



Технологии сбора данных в сети Интернет и их пространственного анализа при принятии стратегических решений в туристической отрасли

И.А. Лакман¹
В.М. Тимирьянова¹

¹ Башкирский государственный университет (Уфа, Россия)

Аннотация

В работе предлагается взгляд на новые возможности, которые открываются в управлении в условиях развития информационных технологий. Целью исследования является выделение преимуществ применения технологий сбора больших данных и пространственной статистики для принятия стратегических решений в туристической отрасли. В работе на примере горнолыжного курорта «Абзаково» показаны технологии сбора данных в интернете, проведена оценка пространственной зависимости данных и сформулированы выводы, позволяющие в последующем принимать стратегические решения как предприятиям, расположенным на данной территории, так и органам государственного управления, заинтересованным в развитии этой туристической зоны. Так, анализ географии участников группы «Абзаково» позволил выделить статистически значимую пространственную зависимость их возраста, проявляющуюся в соседстве более взрослых участников группы вблизи курорта. Результаты анализа рейтинга объектов туристической инфраструктуры в районе горнолыжного курорта «Абзаково» выделили пространственно связанные зоны низких и высоких значений рейтингов.

Ключевые слова: технологии сбора больших данных, пространственная статистика, туристическая отрасль.

Для цитирования:

Лакман И.А., Тимирьянова В.М. (2021). Технологии сбора данных в сети Интернет и их пространственного анализа при принятии стратегических решений в туристической отрасли. *Стратегические решения и риск-менеджмент*, 2(12): 170–177. DOI: 10.17747/2618-947X-2021-2-170-177.

Благодарность

Статья подготовлена в соответствии с государственным заданием Министерства науки и высшего образования Российской Федерации (код научной темы FZWU-2020-0027).

Strategic decisions in the tourism industry based on Internet data collection technologies and their spatial analysis

I.A. Lakman¹
V.M. Timiryanova¹

¹ Bashkir State University (Ufa, Russia)

Abstract

The paper offers a look the new opportunities that open up in management in the context of the development of information technology. The aim of the study is to highlight the benefits of using technologies for collecting big data and spatial statistics for making strategic decisions in the tourism industry. In the work, using the example of the ski resort “Abzakovo”, the technologies for collecting data on the Internet are shown, the spatial dependence of the data is assessed and conclusions are drawn that allow subsequent strategic decisions as to enterprises located in this territory. So are the government bodies interested in the development of this tourist zone. Thus, the analysis of the geography of the members of the “Abzakovo” group made it possible to single out a statistically significant spatial dependence of their age, which manifests itself in the neighborhood of the older members of the group near the resort. The results of the analysis of the rating of tourist infrastructure facilities in the area of the ski resort “Abzakovo” identified spatially related zones of low and high ratings.

Keywords: big data collection technologies, spatial statistics, tourism industry.

For citation:

Lakman I.A., Timiryanova V.M. (2021). Strategic decisions in the tourism industry based on Internet data collection technologies and their spatial analysis. *Strategic Decisions and Risk Management*, 12(2), 170-177. DOI: 10.17747/2618-947X-2021-2-170-177. (In Russ.)

Acknowledgements

This research was funded by the Ministry of Science and Higher Education of the Russian Federation (scientific code FZWU-2020-0027).

1. Введение

«Поток огромных объемов информации различных структур, содержащих знания колоссальной ценности, растет с большой скоростью» [Шведов, 2017]. В условиях конкурентной борьбы выигрывает тот, кто обладает наиболее полной информацией, в связи с чем идет непрекращающийся процесс поиска инструментов сбора и обработки формирующихся данных. Это касается всех отраслей и сфер деятельности, в том числе туристической отрасли.

Туристическая отрасль – одна из наиболее активно развивающихся отраслей мировой экономики [Бутенко, 2015]. Ее развитие оказывает влияние не только на объекты туристской инфраструктуры, но также косвенно сказывается на транспортной сфере, строительстве, легкой и пищевой промышленности. Соответственно, в развитии туризма заинтересованы как владельцы объектов туристской инфраструктуры, так и органы власти, поскольку она положительно отражается на притоке денежных средств, развитии территории и ее инвестиционной привлекательности.

