

УДК 004.048:316.3

Бочарова Татьяна Александровна

кандидат социологических наук,
доцент кафедры информатики,
Тихоокеанский государственный университет
kitaal@yandex.ru

Tatiana A. Bocharova

Candidate of Sociological Sciences,
Associate Professor of Informatics,
Pacific national university,
kitaal@yandex.ru

РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ СОЦИАЛЬНОГО ПОВЕДЕНИЯ С ПОМОЩЬЮ АППАРАТА НЕЙРОННЫХ СЕТЕЙ¹

SOLUTION OF THE PROBLEM OF CLASSIFICATION OF SOCIAL BEHAVIOR BY USING AN EQUIPMENT OF NEURAL NETWORKS

***Аннотация.** Статья посвящена результатам использования аппарата нейронных сетей для решения задачи классификации социального поведения в ситуации конкретного выбора в рамках исследования виртуальной коммуникации. Анализ социологических данных усложняется тем, что большинство представляемых данных плохо поддается формализации. Этот факт затрудняет применение статистических методов обработки и отрицательно влияет на достоверность результатов. Особенности социальных объектов исследования повышают необходимость в разработке более точных и мощных инструментов, обладающих широкими функциональными возможностями.*

***Ключевые слова:** социологические исследования, интеллектуальный анализ данных, классификация, кластеризация.*

***Annotation.** The article is devoted to the results of using the apparatus of neural networks to solve the problem of classifying social behavior in a situation of a particular choice in the study of virtual communication. The analysis of sociological data is complicated by the fact that most of the data presented is difficult to formalize. This fact makes it difficult to use statistical processing methods and negatively affects the reliability of the results. Features of social objects of research increase the need for the development of more accurate and powerful tools with wide functional capabilities.*

Key words: sociological research, intelligent data analysis, classification, clustering.

¹ Статья выполнена в рамках гранта по совместной программе Министерства науки и высшего образования РФ и DAAD «Иммануил Кант»

Социологические исследования являются необходимым средством изучения социальной действительности, с помощью которого осуществляется познание сущности происходящих в обществе процессов и явлений. Полученные выводы позволяют оценить результаты планируемых действий, повысить их эффективность, предотвратить возможные ошибки или минимизировать последствия. Данные, собранные в ходе социологических экспериментов, способствуют разрешению проблем в разных областях жизни и деятельности человека. В последние годы усиливается влияние эмпирической социологии и возрастает роль эмпирических исследований, что приводит к необходимости обрабатывать большие количественные данные и росту качественных составляющих [5].

Возрастающая сложность объектов исследования и увеличивающиеся требования к системам их анализа повышают необходимость в разработке более точных и мощных инструментов, обладающих широкими функциональными возможностями. Разум человека для восприятия больших массивов неоднородной информации и установления взаимосвязей в выборках использует ассоциативное мышление, принципы обучения и самообучения, причинно-следственные связи. Поэтому использование нейронных сетей для задач, не имеющих стандартного алгоритма решения, является наиболее эффективным способом расширения функциональных и аналитических возможностей человеческого мозга. Нейронные сети представляют собой новый инструмент анализа данных. Их применение оправдано, когда формализация способа решения задачи затруднена или невозможна [3].

Нейросети не требуют программирования и предполагают работу по обучению искусственной нейронной сети на наборе примеров [2]. Основная задача обучения нейронной сети, имитирующей работу мозга, состоит в узнавании, умении выделять сходства и различия. На этапе обучения возникают определяющие связи между входными параметрами, и оформление образов, которые в дальнейшем будут использованы при реализации решения на сети. Искусственная нейронная сеть при работе принимает значение входных сигналов и выдает значения выходных параметров. Следовательно, задачи, в которых не определен характер связи между входом и выходом, составляют основной класс задач, решаемых с помощью нейронных сетей. Зависимость между входными и выходными данными может быть установлена в процессе обучения НС.

Решение задач классификации является важнейшей областью применения нейронных сетей [1]. Задачи классификации направлены на определение класса, к которому можно отнести входной набор данных. В рамках настоящего исследования была поставлена задача провести классификацию социальных типов в рамках виртуального сетевого взаимодействия. На основании имеющихся данных определить структуру социальной общности путем выделения в группы респондентов, имеющих схожие результаты опроса, а значит действующих аналогично в социально значимой исследуемой ситуации.

Целью вычислительного эксперимента была разработка методологической базы и технологий применения нейронных сетей при обработке данных социологических исследований на примере анализа виртуальной коммуникации.

Для анализа были использованы материалы социологического исследования, проведенного в рамках гранта по совместной программе Министерства науки и высшего образования РФ и DAAD «Иммануил Кант 2019–2020» по исследованию процессов виртуализации общества.

