

В.М. Замятин¹

¹ Федеральное государственное бюджетное образовательное учреждение высшего образования "Иркутский государственный университет путей сообщения". г. Иркутск. Российская Федерация.

ПРИМЕНЕНИЕ КАРТ КОХОНЕНА В СОЦИОЛОГИЧЕСКОМ ТЕСТИРОВАНИИ

Аннотация. В данной статье изучается применение методов Data Mining, в частности, карт Кохонена, в социологическом тестировании на уровень риска наркотизации. Статья акцентирует внимание на использовании искусственных нейронных сетей для анализа данных и выявления закономерностей, с целью более точного определения индивидуального уровня риска наркотизации. Исследование предлагает новый подход к обработке социологических данных, который может привести к более эффективным методам предотвращения и борьбы с наркотиками.

Ключевые слова: социологическое исследование, карты Кохонена, кластеризация, Deductor Academic, информационная система, социологическое тестирование.

V.M. Zamyatin¹

¹ Federal State Budget Educational Institution of Higher Education "Irkutsk State Transport University." Irkutsk the Russian Federation.

USING OF KOHONEN MAPS IN SOCIOLOGICAL TESTING.

Abstract. This article examines the use of Data Mining methods, in particular, Kohonen maps, in sociological testing for the level of risk of anesthesias. The article focuses on the use of artificial neural networks for data analysis and identification of patterns, in order to more accurately determine the individual risk level of anesthesias. The study suggests a new approach to the processing of sociological data, which may lead to more effective methods of drug prevention and control.

Keywords: sociological research, Kohonen maps, clustering, Deductor Academic, information system, sociological testing.

Введение

В результате развития информационного общества, многие аспекты жизни людей, включая социологические опросы, подверглись автоматизации. Это имело значительное влияние на устранение недостатков проведения социологических опросов.

Среди методов обработки и анализа социологической информации выделяют две группы методов: статистическая обработка и интеллектуальный анализ данных. Традиционный подход при анализе социологических результатов социологических опросов основан на использовании статистических методов, но главным ограничением такого подхода является невозможность обнаружения неоднородных, локальных взаимосвязей, присутствующих в социальных явлениях, поскольку в основе методов статистической обработки обычно лежит концепция усреднения по выборке [1].

Методы интеллектуального анализа данных – это новое, более перспективное направление, стремительно развивающееся в настоящее время. Интеллектуальный анализ данных – это процесс обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретаций знаний, необходимых для принятия решений в различных сферах человеческой деятельности [2].

Социологическое тестирование на уровень риска наркотизации является важной задачей, так как это позволяет определить уязвимые группы населения и разработать эффективные профилактические мероприятия. В последние годы для этой цели все чаще применяются методы Data Mining, основанные на машинном обучении и искусственных нейронных сетях. Одним из таких методов являются карты Кохонена, которые позволяют выявить взаимосвязи между различными факторами и определить наиболее значимые из них [3].

Сбор данных

Для изучения проблемы наркотизации сначала необходимо собрать социологические данные по определенным критериям. Эти данные можно получить из различных источников, таких как опросы, статистические данные и медицинские записи. Центру развития человеческого капитала была поставлена задача выявить социально-психологические факторы риска и защиты немедицинского употребления наркотиков. Для чего был модифицирован и адаптирован тест «Исходная оценка наркотизации» (автор Г.В. Латышев и др.) [4].

Участникам исследования предлагается ответить на вопросы анкеты (79 вопросов). Все личные данные студентов обезличиваются. Респонденты отмечают только свой пол и возраст. В инструкции подчеркивается важность личного мнения каждого обучающегося и необходимость ответов, исходящих из собственных представлений о данной проблеме. Также отмечается необходимость независимой оценки, без вариантов совместного обсуждения.

Для обработки результатов были выделены следующие факторы риска и защиты в контексте изучения наркотизации:

1. Семейные:

- отношения с родителями
- смена места жительства
- система контроля в семье
- конфликтность в семье.

2. Индивидуальные:

- успешность
- отношение к употреблению психоактивных веществ
- отношение к насилию
- протестные реакции
- наличие позитивных жизненных ориентиров
- опыт употребления психоактивных веществ
- наличие кризисных ситуаций
- лозунг жизни.

3. Взаимоотношения со сверстниками:

- влияние среды. Ассоциирование с асоциальным поведением
- социально - психологический климат микросреды.

4. Общественные (социальные):

- участие в общественной деятельности
- отношение к религии
- доступность психоактивных веществ
- социальная «близость» с потребителями психоактивных веществ
- связь с микросоциумом

5. Образовательные:

- успеваемость
- посещаемость
- участие в студенческом сообществе
- организация досуга
- отношения с преподавателями
- социально-психологический климат в студенческом коллективе
- интерес к учебной деятельности
- смена образовательной организации
- связь образовательной организации с семьей.

Ответ на каждый вопрос соответствует количеству баллов по уровню риска: высокий риск – 4 балла, средний риск – 3 балла, низкий риск 2 балла, риска нет – 1 балл. Сумма баллов, полученных за каждый вопрос, и составляет общий уровень риска.

