

## Возможности метода деревьев классификации при обработке социологической информации

© Е.Е. Фомина

Тверской государственный технический университет, Тверь, 170026, Россия

*Для решения задач анализа структуры данных и получения информации о взаимосвязи переменных, описывающих объект или явление, на практике применяют методы многомерного разведывательного анализа данных. К таким методам относятся кластерный, дискриминантный, факторный, логлинейный анализ и ряд других. Одним из наиболее гибких является метод деревьев классификаций, позволяющий решать задачи классификации и прогнозирования. Он может быть использован на практике в различных областях, в том числе в социологии. Рассмотрен алгоритм метода деревьев классификации и его возможности для анализа социологической информации, в частности, результатов анкетирования на тему политической активности молодежи.*

**Ключевые слова:** *деревья классификации, обработка социологической информации, анкетирование*

**Введение.** Математические методы, в частности методы многомерного разведывательного анализа данных, стали широко использоваться во многих областях научного знания. Их популярность объясняется тем, что они позволяют решать различные содержательные задачи и получать предварительную информацию о структуре взаимосвязей между переменными в больших массивах данных.

К таким методам относятся кластерный анализ, предназначенный для разбиения множества объектов на однородные группы [1–4]; дискриминантный анализ, основная задача которого — изучение различий между двумя и более группами объектов по нескольким переменным одновременно [1, 4, 5]; факторный анализ, направленный на выявление скрытых факторов, отвечающих за наличие линейных статистических связей между наблюдаемыми переменными [1, 4, 6–12]; логлинейный анализ, используемый для комплексной оценки взаимосвязей в многомерных таблицах частот [13–19], и другие методы [20, 21].

Наиболее гибким методом многомерного разведывательного анализа, позволяющим прогнозировать принадлежность объектов к определенному классу в зависимости от значения признаков, характеризующих объект, является метод деревьев классификации (ДК).

Метод ДК может быть эффективно использован на практике в различных областях, в том числе в социологии при обработке социологической информации. В частности, он может применяться при анализе информации, полученной в результате опроса или анкетирования, когда необходимо охарактеризовать определенные группы респондентов.

Цель настоящей статьи — демонстрация возможностей метода ДК при обработке социологической информации.

Возможности метода будут рассмотрены при анализе результатов анкетирования на тему «Политическая активность молодежи».

**Методы исследования.** Деревья классификации — метод многомерного разведывательного анализа, позволяющий решать задачи классификации информации в случае, когда зависимая переменная принимает дискретные значения, и задачи прогнозирования принадлежности объекта выборки к определенному классу в случае, когда зависимая переменная принимает непрерывные значения.

В качестве исходных данных выступает таблица, в которой представлены значения признаков (переменных), характеризующих объект. Одна из переменных — зависимая — содержит значения классов, к которым относятся объекты, остальные переменные (независимые) — значения показателей, позволяющих прогнозировать принадлежность объектов к тому или иному классу.

В результате применения метода строится дерево, имеющее структуру графа. Вершина нулевого уровня, или корневая вершина, включающая в себя все объекты выборки, будет делиться на вершины-потомки согласно решающему правилу «если значение переменной больше или равно / меньше или равно ..., то объект относится к ... классу». Вершины-потомки, в свою очередь, согласно соответствующим решающим правилам будут делиться до тех пор, пока не окажутся терминальными, т. е. пока их дальнейшее деление не прекратится (рис. 1). Каждая вершина (в том числе терминальная) будет соответствовать тому или иному классу объектов. Решающие правила позволят выявить некую закономерность, скрытую в массиве данных,