Важность этой отрасли определяет интерес к исследованиям факторов ее развития. В условиях развития больших данных все больший исследовательский интерес вызывают открытые данные. В основном это отзывы и рейтинги, оставляемые туристами на различных сайтах в сети Интернет. Спектр источников данных, равно как и тематика исследований, использующих эту информацию, довольно широк [Alaei et al., 2017; Булгаков, 2018; Пигарева, Шевелев, 2018; Li et al., 2018; Ćurlin et al., 2019; Домбровская, 2020; Chang et al., 2020]. Собираемые данные чаще всего используются для проведения семантического анализа текстов, в том числе в целях определения ключевых моментов, наиболее важных для клиентов [Alaei et al., 2017; Ćurlin et al., 2019], или «тональности постов» [Булгаков, 2018]. В ряде случаев результаты визуализируются на картах [Adhinugroho et al., 2020; Chang et al., 2020]. Однако исследований, предусматривающих еще и пространственный анализ этих данных, очень мало, несмотря на то что современное развитие технологий предполагает их географическое структурирование. Более того, для России в целом свойственно их отсутствие применительно к каким-либо исследованиям в туристической отрасли. Так, С.А. Тархов отмечает, что исследования в России сводятся «даже не к анализу размещения туризма, а к изучению влияния разных факторов на особенности туризма в том или ином регионе или стране» [Тархов, 2019], то есть не предусматривают применение инструментов пространственной статистики.

В то же время в зарубежной литературе рассматривается довольно широкий спектр вопросов с применением этих инструментов и моделей. Так, У. Ли, Л. Пеннингтон-Грей и Ж. Ким исследовали с помощью географически взвешенной регрессии влияние близости к пляжу, центру города и аэропорту на безопасность пищевых продуктов в 308 ресторанах при отелях [Lee et al., 2019]; М. Салас-Олмедо и соавторы изучали пространственную автокорреляцию туристической активности на основе 234 159 твитов, сопоставляя их с данными, получаемыми из открытых источников данных Panoramio и Foursquare [Salas-Olmedo et al., 2018]; С. Жанг, Г. Жанг и Х. Юю оценивали пространственные зако-

номерности развития туризма в бассейне Желтой реки с точки зрения объема туристов (внутренних и въездных) и доходов от туризма (внутреннего и въездного) на уровне провинций и префектур [Zhang et al., 2020]. В целом список таких работ может быть достаточно широк, как широк и спектр направлений использования получаемых результатов.

Если говорить о предприятиях, то такая информация позволяет им принимать решения относительно направлений дальнейшего развития своей деятельности, начиная от базовых маркетинговых планов, учитывающих географическую организацию клиентов, и заканчивая формированием направлений расширения деятельности, принимающих во внимание конкурентное окружение и иерархию туристических мест.

Если говорить о государстве, то здесь спектр направлений работы также широк: обеспечение безопасности и разработка рекомендаций по организации деятельности вблизи объектов притяжения населения (природных, культурных и др.), создание инфраструктуры, привлечение инвестиций на развитие территорий и т. д. Связь результатов с пространственной локацией объектов, их формирующих, очевидна, и ее конкретная количественная оценка в конкретной ситуации позволяет принимать обоснованные решения.

Целью исследования является выделение преимуществ применения технологий сбора больших данных и пространственной статистики для принятия стратегических решений в туристической отрасли. Для достижения поставленной цели были решены следующие задачи: обзор инструментов сбора и пространственного анализа данных, сбор и оценка данных сайтов VK.ru и Google.com на примере горнолыжного курорта «Абзаково», обобщение полученных результатов.

2. Методология исследования и данные

Применяемый в рамках настоящего исследования инструментарий можно разделить на две группы: к первой относятся инструменты сбора и структурирования данных, ко второй – инструменты анализа.

Сбор первичной информации часто занимает львиную долю всего времени, потраченного на исследование. Когда речь идет о микроуровневом анализе (например, о работе отдельных предприятий, об отношениях с клиентами), сбор информации из внешних источников предполагает проведение полевых исследований. Особенно трудоемким является проведение опроса, необходимого для понимания различных поведенческих паттернов. В то же время в условиях развития информационных технологий, когда население само оставляет данные о себе в сети Интернет, а также так или иначе осуществляет оценку деятельности предприятий, на ряд вопросов может быть получен ответ и без этого трудоемкого метода сбора данных. Спектр открытых данных довольно широк, поэтому «с развитием информационных технологий стали создаваться более объемные базы данных на интернет-порталах и сайтах, а также появились инструменты (так называемые парсеры) формирования на основе интернет-данных выборок с заданными характеристиками» [Трифонов и др., 2020]. Именно такой инструмент применялся в настоящей работе для сбора данных.

