение по собственным векторам и (4), получим решение:

$$U = D_u^{-1/2} V, \tag{8}$$

$$R = D_r^{-1} G^T U, \tag{9}$$

где $V = (\vartheta_1, ..., \vartheta_q)$ – матрица, состоящая из первых q собственных векторов матрицы:

$$A := D^{-1/2} S_n^T \left(GD_r^{-1}G^T\right) S_n D_u^{-1/2}.$$

3.2. Фиксированные позиции правил и новые наблюдения. После получения раскладки графа (9) мы предлагаем упростить второй шаг и зафиксировать положение правил в точках полученных решений. В этом случае анализ однородности будет иметь очень полезную черту: можно будет элементарно добавлять новые наблюдения на график без пересчёта всего решения. Минимизировать сумму квадратов длин рёбер (1) при фиксированных позициях правил весьма просто, наблюдение помещаетсяв центр всех правил, которые к нему применяются. В матричном виде решение выглядит так:

$$U = diag(GG^T)^{-1}GR = D_u^{-1}GR. \tag{10}$$

Простой пример приведён на рис.1. На нём представлены статьи (наблюдения) и авторы (правила), полученные на основе анализа структуры данных системы цитирования elibrary.ru. Правила являются бинарными индикаторами, отражающими одну или несколько характеристических черт (в нашем исследовании это авторство статьи).

Для размещения наблюдений мы предлагаем следующий двухшаговый подход.

1. Находим положение правил, минимизируя сумму квадратов длин ребер, при ограничениях (2) и (3).

2. Фиксируем положение правил и размещаем все наблюдения U, минимизируя снова сумму квадратов длин рёбер. Каждое наблюдение размещается в центре всех правил, которые к нему применяются, используя (10).

3. Добавляем новое наблюдение и вычисляем его расположение, используя (10).

В примере, приведённом на рис. 2, показано предсказание новых наблюдений в двухмерных проекциях. Раскрашенные наблюдения соответствуют обучающим наблюдениям с известными кодами цвета (класса области знаний). Светлофиолетовым изображена статья по физике (обозначена литерой A' на рисунке), авторами которой являются физик, математик и информатик. Статья при этом правильно классифицирована и отнесена к области знаний «физика», поскольку это ближайшая обучающая группа. Справа аналогичным образом классифицируется изображённая серым узлом (обозначена на рисунке литерой E') статья по медицине, написанная медиком, специалистом в области ИТ и биологом. Однако она была неверно отнесена к области знаний «химия» (розовый цвет, группа E).

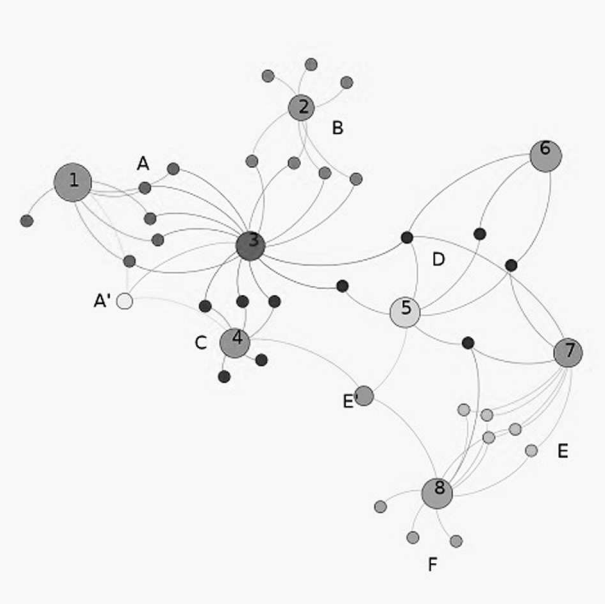

Рис. 2. Классификация вновь добавленных статей

4. Пример.

Теперь попробуем применить анализ однородности для оценки действий экспертов, рецензирующих статьи в научных журналах. И кроме того, попробуем проанализировать некоторые другие интересные характеристики, полученные в ходе эксперимента. Данные о результатах экспертизы статей, принимаемых в рецензируемые журналы по астрофизике, были собраны Нури и Роландом за промежуток с 2002 по 2006 гг. Всего было получено информации о 5 745 оценках статей, оцениваемых 696 экспертами. При каждой оценке возможен один из 3 вариантов: оценивать положительно, оценить отрицательно, отправить на доработку.