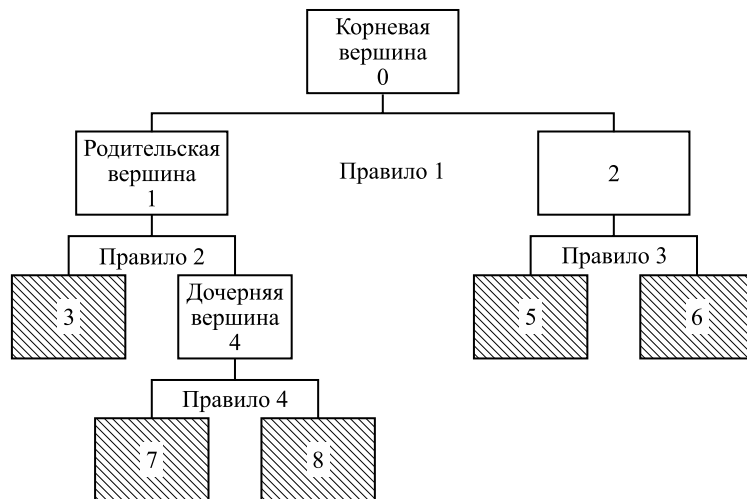


Рис. 1. Структура дерева классификации:

■ — терминальная вершина

провести классификацию нового объекта, т. е. определить, в какую вершину дерева он попадет, или охарактеризовать классы.

Основные преимущества данного метода:

- графическое представление получаемых результатов, которое способствует упрощению интерпретации выводов по сравнению с интерпретацией данных числовой природы и, как следствие, упрощению процедуры принятия решений;
- возможность выбора из множества переменных наиболее значимых, которые используются для построения дерева и формулирования решающих правил;
- устойчивость к выбросам и возможность автоматической обработки пропущенных значений;
- возможность обработки переменных, представленных в любых шкалах;
- отсутствие какого-либо теоретического распределения, так как алгоритм является непараметрическим.

Среди основных недостатков метода можно отметить отсутствие прогнозного уравнения, описывающего модель и возможность построения слишком детализированного дерева, которое вызовет сложность при интерпретации результатов.

Рассмотрим этапы построения ДК.

1. Выбор критерия точности прогноза.

Основная цель построения ДК — возможность осуществлять достаточное точное прогнозирование принадлежности объекта к тому или иному классу значений. Под точностью прогноза в данном случае понимается процент неверно классифицированных объектов, или цена ошибки классификации. Чем меньше цена ошибки классификации, тем выше точность прогноза, следовательно, построенное дерево более качественное.

На точность прогноза оказывает влияние ряд факторов.

Первым ключевым моментом является корректная оценка априорных вероятностей, т. е. вероятностей попадания объекта в тот или иной класс. При решении данной проблемы возможны следующие варианты:

- априорные вероятности полагаются равными. Такая ситуация возникает, если исследователю заранее известно, что каждый класс содержит приблизительно равное число объектов;
- априорные вероятности берутся пропорциональными размерам классов. Данная ситуация возможна, если исследователь располагает какой-либо предварительной информацией о структуре выборки и частотах встречаемости.

Вторым ключевым моментом является тот факт, что в ряде исследований необходимо получить более точный прогноз для определенных классов объектов. В этом случае одним классам будет соответствовать меньшая, а другим — большая цена ошибок классификации.

## 2. Выбор типа ветвления.

Выбор подходящего для решаемой задачи типа ветвления — один из самых важных этапов, на котором будет построено ДК. Сегодня разработан достаточно большой набор методов построения деревьев классификации. Рассмотрим те из них, которые реализованы в таких программных продуктах, как SPSS, STATISTICA, MATLAB.

*Алгоритм CHAID.* Используется для построения ДК в случае любого типа независимых переменных. Алгоритм основан на применении критерия  $\chi^2$ , который позволяет оценить статистическую значимость различий двух (или нескольких) переменных.

При расщеплении узла (начиная с корневой вершины) ищут ту независимую переменную, которая наилучшим образом объясняет различия между значениями зависимой переменной, используя критерий  $\chi^2$ .

*Алгоритм QUEST.* Применяется для построения двоичного дерева классификации в случае любого типа независимых переменных. В процессе построения дерева на каждом этапе выбирается та терминальная вершина, которая будет расщеплена, и наиболее значимая переменная, которая будет использоваться для составления решающего правила в вершине. При решении этой задачи вычисляются  $p$ -уровни для проверки значимости взаимосвязей между принадлежностью объектов к классам и уровнями каждой из независимых переменных [22–25]. Для проверки гипотезы независимости принадлежности к классам от уровня категориальной переменной в данном узле дерева рассчитывается критерий  $\chi^2$  (если переменные категориальные) или проводится дисперсионный анализ (если переменные порядковые). Далее для выделения двух классов применяется кластеризация методом  $k$ -средних [22–24].

