

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В УПРАВЛЕНИИ МИГРАЦИОННЫМИ ПРОЦЕССАМИ

FEATURES OF THE USE OF DATA MINING IN THE MANAGEMENT OF MIGRATION PROCESSES

Yu. Dmitrieva

Summary. The aim of the study is to assess the possibility of managing migration processes through the use of data mining. The article determines the possibility of using new technologies in migration management ("migration management"), which will allow to get the greatest effect from labor costs, when using new tools in process management. It is shown that the optimization and transformation of the data flow in process control will make it possible to abandon inefficient and outdated tools in the analysis, which will create conditions for facilitating work in the control system and affect the efficiency of the system as a whole. The possibilities of using new technologies in the process of managing migration processes are substantiated, using algorithms to study data and extract meaningful information from them to predict future trends. The scientific novelty lies in the substantiation of the possibilities of using new technologies in the analysis of data on population migration. As a result of the study, it was proved that the use of new technologies will contribute to the acceleration and high accuracy of scientific research in the field of population migration using information and communication both within the scientific field and between fields of knowledge. In addition, the practical use of data mining in the development of socio-economic and demographic policy measures will reduce errors and inaccuracies in the use of strategically ineffective tools.

Keywords: migration management, artificial intelligence, big data, machine learning.

Дмитриева Юлия Викторовна

Доктор экономических наук, доцент, ведущий научный сотрудник

Лаборатория проблем воспроизводства населения, Института социально-экономических проблем народонаселения Федерального научно-исследовательского социологического центра Российской академии наук
Москва
julia9770@mail.ru

Аннотация. Цель исследования дать оценку возможности управления миграционными процессами через использование интеллектуального анализа данных. В статье определяется возможность применения новых технологий в миграционном управлении («миграционный менеджмент»), что позволит получить наибольший эффект от трудовых затрат, при использовании новых инструментов в управлении процессом. Показано, что оптимизация и преобразование потока данных в управлении процессом, сделает возможным отказаться от малоэффективных и устаревших инструментов в анализе, что создаст условия для облегчения работы в системе управления и повлияет на эффективность работы системы в целом. Обоснованы возможности применения новых технологий в процессе управления миграционными процессами, используя алгоритмы для исследования данных и извлечения из них осмысленной информации для предсказания будущих тенденций. Научная новизна заключается в обосновании возможностей применения новых технологий в анализе данных о миграции населения. В результате исследования доказано, что применение новых технологий будет способствовать ускорению и высокой точности научных изысканий в области миграции населения с использованием информации и связи как внутри научной области, так и между областями знаний. Кроме того, практическое использование интеллектуального анализа данных при разработке мероприятий социально-экономической и демографической политики позволит сократить ошибки и неточности в применении стратегически неэффективных инструментов.

Ключевые слова: миграционное управление, искусственный интеллект, большие данные, машинное обучение.

Введение

Управление миграционными процессами носит характер решения насущных проблем, с которыми мы сталкиваемся в процессе управления. Однако эффективное управление любым процессом — это управление, направленное на поиск новых воз-

можностей. Цель исследования дать оценку возможности управления миграционными процессами через использование интеллектуального анализа данных. Основными задачами исследования являются: — определить возможность применения новых технологий в миграционном управлении («миграционный менеджмент»); — обосновать, что оптимизация и преобра-

зование потока данных в управлении процессом, позволит отказаться от малоэффективных и устаревших инструментов в анализе любого процесса, в данном случае миграционного; — дать оценку эффективности новых технологий в процессе управления миграционными процессами, используя алгоритмы для исследования данных и извлечения из них осмысленной информации для предсказания будущих тенденций; — показать возможности практического использования интеллектуального анализа данных при разработке мероприятий социально-экономической и демографической политики.

В статье применяются следующие методы исследования: научных обобщений, системного, экономического, исторического, критического, статистического и социологического анализа. Теоретико-методологическую основу исследования составили труды ведущих демографов, экономистов, социологов, а также агрегатор статей из ведущих блогов по Data Science, который охватывает такие ресурсы, как Simplified Analytics, Revolution Analytics, BrightPlanet и Big Data University. В работе использованы результаты научных и практических исследований отечественных и зарубежных ученых по исследуемой проблематике.