При решении задачи были задействованы 6 признаков, отражающих выбор типа виртуальной площадки опрашиваемого. Из всей совокупности респондентов на первом этапе эксперимента использованы данные 110 человек. Обучающий набор был представлен в форме результатов наблюдения с указанием значения входных и выходных переменных. Использование большего числа входных переменных будет способствовать получению более точного прогноза. Однако в рамках решения поставленной задачи общее множество было сокращено до шести основных с помощью корреляционного анализа.

Экспериментальные данные первоначально носили нечисловой характер и имели вид номинальных переменных, например:

$$\text{пол} = \begin{cases} \text{мужской;} \\ \text{женский.} \end{cases} \quad \text{образование} = \begin{cases} \text{неполное среднее;} \\ \text{среднее;} \\ \text{среднее специальное;} \\ \text{незаконченное высшее;} \\ \text{высшее;} \\ \text{наличие ученой степени.} \end{cases}$$

Нейронные сети лучше работают с числовыми данными, лежащими в определенном интервале. В связи с этим номинальные значения переменных были представлены в числовом виде.

X1 – возраст (лет), значения:

- | | |
|------------|------------|
| 1 – до 20; | 5 – 36-40; |
| 2 – 21-25; | 6 – 41-50; |
| 3 – 26-30; | 7 – 51-60. |
| 4 – 31-35; | |

X2 – пол, значения:

- | | |
|--------------|--------------|
| 1 – мужской; | 2 – женский. |
|--------------|--------------|

X3 – образование, значения:

- | | |
|--------------------------|-----------------------------|
| 1 – неполное среднее; | 4 – незаконченное высшее; |
| 2 – среднее; | 5 – высшее; |
| 3 – среднее специальное; | 6 – наличие ученой степени. |

X4 – род занятий, значения:

- | | |
|--|----------------------------------|
| 1 – неработающий (безработный, домохозяйка); | |
| 2 – учащийся (студент); | 3 – неквалифицированный рабочий; |
| 4 – квалифицированный рабочий; | |
| 5 – работник сферы обслуживания; | |
| 6 – работник образовательной сферы; | |

- 7 –военнослужащий (МВД, налоговая полиция, таможня);
- 8 – индивидуальный предприниматель (самозанятость);
- 9 – пенсионер (неработающий).

X5 – наличие детей, значения:

- 1 – нет;
- 2 – 1-2 детей;
- 3 – 3-е детей и больше.

X6 – тип форумов, значения:

- 1 – универсальные;
- 2 – профессиональные;
- 3 – развлекательные (тематические по интересам).

В качестве инструментария был использован программный комплекс «Нейронная сеть для анализа (прогнозирования) социального поведения респондентов в ситуации конкретного выбора», разработанный автором (свидетельство о государственной регистрации программы для ЭВМ № 2019661470).

Решение задачи сводится к решению задачи кластеризации: необходимо выделить однородные группы респондентов на основе результатов проведенного опроса, занесенных в базу данных. В процессе обучения сеть учится понимать структуру данных и самостоятельно определяет классы для разделения обучающего множества, что позволяет найти неявные или скрытые закономерности в исходных данных, установить факторы, оказывающие максимальное влияние на полученные классы.

Респонденты были классифицированы на 9 групп. Результаты кластерного анализа можно представить в форме таблицы 1, характеризующей количественные показатели каждого кластера.

Таблица 1.

Количественные показатели кластеров

Кластер	Количество респондентов	Среднее значение входных признаков (центр кластера)					
		X1	X2	X3	X4	X5	X6
1	38	4,71	1	4,55	7,74	2,15	2
2	25	4,16	1	5,08	5,32	1	2
3	7	6	1	3	4	2	3
4	39	1,05	1	2,05	1,87	1	3
5	25	4,32	2	4,8	6,12	1,1	2
6	20	1,45	2	2,45	2,45	1	3
7	20	5,6	1	4,35	8,65	2,35	3
8	10	7	2	6	9	1	3
9	66	3,38	1,7	4,95	5,5	2,15	1

Полученные данные позволяют дать следующую интерпретацию результатов опроса. По типам предпочитаемых сетевых сообществ (показатель X6) всех участников опроса можно разделить на три группы:

- 1) пользователи, предпочитающие форумы универсальной направленности (35,2 %);
- 2) пользователи, посещающие профессиональные форумы (26,4 %);
- 3) пользователи, зарегистрированные на развлекательных площадках, посвященных хобби (38,4 %).

Согласно проведенному кластерному анализу, в первую группу входят респонденты 9 кластера, всего 66 человек. Особенности состава кластера 9: средний возраст 28,3 лет, преимущественно женщины с высшим образованием, работающие в сфере обслуживания или образования, имеющие детей.