*Алгоритм CART.* Применяется для построения двоичного дерева классификации по категориальным и порядковым независимым переменным. Метод основан на идее уменьшения неопределенности в узле. На каждом шаге построения ДК решающее правило делит выборку на две части: в одной из них правило выполняется, а в другой — нет. При этом перебираются все возможные варианты ветвления для каждого узла и выбирается та независимая переменная, при которой оценочная функция дает наилучший показатель. В качестве оценочной функции или критерия согласия используется мера Джини однородности вершины [22–24].

Выбор метода является прерогативой исследователя, но надо отметить, что алгоритм CHAID следует использовать в том случае, когда необходимо найти переменные, которые максимизируют связь с откликом. Другие алгоритмы рекомендуется применять, если необходимо выделить гомогенные группы, которые хорошо будут предсказываться в дальнейшем.

## 3. Определение момента прекращения ветвления.

Построенное дерево должно не только давать точные результаты прогнозирования, но и быть наглядным и легко анализируемым, пре-

доставлять исследователю возможность построения некоего алгоритма для классификации объектов и анализа групп. Построение дерева и решающих правил должно опираться на те признаки, которые являются наиболее информативными и увеличивают точность прогноза (незначимые факторы можно проигнорировать). Оно должно способствовать более глубокому пониманию анализируемого процесса.

Это требование напрямую связано с определением момента прекращения дальнейшего ветвления, так как слишком детализированное дерево не позволит выделить и охарактеризовать классы объектов, выявить скрытую закономерность в структуре данных.

Критериями останова могут быть [23, 24]:

- число неклассифицированных объектов. Согласно данному правилу, разделение узлов продолжается, пока количество объектов в каждой терминальной вершине превышает заданное число;
- доля неклассифицированных объектов. Согласно данному правилу, разделение узлов продолжается, пока количество объектов в каждой терминальной вершине превышает заданную долю;
- прямая остановка по методу FАСТ. Согласно данному правилу, разделение узлов продолжается до тех пор, пока каждая терминальная вершина не окажется «чистой» или количество объектов в ней из одного или нескольких классов не станет меньше заданной доли.

4. Кросс-проверка построенного дерева на независимой выборке. При использовании метода для решения конкретной практической задачи ДК строится по ограниченному набору объектов, называемому обучающей выборкой. Однако ценность дерева заключается в том, что оно должно давать хорошие прогнозы и на независимой (или тестовой) выборке. Для оценки качества построенного ДК применяют следующие методы [22–24]:

- построение ДК на обучающей выборке и дальнейшее его исследование или кросс-проверка тестовой выборки. Этот метод можно реализовать, если размер исходной выборки является достаточным для ее разбиения случайным образом на обучающую и тестовую. Если в результате применения построенного ДК к тестовой выборке окажется, что цена ошибки классификации больше, чем на обучающей выборке, необходимо изменить параметры примененного метода ветвления или попробовать использовать другой метод;
- $V$ -кратная кросс-проверка. Этот вариант используется при большом объеме исходной совокупности и отсутствии тестовой выборки. Исследователь задает число  $V$ , определяя при этом число подвыборок, на которое разбивается исходное множество объектов. Дерево строится  $V$  раз, при этом одна из подвыборок выступает в качестве обучающей для проведения кросс-проверки. Для каждого дерева вычисляется цена ошибки и на основании полученных данных рассчитывается ее среднее значение;

- глобальная кросс-проверка. Данный вариант предполагает проведение  $n$  итераций, на каждой из них происходит отсечение небольшой доли объектов от исходной выборки, которая в дальнейшем выступает в качестве тестовой выборки;
- кросс-проверка по минимальной цене-сложности. В данном случае проводится расчет функции потерь, равной доли неправильно классифицированных объектов при оцениваемых априорных вероятностях и одинаковых ценах ошибок классификации.