Основная часть

Если направить поиск новых возможностей в миграционном управлении, назовем это «миграционный менеджмент», то скорее всего получим и наибольший эффект от наших трудозатрат, т.к. приобретем новые инструменты в управлении процессом. Оптимизация всех ресурсов в управлении, сделает возможным отказаться от малоэффективных и устаревших инструментов и механизмов управления, по средствам вычленения или перенаправления ненужных действий со стороны системы управления. Таким образом, отказываясь от малоэффективных и устаревших инструментов управления, можно создать условия для облегчения работы в системе управления, что повлияет на производительность труда и эффективность работы системы в целом.

На первое место в эффективной системе управления любым процессом выходит маркетинг — управленческий процесс, который выявляет, предвидит и удовлетворяет требования потребителя (в данном случае государство) при соблюдении принципов эффективности и прибыльности (в данном случае от управления миграционными процессами). Безусловно, любой процесс, и в первую очередь, управление им, требует постоянного совершенствования и использования знаний. Кроме того, необходимо следить за тем, чтобы

происходящие события в системе управления не определяли, что делать, что решить, что отложить, что принять — нужно предупреждать события и быть к ним готовым.

Прогнозирование миграционных процессов, предупреждение негативных событий и явлений возможно и необходимо. Эффективность данного инструмента во многом будет зависеть от того, насколько руководству организации удастся систематизировать данные по процессу и преобразовать в инструмент, который способен сделать управление интеллектуально грамотным, качественным и, что самое важное, повысить скорость принятия решений. Миграционные процессы — это процессы, где изменения наступают быстро в зависимости от социально-экономической и политической обстановки в обществе, а управленческие решения вырабатываются годами. Поэтому применение новых технологий, а именно современных программ и методов обработки данных, в процессе миграционного управления необходимо.

В настоящее время для аналитического процесса используются программы статистической обработки данных, одной из которых является SPSS для Windows. Этот программный продукт предназначен для всех этапов аналитического процесса — планирования, сбора данных, доступа к данным и управления данными, анализа, создания отчетов и распространения результатов. SPSS позволяет проводить частотный анализ, корреляционный анализ, дисперсионный анализ, кластерный анализ, факторный анализ, а также регрессионный анализ данных. [1] Казалось бы совершенный продукт, лучшее программное обеспечение, позволяющее решить исследовательские задачи, используя статистические методы. Однако, стремительное развитие технологий и постоянно увеличивающийся объем данных, которые необходимо анализировать, стали причиной радикальных изменений, которых мы иногда просто не замечаем. Именно благодаря новым технологиям поисковая машина понимает, какие результаты показывать в ответ на ваш запрос в интернете, а ваша электронная почта очищена от спама с помощью самообучающихся алгоритмов.

«Большие данные» этот термин получает все большее распространение и уже обозначает новую прикладную область, которая формируется на поиске способов быстрого анализа огромных объемов информации. Самым распространенным и перспективным подходом к анализу больших данных считается машинное обучение. Использование машинного обучения, на наш взгляд, способно кардинально изменить систему подходов к анализу данных. Статистические методы исследования предназначены для того, чтобы до-

казать связь между переменными или делать выводы из существующих данных, тогда как машинное обучение способно составить максимально точный прогноз события, явления, процесса и т.п., используя огромные объемы разнородной информации. Еще одно существенное различие между этими двумя подходами состоит в том, что статистический анализ данных использует описательный, исследовательский, прогнозирующий, причинный аспект анализа, а машинное обучение фокусируется только на интеллектуальном моделировании. Другими словами, компьютер сможет найти в массивах данных изначально неизвестные взаимосвязи и закономерности. Статистические модели — это вероятностные модели, любой анализ начинается с предположения. [2] Машинное обучение не привязано к вероятности и может использовать другие подходы к решению проблемы. Таким образом, машинное обучение — это процесс, который обучает искусственный интеллект принимать правильные решения, т.е. компьютер должен не просто использовать составленный алгоритм, а сам обучаться решению поставленной задачи. Машинное обучение состоит из данных, признаков и алгоритмов. Безусловно, данные, это самое важное в машинном обучении — каким бы не был используемый алгоритм, если данные не качественные, то и результат будет соответствующим. Здесь следует отметить, что с помощью машинного обучения мы можем анализировать изображения. Раньше это не мог сделать ни один компьютер. Сейчас компьютер может видеть и читать одновременно.