Вторая группа, в свою очередь, состоит из трех подгрупп – кластеры 1, 2, 5, всего 88 человек. Кластер 1 характеризуется такими показателями: средний возраст 34 года; исключительно мужчины с детьми, имеющие образование не ниже незаконченного высшего; служащие в ведомственных структурах (МВД, прокуратура, налоговая полиция, таможня). Особенности кластера 2: средний возраст 31,1 года; мужчины, имеющие в основном высшее образование, занятые в сфере обслуживания и не имеющие детей. Кластер 5: женщины в возрасте около 30 лет с высоким уровнем образования (высшее, незаконченное высшее), не имеющие детей.

В третью группу входит 5 кластеров: 3, 4, 6, 7, 8, всего 96 человек. Особенности кластера 3: самый немногочисленный кластер, объединяющий мужчин в возрасте около 45 лет, имеющих рабочую профессию (образование среднее специальное) и занятых на производстве, количество детей не больше 2. Кластер 4: в него вошли молодые люди (исключительно мужского пола) в возрасте до 20 лет, студенты учебных заведений разного уровня, не имеющие детей. Кластер 6: имеет схожую возрастную структуру с 4 кластером, т. е. молодежь, не достигшая 20-ти летнего возраста, однако, половая принадлежность противоположная – только девушки, также обучающиеся в различных учебных заведениях (ВУЗы, СУЗы), подрабатывающие в сферах, не требующих квалификации, без детей. Кластер 7: мужчины, средний возраст около 40 лет, образование – среднее специальное или незаконченное высшее, представители самозанятой прослойки населения (малого бизнеса), детей не больше 2. Кластер 8: женщины пенсионного возраста с высоким уровнем образования (наличие ученой степени), не работающие и не имеющие детей. Относительно 8 группы можно предположить, что отсутствие детей означает отсутствие детей, находящихся на иждивении. В силу возрастной и гендерной особенности и указанной группы можно заключить, что ее представители имеют детей, однако те уже выросли и живут самостоятельно.

Приведенные результаты имеют достаточно высокую точность с погрешностью порядка 6 % и совпадают с данными, полученными автором в 2018 году в результате обработки ИНС типа самоорганизующейся карты Кохонена, реализуемой в пакете Deductor Studio (разработка компании BaseGroup Labs, г. Рязань, адрес в Интернете <https://basegroup.ru/>).

Меры в социологических исследованиях носят эмпирический характер [4]. Постановка задачи усложняется многообразием в формализации

эмпирических закономерностей и сложностью выбора подходящей процедуры классификации. Основанием для классификации может выступать не только критерий «схожести» элементов, но и определенная закономерная зависимость между характерными признаками, всеми или целевыми. Методы классификации с помощью параметров ограничений реализуют алгоритм классификации, осуществляющий разбиение объектов на классы после проверки наличия искомой закономерности.

Литература:

1. Богославский С. Н. *Современные методы нейросетевого исследования* // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ). 2006. №06(022). С. 48–61.

2. Касторнова В.А. *Искусственные нейронные сети как современные средства информатизации* // Информационная среда образования и науки. 2012. №7. С. 1–17.

3. Мальцева А В., Шилкина Н. Е., Махныткина О. В. *DATA MINING в социологии: опыт и перспективы проведения исследования* // Социологические исследования. 2016. № 3. С. 35–44.

4. Толстова Ю. Н. *Социология и компьютерные технологии* // Социологические исследования. 2015. № 8. С. 3–13.

5. Тощенко Ж. Т. *Теоретические и прикладные проблемы исследования новых явлений в общественном сознании и социальной практике* // Социологические исследования. 2010. N 7. С. 3 – 6.

Bibliography:

1. Bogoslavskij S. N. *Modern methods of neural network research. Politematicheskij setevoy elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Political Internet electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubSAU)]*, 2006, no. 06 (022), pp. 48–61.

2. Kastornova V.A. *Artificial neural networks as modern means of informatization. Informacionnaya sreda obrazovaniya i nauki [Information environment of education and science]*, 2012, no. 7, pp. 1–17.

3. Mal'ceva A V., SHilkina N. E., Mahnytkina O. V. *DATA MINING in Sociology: Experience and Prospects for Research. Sociologicheskie issledovaniya [Sociological researches]*, 2016, no. 3, pp. 34–44.

4. Tolstova YU. N. *Sociology and computer technology. Sociologicheskie issledovaniya [Sociological researches]*, 2015, no. 8, pp. 3–13.

5. Toshchenko ZH. T. *Theoretical and applied problems of the study of new phenomena in public consciousness and social practice. Sociologicheskie issledovaniya [Sociological researches]*, 2010, no. 7, pp. 3–6.