**Материалы исследования.** Рассмотрим возможности метода на примере обработки анкет на тему «Политическая активность молодежи». Были опрошены студенты 1–4-го курса. Используя метод ДК, проанализируем состав групп респондентов с активной и пассивной политической позицией.

Анкета содержала следующие вопросы.

1. Интересуетесь ли Вы политикой и принимаете ли участие в выборах в различные властные структуры?
  - 1) да;
  - 2) нет.
2. Как часто?
  - 1) постоянно;
  - 2) от случая к случаю;
  - 3) редко.
3. Принимаете ли Вы участие в общественной жизни?
  - 1) да;
  - 2) нет.
4. Каким образом?
  - 1) состою в политической партии;
  - 2) участвую в митингах, акциях протеста;
  - 3) регулярно участвую в выборах;
  - 4) участвую в органах студенческого самоуправления;
  - 5) участвую в общественных организациях;
  - 6) иное.
5. Укажите Ваш пол:
  - 1) мужской;
  - 2) женский.
6. Укажите Вашу специальность.
7. Укажите Ваш курс.
8. Ваше материальное положение.
  - 1) мы едва сводим концы с концами, денег не хватает даже на питание;
  - 2) на продукты денег хватает, но покупка одежды вызывает затруднения;
  - 3) денег хватает на продукты и одежду, но покупка вещей длительного пользования вызывает затруднения;
  - 4) мы можем приобретать вещи длительного пользования, но не можем совершать дорогие покупки (машины);

- 5) без труда приобретаем вещи длительного пользования, можем накопить деньги на недорогую машину, но купить квартиру или коттедж не можем;
- 6) у нас нет материальных проблем, можем приобрести недвижимость, дорогой автомобиль, дорогой заграничный тур.

Анкета разработана на кафедре социологии и социальных технологий Тверского государственного технического университета. Объем выборки составил 100 респондентов.

Отметим, что среди опрошенных не было респондентов с материальным положением, соответствующим п. 1 и 2 вопроса 8.

Обработка результатов проводилась с помощью программы SPSS [5, 18].

Перед применением метода определим зависимую и независимые переменные, по которым будет осуществляться ветвление.

В качестве зависимой дихотомической переменной будет выступать вопрос 1, при ответе на который респонденты определяли свою политическую активность. Независимые переменные (включают ряд социально-демографических признаков) — вопросы 3, 5–8.

На первом шаге применения метода определяются априорные вероятности и цены ошибок классификации. В рассматриваемом примере априорные вероятности полагались оцениваемыми, а цены ошибок классификации — равными.

Для построения дерева классификации использовался метод CART. В качестве условия прекращения ветвления была выбрана прямая остановка по методу FACT.

Результат применения метода — дерево классификации, позволяющее провести наглядную и содержательную интерпретацию результатов (рис. 2). Дерево содержит 8 терминальных вершин, или узлов (вершины 7–14), и 7 решающих правил.

Прокомментируем дерево, начиная с корневой вершины, в которой выборка делится на две части в зависимости от активности в общественной жизни: вершина 1 — принимающие участие в политической жизни (56 человек; активны в общественной жизни) и вершина 2 — не принимающие участие в политической жизни (44 человека; не активны в общественной жизни).

Вершина 1 в зависимости от курса разделяется на две ветви: вершина 3 — принимающие участие в политической жизни (13 человек; студенты 2-го курса) и вершина 4 — принимающие участие в политической жизни (43 человека; студенты старших курсов).

Вершина 3 в зависимости от материального положения разделяется на терминальные вершины 7 — принимающие участие в политической жизни (материальное положение, соответствующее вариантам 2–4 ответа на вопрос 8; 11 человек) и 8 — не принимающие участие в политической жизни (материальное положение, соответствующее вариантам 5, 6 ответа на вопрос 8; 2 человека).