В настоящее время в совершенно разных сферах деятельности решаются различные прикладные задачи, которые связаны с прогнозированием, регрессией и классификацией. Кроме того, с каждым годом растет количество данных запросов. Так, например, в задачах медицинской диагностики, которые приводит в своем исследовании К.В. Воронцов [3; 9–10], в роли объектов выступают пациенты. Признаки — это результаты обследований, симптомы заболевания и применявшиеся методы лечения. Примеры бинарных признаков — пол, наличие симптомов — головной боли, слабости, тошноты, и т.д. Порядковый признак — тяжесть состояния (удовлетворительное, средней тяжести, тяжелое, крайне тяжелое). Количественные признаки — возраст, пульс, артериальное давление, содержание гемоглобина в крови, доза препарата, и т.д. Признаковое описание пациента является, по сути дела, формализованной историей болезни. Накопив достаточное количество прецедентов, можно решать различные задачи: классифицировать вид заболевания (дифференциальная диагностика); определять наиболее целесообразный способ лечения; предсказывать длительность и исход заболевания; оценивать риск осложнений; находить синдромы — наиболее характерные для данного за-

болевания совокупности симптомов. Этот пример наглядно демонстрирует, что машинное обучение может моментально обобщать и анализировать очень большое количество информации (или прецедентов), что, безусловно, для человека будет нелегко, а возможно и недоступно.

Однако, необходимо отметить, некоторые особенности, которые были определены учеными в процессе работы с демографическими данными (миграционные процессы относятся к демографическим данным), с использованием технологий искусственного интеллекта и машинного обучения. [4] Отмечено, что в демографических данных существуют значительные пропуски или полное отсутствие информации, при этом они являются яркими представителями временных рядов, поэтому основное свойство демографических данных — временная упорядоченность, что необходимо использовать при работе с недостатком или отсутствием данных. Еще одной особенностью демографических данных является последствие их изменения и отсутствие перепадов между ближайшими значениями наблюдаемых данных. Влияние процессов социально-экономического и политического характера в разные исторические периоды на демографические данные также является важной особенностью для анализа с использованием машинного обучения. Таким образом, данные особенности необходимо принимать во внимание для повышения качества и устранения недостатков в использовании демографических данных в машинном обучении.

В связи с вышесказанным, актуальной темой машинного обучения должна стать тема прогнозирования в области демографических явлений, миграционных процессов, что, безусловно, будет являться эффективным и качественным инструментом управления этими процессами. Практическое использование интеллектуального анализа данных при разработке мероприятий социально-экономической и демографической политики позволит сократить ошибки и неточности в применении стратегически неэффективных инструментов, которые в полной мере не могут учитывать все признаковые описания для объекта миграционный процесс. Например, у нас есть данные о 100 тыс. мигрантов в Москве, причем, нам известна их возрастная, половая структура, национальный состав, причины нахождения на территории России, частота переездов с места на место, места выхода и входа мигрантов и т.п. Кроме того, у нас есть исторические данные о количестве мигрантов. Нашей задачей будет построить модель, которая на основе данных признаков будет предсказывать количество мигрантов в будущем. Это пример решения задачи с «учителем», где есть данные по населению (возраст, пол, национальность и т.п.) и отклики

(количество мигрантов). Мы сможем предвидеть рост или снижение миграционного потока заранее, что позволит принимать взвешенные решения для будущих событий, другими словами, будем способны управлять рисками, снижая негативный эффект и используя открывающиеся возможности. Это задача машинного обучения решается с помощью «учителя» и называется регрессией. Можно привести и другие примеры: на основании показателей уровня образования мигрантов, определить какие качественные характеристики будут преобладать в том или ином миграционном потоке.

В целом, задачи машинного обучения можно разделить на два вида — это задачи обучения с «учителем» и задачи обучения «без учителя» соответственно. [5] Задачи, которые решаем с «учителем» (см. пример выше) — это когда у нас есть данные, на основании которых мы будем делать прогноз и есть некоторые гипотезы. Задачи, решаемые «без учителя», используют только данные, свойства которых мы должны узнать. Необходимо отметить, что самое сложное в данной области — это собрать всю информацию в единое целое. С каждым днем объем информации растет и самый легкий способ раскрыть тему исследования полностью — это выбрать узкую область специализации, чтобы не потеряться в информационном потоке данных. Именно здесь на помощь приходит машинное обучение, которое способно обработать литературу в поисках нужной информации, а также найти связь между областями и внутри них.