При анализе этих данных нас будет интересовать вопрос о том, можно ли установить связь между оценкой статьи и принадлежностью эксперта определённому научному центру. Это позволит нам сделать ряд определённых выводов, касающихся предвзятой оценки своих коллег, насколько одинаково оценивают статьи эксперты из одного научного центра, существуют ли внутри экспертов определённые подгруппы. Так же нас будет интересовать вопрос, есть ли связь между национальностью эксперта и его оценками статей.

Все эксперты принадлежали одному из 8 научных центров, первый будем обозначать светло-голубым, второй – тёмно-синим, третий – жёлтым, четвёртый – зелёным, пятый – красным, шестой – оранжевым, седьмой – коричневым, восьмой – чёрным.

Классификатор случайный лес со 100 деревьями и параметрами, настроенными на обучающей выборке. Принадлежность эксперта тому или иному научному центру определяется при помощи случайного леса с ошибкой примерно в 10 %. Двухмерные вложения нам интересны для выявления связей между научными центрами и выявления экспертов, ведущих себя значительно отлично от своих коллег.

Многомерное шкалирование. Для визуализации результатов анализа, полученных при помощи случайного леса, обычно применяют многомерное шкалирование близостей [3; 9]. Близость двух наблюдений определяется как соотношение деревьев в случайном лесе, для которого оба наблюдения попадают в листовой узел. Матрица близостей в наших обозначениях может быть вычислена при помощи следующего выражения:

$$T^{-1}GG^T,$$

где T – количество деревьев в ансамбле. Расстояние, определяемое как 1, меньше чем близость. Тогда матрица расстояний вычисляется при помощи выражения:

$$1_n - T^{-1}GG^T,$$

или при помощи монотонного преобразования, такого как квадратный корень, используемого в [11]. Здесь мы используем неметрическое многомерное шкалирование, а именно isoMDS [13–15], которое делает результаты инвариантными относительно любого монотонного преобразования расстояний. MDS генерирует двухмерные вложения для всех наблюдений, в нашем случае экспертов, которые можно видеть слева на рис. 3.

Рис. 3. Применение анализа однородностей для выявления групп экспертов

Хотя это даёт представление о близости между институтами, всё же научные центры не располагаются как когерентные блоки и особенно это заметно на экспертах из 7 и 8 научных центров (черный и коричневый), которые весьма широко разбросаны.

Анализ однородности. Справа на рисунке 3 представлены результаты обработки тех же данных при помощи анализа однородности. При этом G является матрицей размерности $n \times m$, в которой каждый из m столбцов соответствует листовому узлу случайного леса. Значение 0 соответствует случаю, когда эксперт не попал в лист, 1 — попал. Теперь заметно, что все научные центры имеют свои точки притяжения и располагаются более компактно, формируя треугольник, как показано на рисунке 3 справа. На рисунке отчётливо видно, что некоторые из научных центров имеют достаточно сильные взаимосвязи. При использовании этого подхода процент неверной классификации составляет примерно 19 % и есть ещё над чем работать (для случая использования метода «ближайший сосед» в двухмерных вложениях).

Однако, даже эти результаты дают весьма интересную картину. Так можно сказать, что эксперты из институтов 2, 3, 7, 8 примерно одинаково оценивают все статьи, в то время как эксперты из 4 и 6 институтов некоторые статьи оценивают одинаково, а относительно других статей их мнения весьма разнятся.

Выводы.

В работе было предложено использование метода однородности для визуализации и классификации данных. Подход заключается в вычислении начальных расположений правил на основе обучающей выборки и формировании начальных групп наблюдений на основе метода наименьших квадратов. Далее мы фиксируем расположение правил и поступающие новые наблюдения располагаем относительно уже зафиксированных расположений правил, а классификация нового наблюдения производится на основе принципа «ближайшего соседа» по отношению к классам наблюдений, сформированным обучающими данными.

Данный подход позволяет избежать полного пересчёта всей модели при добавлении в модель новых наблюдений и увеличить скорость работы алгоритма.

Возможности подхода продемонстрированы на модельном примере, оценивающем связи институтов на основе действий экспертов, рецензирующих статьи.