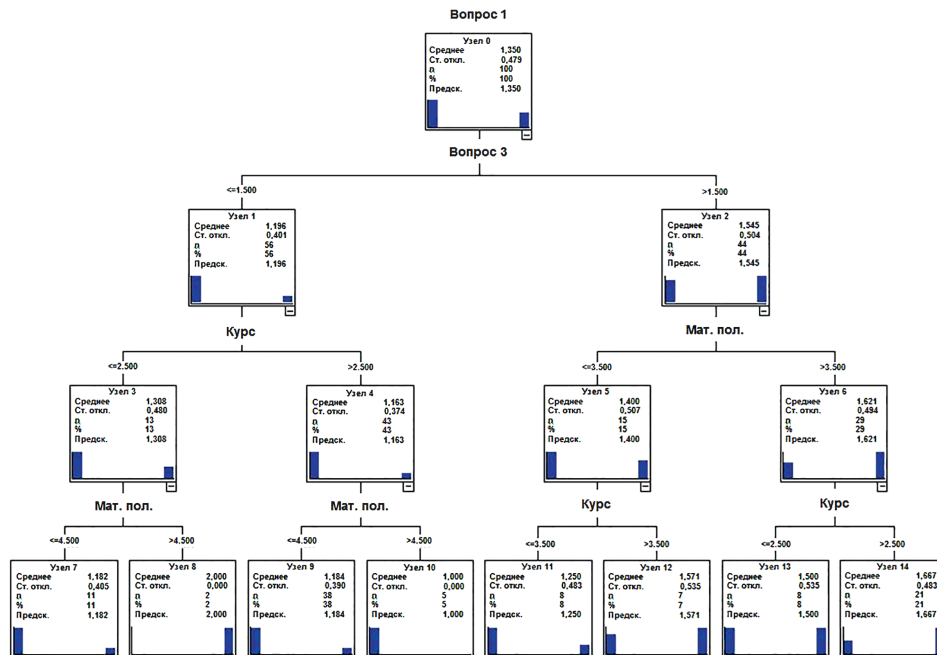


Рис. 2. Дерево классификации:

Мат. пол. — материальное положение

Вершина 4 разделяется на две терминальные вершины (9 и 10) в зависимости от материального положения, но обе они содержат респондентов, принимающих участие в политической жизни.

Вершина 5 в зависимости от курса разделяется на две терминальные: вершина 11 — принимающие участие в политической жизни (8 человек; студенты 2-го и 3-го курса) и вершина 11 — не принимающие участие в политической жизни (7 человек; студенты 4-го курса).

Вершина 6 в зависимости от курса разделяется на две терминальные: вершина 13 — принимающие участие в политической жизни (8 человек; студенты 2-го курса) и вершина 14 — не принимающие участие в политической жизни (7 человек; студенты 3-го и 4-го курса).

Анализируя дерево, можно прийти к выводу, что группа, включающая в себя респондентов, принимающих участие в политической жизни, — это студенты 2-го и 3-го курса, материальное положение которых соответствует вариантам 3, 4 ответа на вопрос 8 (средний уровень достатка).

Группа, включающая в себя респондентов, не принимающих участие в политической жизни, — это студенты старших курсов, материальное положение которых соответствует вариантам 5, 6 ответа на вопрос 8 (высокое материальное положение).

Значимость переменных распределяется следующим образом: наиболее значимая (т. е. переменная, оказывающая наибольшее влияние на политическую активность) — ответ на вопрос 3 (ранг 100), следующая по значимости — материальное положение (ранг 80,7), далее — курс (ранг 38,5) и пол (ранг 10), наименее значимая пере-



менная — специальность (ранг 0). Таким образом, переменные «пол» и «специальность» не участвуют в формировании решающих правил, т. е. не влияют на политическую активность.

*V*-кратная кросс-проверка с параметром  $V = 10$  показала, что общий процент неверных предсказаний, или риск, равен 0,147. Это значение говорит о приемлемом качестве дерева.

Построенное дерево классификации позволило охарактеризовать группы молодежи, принимающей и не принимающей участие в политической жизни. Полученные результаты содержательно адекватны и поддаются интерпретации.

**Заключение.** Итак, ДК позволяют: прогнозировать принадлежность объектов к определенной группе в зависимости от значения признаков, которые характеризуют объект; выделять наиболее существенные признаки, участвующие в формировании решающих правил, согласно которым объект будет отнесен к той или иной группе; проводить описание выделенных групп.