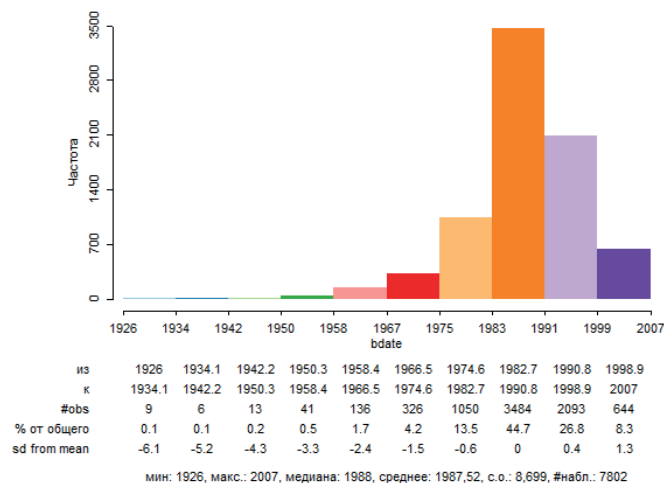
Парсинг определяется как «автоматизированный сбор неструктурированной информации, ее преобразование и выдача в структурированном виде» [Селиверстов и др., 2019]. «Технологии парсинга позволяют извлекать большой объем фактических данных, который обрабатывается инструментами дата-майнинга» [Днепровская, 2020]. Этот инструмент активно начал применяться в России последние пять лет, в то время как в ряде зарубежных стран результаты его использования нашли отражение в научных работах еще двадцать лет назад [Piccoli, 2001]. В целом он имеет довольно широкий круг применения. В частности, он позволил собрать данные для моделирования рыночной стоимости автомобилей на основе объявлений о продаже автомобилей [Трифонов и др., 2020], оценки состояния транспортных магистралей Северо-Западного федерального округа России [Селиверстов и др., 2019], анализа запросов для привлечения абитуриентов [Плотников, 2016] и данные электронных государственных закупок [Цыганова и др., 2020]. В туристическом секторе такой метод сбора применялся для анализа отзывов посетителей отелей на сайте Tripadvisor [Chang et al., 2020], сообщений туристов в Twitter [Alaei et al., 2017; Ćurlin et al., 2019], ВКонтакте [Домбровская, 2020], Facebook [Булгаков, 2018], Instagram [Пигарева, Шевелев, 2018], Yahoo [Li et al., 2018] и т.д.

Для парсинга, как правило, применяют специализированные программы, а также различные библиотеки под пакеты Python и R. При этом необходимо учитывать, что крупные сайты, владеющие пользующимися спросом большими данными, предоставляют доступ к ним через API¹. В этих случаях, как правило, под каждый из них формируются собственные специализированные библиотеки. В частности, в рамках данного исследования применялись библиотеки vkR, googleway, а также требовалось подключение соответствующих API, так как и социальная сеть ВКонтакте, и поисковая система Google имеют четко регламентированный набор готовых классов, функций или структур для работы с имеющимися данными. Следует отметить, что сбор данных с сайтов в целом не противоречит законодательству, в случае если владельцем сайта в лицензии robots.txt не прописан запрет на него, как, например, в социальных сетях Facebook и Instagram, а также в поисковой сети Yandex. И если в Yandex возможность парсинга допускается в случае получения лицензии, в том числе коммерческой, то Facebook на текущий момент жестко придерживается политики ограждения личных данных пользователей от какой-либо обработки. Социальная сеть ВКонтакте также имеет ограничения на сбор, которые прописаны в правилах платформы (п. 2), в ряде случаев требующие получения согласия от администрации сайта.

В рамках данного исследования проводился анализ горнолыжного курорта «Абзаково», расположенного в создаваемой на территории Республики Башкортостан туристической зоны «Урал». С целью выявления пространственных особенностей, связанных с этим курортом, была проанализирована группа в социальной сети ВКонтакте «Абзаково»², а также рейтинги объектов общественного питания и размещения горнолыжного центра в картах Google. Сбор данных осуществлялся в апреле 2021 года.

Рис. 1. Гистограмма возраста участников группы «Абзаково», включенных в анализ

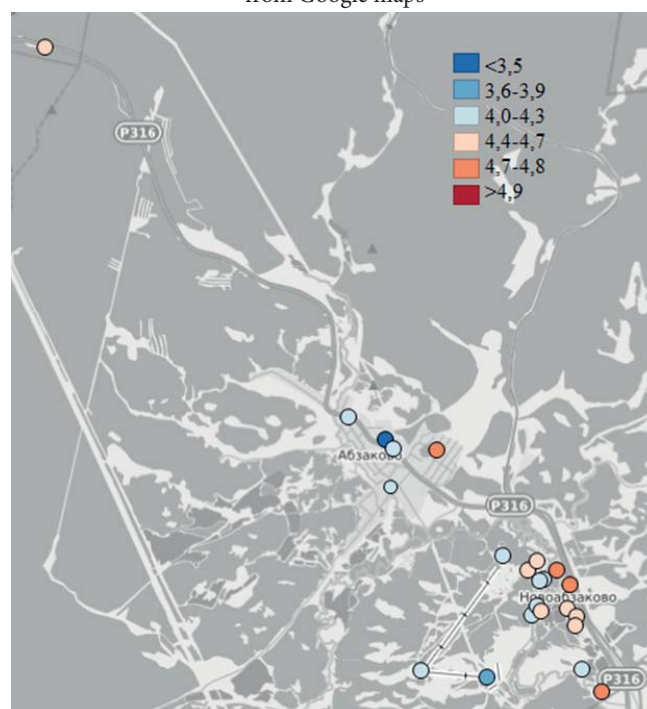
Fig. 1. Histogram of the age of the Abzakovo group participants included in the analysis