Необходимо обратить внимание на то, что существуют важные особенности при внедрении машинного обучения в ту или иную отрасль, которые необходимо учитывать при работе с данными в процессе подготовки и анализа. [6] Проблема заключается в том, что машинное обучение работает с огромными базами данных, но, при этом, эти данные представлены не в полном объеме или они отсутствуют в исследуемых областях. Кроме того, сама постановка задачи исследования в машинном обучении может быть некорректной, т.к. система ограничена данными, по которым ее обучали. Поэтому проблема корректной формулировки задач является важной. Кроме того, важным является органи-

зация процесса сбора данных для анализа, их полнота и применимость для решения поставленной задачи.

Заключение

Использование машинного обучения становится все более и более популярным методом аналитического анализа, прогнозирования и мониторинга, что связано с ростом объема данных и необходимости анализа больших массивов информации. Формирование необходимой инфраструктуры для обеспечения анализа данных с применением методов машинного обучения тоже не стоит на месте — многие крупные компании строят гипермасштабные ЦОДы (Центр Обработки Данных), разрабатывают собственные системы искусственного интеллекта, увеличивают финансирование исследований в области искусственного интеллекта и машинного обучения. Так, в докладе консалтинговой компании Alorithmia за 2021 г. отмечается, что 83% организаций-респондентов значительно наращивают бюджеты на искусственный интеллект и машинное обучение. [7]

Таким образом, использование искусственного интеллекта в управлении миграционными процессами выведет эту отрасль управления совершенно на технологически новый уровень, что позволит решать комплекс миграционных проблем России более эффективно и результативно, корректно выделяя приоритеты и формулируя текущие задачи, учитывая национальные, прежде всего, геополитические интересы страны в целом. Машинное обучение уже сегодня может в режиме реального времени предлагать информацию, что приведет к более быстрым и эффективным решениям. В связи с этим, необходима стратегическая основа для реализации внедрения машинного обучения в управление как в целом демографическими процессами, так и миграционными процессами, в частности. Разработка стратегического плана внедрения системы машинного обучения в управление демографическими и миграционными процессами — способно вывести на новый, высокий уровень в целом систему прогнозирования, мониторинга, анализа данных в данной области исследования.

ЛИТЕРАТУРА

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.
2. Бенедикт Эванс: «Мы не до конца понимаем возможности машинного обучения». Интернет-ресурс: <https://rb.ru/story/how-to-think-about-ml/>
3. Воронцов К.В. Математические методы обучения по прецедентам. МФТИ. 2006. Интернет-ресурс: www.MachineLearning.ru
4. Соловьев А.И., Соловьев С.А. О подходах к анализу демографических данных с использованием машинного обучения/Modern Information Technologies and IT-Educatio. Vol. 14, no 4. 2018. Интернет-ресурс: sitito.cs.msu.ru
5. Что такое машинное обучение простыми словами. Интернет-ресурс: <https://www.reg.ru/support/vps-server/oblastnie-serveri-vps/usluga-oblastnyye-servery/chto-takoe-mashinnoe-obuchenie-prostymi-slovami>

6. Государство. Бизнес. Технологии. Машинное обучение. Machine Learning. 04/04/2022. Интернет-ресурс: [https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_\(Machine_Learning\)](https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_(Machine_Learning))
7. Как использовать машинное обучение при управлении структурой ЦОД? Centers.ru Интернет-ресурс: <https://alldc.ru/news/6437.html>
8. Машинное обучение: как оно применяется в жизни. Интернет-ресурс: <https://rb.ru/opinion/mashinnoe-obuchenie/>
9. Машинное обучение. Интернет –ресурс: http://www.machinelearning.ru/wiki/index.php?title=Machine_Learning
10. Mitchell T. Machine Learning. — McGraw-Hill Science/Engineering/Math, 1997.

© Дмитриева Юлия Викторовна (julia9770@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»



Российская академия наук