В социологии метод может применяться при анализе результатов анкетирования или опроса, когда необходимо осуществить поиск скрытых закономерностей в исходном массиве данных.

## ЛИТЕРАТУРА

- [1] Буреева Н.Н. *Многомерный статистический анализ с использованием ППП «STATISTICA»: учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики»*. Нижний Новгород, Нижегородский государственный университет им. Н.И. Лобачевского, 2007, 112 с.
- [2] Дубров А.М. *Многомерные статистические методы*. Москва, Финансы и статистика, 2003, 352 с.
- [3] Дюран Б. *Кластерный анализ*. Москва, Статистика, 1977, 128 с.
- [4] Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. *Факторный, дискриминантный и кластерный анализ*. Москва, Финансы и статистика, 1989, 215 с.
- [5] Тюрин В.В., Щеглов С.Н. *Дискриминантный анализ в биологии*. Краснодар, Кубанский государственный университет, 2015, 126 с.
- [6] Бессокирная Г.П. Факторный анализ: традиции использования и новые возможности. *Социология: методология, методы, математическое моделирование*, 2000, № 12, с. 142–153.
- [7] Губанова Н.Ю. Факторный анализ данных и возможности его использования при обработке данных психологического эксперимента. *Sochi Journal of Economics*, 2008, № 3–4, с. 178–187.
- [8] Жуковская В.М., Мучник И.Б. *Факторный анализ в социально-экономических исследованиях*. Москва, Статистика, 1976, 152 с.
- [9] Иберла К. *Факторный анализ*. Москва, Статистика, 1980, 389 с.
- [10] Фомина Е.Е., Жиганов Н.К. Методика обработки результатов анкетирования с использованием методов многомерной и параметрической статистики. *Вестник Пермского национального исследовательского политехнического университета. Социально-экономические науки*, 2017, № 1, с. 106–115.

- [11] Фомина Е.Е. Факторный анализ и категориальный метод главных компонент: сравнительный анализ и практическое применение для обработки результатов анкетирования. *Гуманитарный вестник*, 2017, вып. 10 DOI: 10.18698/2306-8477-2017-10-473
- [12] Фомина Е.Е. Применение факторного анализа для обработки результатов анкетирования. *Социосфера*, 2016, № 3, с. 122–127.
- [13] Анализ качественных признаков на основе логлинейной модели. URL: [https://studopedia.ru/6\\_28922\\_analiz-kachestvennih-priznakov-na-osnove-loglineynou-modeli.html](https://studopedia.ru/6_28922_analiz-kachestvennih-priznakov-na-osnove-loglineynou-modeli.html) (дата обращения 26.01.2018).
- [14] Аптон Г. *Анализ таблиц сопряженности*. Москва, Финансы и статистика, 1982, 143 с.
- [15] Буре В.М., Гливинская О.А., Сотников А.В. Логлинейный анализ базы данных по инфаркту миокарда у больных молодого и среднего возраста. *Вестник Санкт-Петербургского университета*. Сер. 10, 2010, вып. 1, с. 35–41.
- [16] Крымзин Д.Н. Применение логлинейного анализа для исследования зависимости оценки кадрового потенциала вуза от характеристик преподавателей. *Вестник НГУЭУ*, 2014, № 2, с. 134–141.
- [17] Логлинейный анализ. URL: <http://www.statmethods.ru/konsalting/statistics-metody/114-loglinejnyj-analiz.html> (дата обращения 21.01.2018).
- [18] Толстова Ю.Н., Рыжова А.В. Анализ таблиц сопряженности: использование отношения преобладаний и логлинейных моделей. *Социология 4М*, 2003, № 16, с. 150–164.
- [19] Трофимов Д.А. Логлинейный анализ таблиц мобильности: обзор основных моделей. *Социология 4М*, 2008, № 26, с. 119–138.
- [20] Фомина Е.Е. Подготовка и анализ результатов анкетирования с применением математических методов. *Социосфера*, 2018, № 2, с. 194–198.
- [21] Фомина Е.Е. Использование методов многомерной статистики для анализа социальной и экономической информации. *Экономика. Социология. Право*, 2018, № 2, с. 61–67.
- [22] Бова А. Деревья решений как техника добычи данных. *Социология: теория, методы, маркетинг*, 2002, № 1, с. 128–136.
- [23] Деревья классификации. URL: [http://www.statlab.kubsu.ru/sites/project\\_bank/trees.pdf](http://www.statlab.kubsu.ru/sites/project_bank/trees.pdf) (дата обращения 17.09.2018).
- [24] Деревья классификации. Основные идеи. URL: <http://statsoft.ru/home/textbook/modules/stclatre.html> (дата обращения 17.09.2018).
- [25] Толстова Ю.Н. *Анализ социологических данных. Методология, дескриптивная статистика, изучение связей между номинальными признаками*. Москва, Научный мир, 2000, 352 с.