Список литературы

1. Goldberger J. Neighbourhood Components Analysis / Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R. Advances in Neural Information Processing System, 2005. Vol. 17. P. 513—520.

2. Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis // The Journal of Machine Learning Research. 2007. № 8. P. 1061.

3. Breiman L. Random Forests // Machine Learning. 1997. № 45. P. 5—32.

4. Breiman L. Bagging Predictors // Machine Learning. 1996. № 24. P. 123—140.

5. Friedman J. Predictive Learning via Rule Ensembles /J. Friedman and B. Popescu // The Annals of Applied Statistics. 2008. № 2. P. 916—954.

6. De Leeuw J. Homogeneity Analysis in R: The Package Homals / De Leeuw J. and Mair P. // Journal of Statistical Software. 2008. № 31. P. 1—21.

7. Michailidis G. The Gifi System of Descriptive Multivariate Analysis / Michailidis G. and De Leeuw J. // Statistical Science. 1998. Vol. 13. № 4. P. 307—336.

8. Breiman L. Classification and Regression Trees / L. Breiman, J. Friedman, R. Olshen and C. Stone 1998. Belmont: Wadsworth

9. Liaw A. Classication and Regression by Random Forest / A. Liaw and M. Wiener // R News. 2002. № 2. P. 18—22.

10. Lin Y. Random Forests and Adaptive Nearest Neighbors / Y. Lin and Y. Jeon // Journal of the American Statistical Association. 2006. № 101. P. 578—590.

11. Shi T. Unsupervised Learning With Random Forest Predictors / Shi T. and Horvath S. // Journal of Computational and Graphical Statistics. 2006. № 15. P. 118—138.

12. Urbanek S. Visualizing Trees and Forests // in Handbook of Data Visualization. 2008. Berlin, Heidelberg: Spinger. P. 243—264.

13. Borg I. and Groenen P. Modern Multidimensional Scaling: Theory and Applications 1997. New York.

14. Kruskal J. Nonmetric Multidimensional Scaling: A Numerical Method // Psychometrika. 1964. № 29. P. 115—129.

15. Kruskal J., Wish M. Multidimensional Scaling. 1978. Beverly Hills: Sage Publications.

References

1. Goldberger J. Neighbourhood Components Analysis / Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R. Advances in Neural Information Processing System, 2005. Vol. 17. P. 513—520.

2. Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis // The Journal of Machine Learning Research. 2007. № 8. P. 1061.

3. Breiman L. Random Forests // Machine Learning. 1997. № 45. P. 5—32.

4. Breiman L. Bagging Predictors // Machine Learning. 1996. № 24. P. 123—140.

5. Friedman J. Predictive Learning via Rule Ensembles / J. Friedman and B. Popescu // The Annals of Applied Statistics. 2008. № 2. P. 916—954.

6 De Leeuw J. Homogeneity Analysis in R: The Package Homals / J. De Leeuw and P. Mair // Journal of Statistical Software. 2008. № 31. P. 1—21.

7. Michailidis G. The Gifi System of Descriptive Multivariate Analysis / Michailidis G. and De Leeuw J. // Statistical Science. 1998. Vol. 13. № 4. P. 307—336.

8. Breiman L. Classification and Regression Trees / L. Breiman, J. Friedman, R. Olshen and C. Stone. 1998. Belmont: Wadsworth

9. Liaw A. Classication and Regression by Random Forest /A. Liaw and M. Wiener// R News. 2002. № 2. P. 18—22.

10. Lin Y. Random Forests and Adaptive Nearest Neighbors /Y. Lin and Y. Jeon // Journal of the American Statistical Association. 2006. № 101. P. 578—590.

11. Shi T. Unsupervised Learning With Random Forest Predictors /T. Shi and S. Horvath // Journal of Computational and Graphical Statistics. 2006. № 15. P. 118—138.

12. Urbanek S. Visualizing Trees and Forests // in Handbook of Data Visualization. 2008. Berlin, Heidelberg: Spinger. P. 243—264.

13. Borg I. and Groenen P. Modern Multidimensional Scaling: Theory and Applications 1997. New York.

14. Kruskal J. Nonmetric Multidimensional Scaling: A Numerical Method // Psychometrika. 1964. №29. P. 115–129.

15. Kruskal J., Wish M. Multidimensional Scaling. 1978. Beverly Hills: Sage Publications.

Статья поступила в редакцию 10.04.2014