В качестве респондентов исследования принимали участие студенты ИрГУПС в количестве 2087 человек, и студенты БГУ в количестве 2661 человек. По результатам обработки было выявлено два респондента с высоким уровнем риска, чьи данные были отправлены эксперту психологу, для дальнейшей личной беседы. Для социологического исследования использовалась программно-информационная система АССОЛЬ, которая формирует статистическую сводку по респондентам, далее её получает и обрабатывает эксперт-социолог [5].

Предварительная обработка данных

Предварительная обработка данных включает в себя очистку, нормализацию и трансформацию данных. Целью этого этапа является преобразование сырых данных в формат, подходящий для анализа с помощью карт Кохонена. На этом этапе респонденты, ответившие не на все вопросы, исключались из выборки данных.

Обучение карт Кохонена

На этапе обучения карт Кохонена, искусственная нейронная сеть настраивается на имеющиеся данные. Это позволяет выявить закономерности и взаимосвязи между различными факторами, которые могут влиять на уровень риска наркотизации.

Алгоритм функционирования самообучающихся карт (Self-Organizing Maps – SOM) представляет собой один из вариантов кластеризации многомерных векторов.

Примером таких алгоритмов может служить алгоритм k-средних. Важным отличием алгоритма SOM является то, что в нем все нейроны упорядочены в некоторую структуру.

При этом в ходе обучения модифицируется не только нейрон-победитель, но и его соседи. За счет этого SOM можно считать одним из методов проецирования многомерного пространства в пространство с более низкой размерностью [6-9].

В качестве функции расстояния была использована Гауссова функция.

При анализе карт рекомендуют использовать сразу несколько экземпляров (рис. 1).

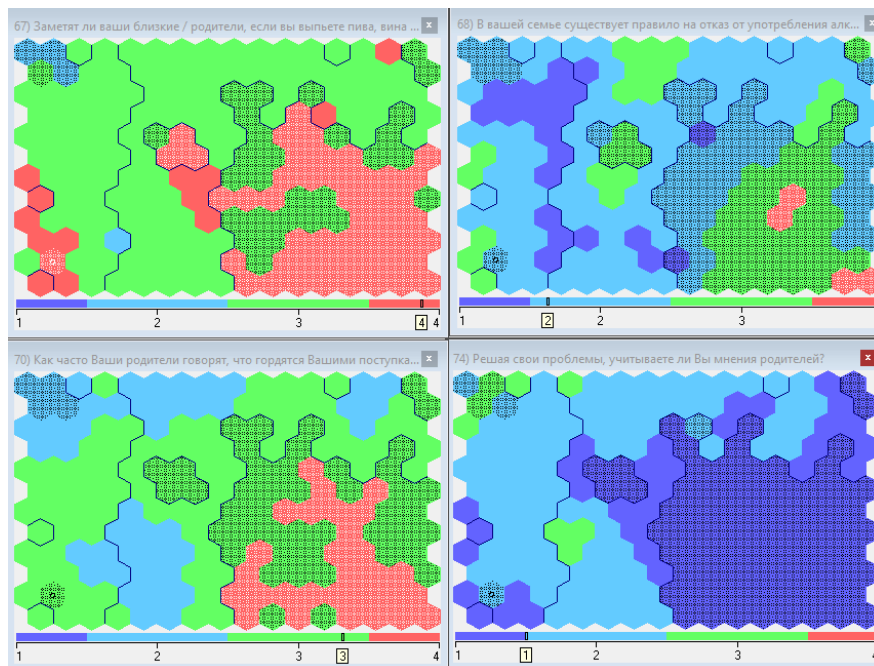


Рис. 1 Карты четырех входов

Например, на карте, отображающей результаты вопроса «Как часто Ваши родители говорят, что гордятся Вашими поступками?», ответы «всегда» расположены в нижнем правом углу. Рассматривая одновременно три карты, мы можем сказать, что эти же объекты

имеют похожие показатели на карте шестьдесят семь, а на карте семьдесят четыре обратная картина. Это лишь фрагмент вывода, который можно сделать, исследуя карту.

Результаты

После обучения карт Кохонена можно анализировать полученные результаты и выявить наиболее значимые факторы, влияющие на уровень риска наркотизации. Такие факторы могут включать, например, социально-экономический статус, уровень образования, наличие семейных проблем или определенных психологических особенностей. Выявление этих факторов позволит разработать более эффективные профилактические мероприятия и снизить риск наркотизации в уязвимых группах населения.

Результаты исследования показывают, что применение методов Data Mining, в частности, карт Кохонена, может быть полезным инструментом в социологическом тестировании на уровень риска наркотизации. Искусственные нейронные сети способны обрабатывать большие объемы данных и выявлять сложные взаимосвязи между различными факторами, что может помочь в более точном определении индивидуального уровня риска.