Статья поступила в редакцию 16.11.2018

Ссылку на эту статью просим оформлять следующим образом:

Фомина Е.Е. Возможности метода деревьев классификации при обработке социологической информации. *Гуманитарный вестник*, 2018, вып. 11 <http://dx.doi.org/10.18698/10.18698/2306-8477-2018-11-574>

**Фомина Елена Евгеньевна** — канд. техн. наук, доцент кафедры «Информатика и прикладная математика» Тверского государственного технического университета. e-mail: [f-elena2008@yandex.ru](mailto:f-elena2008@yandex.ru)

## **The potential of the method of classification trees in the sociological information processing**

© E.E. Fomina

Tver State Technical University, Tver, 170026, Russia

*Methods of multidimensional intelligence and reconnaissance data analysis are used in practice to solve the problems of analyzing the data structure and obtaining information about the dependencies between variables describing an object or a phenomenon. Such methods include cluster, discriminant, factor, log linear analyses and a number of other methods. One of the most flexible methods is the classification tree method, which allows solving problems of classification and forecasting. The method can be practically used in various fields, including sociology. The article considers the algorithm of the classification tree method and its possibilities for analyzing sociological information, in particular, for analyzing the results of a questionnaire design of youth political activity.*

**Keywords:** *classification tree, processing sociological information, questionnaire design*

### REFERENCES

- [1] Bureeva N.N. *Mnogomernyy statisticheskiy analiz s ispolzovaniem PPP «STATISTICA»* [Multivariate statistical analysis using the software “STATISTICA”]. Nizhny Novgorod, Lobachevsky Nizhny Novgorod State University, 2007, 112 p.
- [2] Dubrov A.M. *Mnogomernye statisticheskie metody* [Multivariate statistical methods]. Moscow, Finansy i statistika Publ., 2003, 352 p.
- [3] Duran B., Odell P. *Cluster analysis*. A survey. Berlin – Heidelberg – N.Y. Springer-Verlag, 1974 [In Russ.: Duran B., Odell M. Cluster analysis. Moscow, Statistika Publ., 1977, 128 p.].
- [4] Kim Jae-On, Mueller Ch.W., Klecka W.R., *Factor, Discriminant, and Cluster Analysis*. Beverly Hills, CA, Sage Publications, 1989 [In Russ.: Kim Jae-On, Mueller Ch.W., Klecka W.R., Faktornyy, discriminantnyy i klasternyy analiz. Moscow, Finansy i statistika Publ., 1989, 215 p.].
- [5] Tyurin V.V., Shcheglov S.N. *Diskriminantnyy analiz v biologii* [Discriminant analysis in biology]. Krasnodar, Kuban State University Publ., 2015, 126 p.
- [6] Bessokirnaya G.P. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie (4M) — Sociology: methodology, methods, mathematical modeling (4M)*, 2000, no. 12, pp. 142–153.
- [7] Gubanov N.Yu. *Sochi Journal of Economy*, 2008, no. 3-4, pp. 178–187.
- [8] Zhukovskaya V.M., Muchnik I.B. *Faktornyy analiz v sotsialno-ekonomicheskikh issledovaniyakh* [Factor analysis in socio-economic research]. Moscow, Statistika Publ., 1976, 152 p.
- [9] Überla K. *Faktorenanalyse*. Berlin, Springer Verlag Publ., 1968 [In Russ.: Überla K. Factor analysis. Moscow, Statistika, 1980, 389 p.].
- [10] Fomina E.E., Zhiganov N.K. *Vestnik Permskogo natsionalnogo issledovatel'skogo politekhnicheskogo universiteta. Sotsialno-ekonomicheskie nauki — PNRPU Sociology and Economics Bulletin*, 2017, no. 1, pp. 106–115.
- [11] Fomina E. E. *Gumanitarnyy vestnik — Humanities Bulletin of BMSTU*, 2017, no. 10 (60). DOI: 10.18698/2306-8477-2017-10-473
- [12] Fomina E. E. *Sotsiosfera — Sociosphere*, 2016, no. 3, pp. 122–127.
- [13] *Analiz kachestvennykh priznakov na ocnove loglineynoy modeli* [Analysis of qualitative features based on loglinear model]. Available at:

- [https://studopedia.ru/6\\_28922\\_analiz-kachestvennih-priznakov-na-osnove-loglineynoy-modeli.html](https://studopedia.ru/6_28922_analiz-kachestvennih-priznakov-na-osnove-loglineynoy-modeli.html) (accessed January 26, 2018).
- [14] Upton G.J.G. *The Analysis of Cross-tabulated Data*. Chichester, New York, Brisbane, Toronto, John Wiley & Sons Publ., 1978 [In Russ.: Upton G. Analiz tablits sopryazhennosti. Moscow, Finansy i statistika Publ., 1982, 143 p.].
- [15] Bure V.M., Glivinskaya O.A., Sotnikov A.V. *Vestnik Sankt-Peterburgskogo Universiteta — Vestnik of Saint Petersburg University*, 2010, Ser. 10, no. 1, pp. 35–41.
- [16] Krymsin D.N. *Vestnik NGUEU — Vestnik of NSUEM*, 2014, no. 2, pp. 134–141.
- [17] *Loglinear analysis* [Electronic resource]. Available at: <http://www.statmethods.ru/konsalting/statistics-metody/114-loglinejnyj-analiz.html> (accessed January 21, 2018).
- [18] Tolstova Yu.N., Ryzhova A.V. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie (4M) — Sociology: methodology, methods, mathematical modeling (4M)*, 2003, no. 16, pp. 150–164.
- [19] Trofimov D.A. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie (4M) — Sociology: methodology, methods, mathematical modeling (4M)*, 2008, no. 26, pp. 119–138.
- [20] Fomina E.E. *Sotsiosfera — Sociosphere*, 2018, no. 2, pp. 194–198.
- [21] Fomina E.E. *Ekonomika, Sotsiologiya, Pravo (Economy. Sociology. Law)*, 2018, no. 2 (10), pp. 61–67.
- [22] Bova A. *Sotsiologiya: teoriya, metody, marketing (Sociology: theory, methods, marketing)*, 2002, no. 1, pp. 128–136.
- [23] *Derevya klassifikatsii* [Classification trees]. Available at: [http://www.statlab.kubsu.ru/sites/project\\_bank/trees.pdf](http://www.statlab.kubsu.ru/sites/project_bank/trees.pdf) (accessed September 17, 2018).
- [24] *Derevya klassifikatsii. Osnovnye idei* [Classification trees. Basic ideas]. Available at: <http://statsoft.ru/home/textbook/modules/stclatre.html> (accessed September 17, 2018).
- [25] Tolstova Yu.N. *Analiz sotsiologicheskikh dannykh. Metodologiya, diskriptivnaya statistika, izuchenie svyazey mezhdu nominalnymi priznakami* [Analysis of sociological data. Methodology, descriptive statistics, study of relationships between nominal characteristics]. Moscow, Nauchnyy mir Publ., 2000, 352 p.

**Fomina E.E.**, Cand. Sc. (Eng.), Assoc. Professor, Department of Informatics and Applied Mathematics, Tver State Technical University.  
e-mail: f-elena2008@yandex.ru