Однако стоит отметить, что результаты исследования могут зависеть от качества собранных данных и выбора параметров обучения карт Кохонена. В дальнейшем исследовании можно использовать дополнительные методы Data Mining для сравнения результатов и улучшения точности прогнозирования уровня риска наркотизации [10, 11].

Заключение

В данной статье был рассмотрен подход к применению методов Data Mining, основанных на картах Кохонена, в социологическом тестировании на уровень риска наркотизации. Результаты исследования показывают, что такой подход может быть полезным для выявления наиболее значимых факторов и разработки эффективных профилактических мероприятий. В дальнейшем, возможно развитие данного направления исследований, включая использование других методов Data Mining и совершенствование подходов к обучению искусственных нейронных сетей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Обработка результатов социологических опросов граждан с помощью методов интеллектуального анализа данных, Подвесовский А. Г., Лагерев Д. Г., Бабурин А. Н.
2. Паклин, Н. Б. Бизнес-аналитика: от данных к знаниям: / Н. Б. Паклин, В. И. Орешков / Учебное пособие. 2-е изд., испр. – СПб: Питер, 2013. – 704 с.
3. Информация из ИНТЕРНЕТ. «Областной центр диагностики и консультирования» URL: <http://concord.websib.ru> (дата обращения 01.05.21).
4. Организация профилактики наркозависимости в условиях сельского образовательного учреждения: Сборник методических рекомендаций. Авторский коллектив: Латышев Г.В., Речнов Д.Д., Титова О.А., Орлова М.В., Яцышин С.М., Серeda В.М., Уминская Л.Л., Гериш А.А. и др., Санкт-Петербург, 2001.
5. Замятин, В. М. Разработка по для тестирования обучающихся и персонала ИРГУПС / В. М. Замятин, В. В. Федоров // Молодая наука Сибири. – 2020. – № 3(9). – С. 190-196.
6. T.Kohonen, «Self-Organizing Maps» (2-nd edition), Springer, 1997.
7. Информация из ИНТЕРНЕТ. Описание платформы - BaseGroup Labs. URL: <https://basegroup.ru> (дата обращения 01.05.21).
8. Л.И. Дубровская, Г.Б. Князев, Компьютерная обработка естественно-научных данных методами многомерной прикладной статистики, 2011
9. Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И. Методы и модели анализа данных: OLAP и Data Mining. -СПб.: БХВ-Петербург, 2004.
10. Ю.А. Кузнецов, В.И. Перова, Д.С. Семиков, Информационные и коммуникационные технологии как фактор развития цифровой экономики в российской федерации, 2017

11. Мальцева А.В., Использование кластерного анализа в парадигме Data Mining для изучения структуры рынка труда, Алтайский государственный университет

REFERENCES

1. Processing of the results of sociological surveys of citizens using data mining methods, Podesovsky A. Year., Lagerev D. Year. Baburin A. N.
2. Paklin, N. B. Business analytics: from data to knowledge: / N. B. Paklin, V. I. Oreshkov / Study guide . 2nd ed., ispr. St. Petersburg: Peter, 2013. - 704 P.
3. Information from the INTERNET. "Regional Center for Diagnostics and Counseling" URL: <http://concord.websib.ru> (01.10.22).
4. Organization of drug addiction prevention in a rural educational institution: A collection of methodological recommendations. Team of authors: Latyshev G.V., Rechnov D.D., Titova O.A., Orlova M.V., Yatsyshin S.M., Sereda V.M., Uinskaya L.L., Gerish A.A. and others, St. Petersburg, 2001.
5. Zamyatin, V. M. Software development for testing students and staff of IRGUPS / V. M. Zamyatin, V. V. Fedorov // Young Science of Siberia. – 2020. – № 3(9). – Pp. 190-196.
6. T.Kohonen, "Self-Organizing Maps" (2nd edition), Springer, 1997.
7. Information from the INTERNET. Description of the platform - BaseGroup Labs. URL: <https://basegroup.ru> (01.10.22).
8. L.I. Dubrovskaya, G.B. Knyazev, Computer processing of natural science data by methods of multidimensional applied statistics, 2011
9. Barseghyan A. A., Kupriyanov M. S., Stepanenko V. V., Kholod I. I. Methods and models of data analysis: OLAP and Data Mining. -St. Petersburg: BHV-Petersburg, 2004.
10. Yu.A. Kuznetsov, V.I. Petrova, D.S. Semikov, Information and communication technologies as a factor in the development of the digital economy in the Russian Federation, 2017
11. Maltseva A.V., Using cluster analysis in the Data Mining paradigm to study the structure of the labor market, Altai State University

Информация об авторах

В.М. Замятин – аспирант группы ИВТ.1-20-1 ФГБОУ ВО ИрГУПС, сотрудник управления информатизации ФГБОУ ВО ИрГУПС. Email: zamiatin_vm@irgups.ru.

Information about the authors

V.M. Zamyatin is a graduate student of the group IVT.1-20-1 IRGUPS and an employee of the " department of informatization" department of Irkutsk State Transport University. Irkutsk the Russian Federation. Email: zamiatin_vm@irgups.ru.