Группа в социальной сети ВКонтакте «Абзаково» на момент сбора данных насчитывала 36 755 чел., среди которых 54,6% указали свою принадлежность женскому полу, а 19 чел. не указали или скрыли свой пол. 54,3% всех участников отметили город своего проживания, 21,4% – год своего рождения. Отсутствие данных по ряду участников определяется применяемыми ими настройками приватности, изменяющимися в диапазоне от «видят все пользователи»

Рис. 2. Рейтинги мест размещения и объектов питания, представленные на картах Google

Fig. 2. Ratings of accommodation and food facilities from Google maps



¹ Совокупность программных средств для работы с данными посредством HTTP-запросов.

² Сбор данных осуществлялся с согласия группы.

до «только я». География участников группы очень широкая, выходящая за границы России. Есть участники, отметившие свое расположение в Абу-Даби, Леоне, Лондоне и т.д. Анализ возраста участников группы проводился только по 7802 наблюдениям в России (рис. 1). Следует согласиться с тем, что достоверность указанных данных сложно проверить, и это должно учитываться в анализе [Salas-Olmedo et al., 2018; Домбровская, 2020]. Так, возвращаясь к возрасту участников группы, мы не можем исключать того, что часть из них завысила свой возраст в связи с ограничениями по возрасту при регистрации в социальной сети ВКонтакте, часть участников могла занижить свой возраст или указать неверную дату по каким-либо причинам.

С карт Google были взяты данные о 23 объектах туристической инфраструктуры, среди которых семь кафе, один ресторан, одна столовая, четырнадцать отелей и домов отдыха, расположенных в непосредственной близости к горнолыжному центру в деревнях Абзаково и Новоабзаково, а также на трассе Р316, ведущей в г. Белорецк (рис. 2).

Анализ данных проводился наиболее известным методом пространственной статистики и включал оценку локальных индексов Морана (LISA – Local Index Spatial Autocorrelation) [Moran, 1948; Anselin et al., 2002]:

$$L_{Imi} = N \frac{(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{\sum_j (x_j - \bar{x})^2}, \quad (1)$$

где \bar{x} – среднее значение показателя, w_{ij} – пространственная матрица, N – число исследуемых единиц.

Для анализа пространственного распределения участников группы социальной сети ВКонтакте и для анализа рейтингов объектов инфраструктуры применялись разные матрицы (w_{ij}), учитывающие соседство i -го и j -го объектов наблюдения. В случае анализа рейтингов объектов инфраструктуры применялась простая матрица соседства первого порядка, построенная по критерию ферзя. В случае анализа участников социальной группы ВКонтакте использовалась матрица расстояний с радиусом 0,4°. Применение других типов матриц во втором случае было невозможно в силу особенностей данных. Так, наличие городов с единичными участниками делало нецелесообразным использование матрицы типа k – ближайших соседей, так как в этом случае в качестве соседей могли учитываться наблюдения из других городов, находящихся на достаточно большом удалении. В городах со значительным количеством участников матрица соседства, равно как и матрица типа k – ближайших соседей, не позволяла охватить всю совокупность связей, возможно, существующих между наблюдениями внутри города.

При интерпретации полученных результатов расчета локального индекса Морана важны и знак, и полученное значение. При $L_{Imi} < 0$ имеется отрицательная пространственная автокорреляция, то есть объект наблюдения по этому значению существенно отличается от соседних (outlier). При $L_{Imi} > 0$ пространственная автокорреляция положительная, то есть это наблюдение по данному значению подобно соседним (cluster). На основе значений локального индекса Морана и их статистической значимости могут быть выделены ядра кластеров.

1. Высокий – высокий – наблюдения имеют относительно высокие собственные значения анализируемого показателя,

окружены наблюдениями также с относительно высокими значениями анализируемого показателя. Пространственная автокорреляция положительная.

2. Низкий – низкий – наблюдения имеют относительно низкие собственные значения анализируемого показателя, окружены наблюдениями также с относительно низкими значениями анализируемого показателя. Пространственная автокорреляция положительная.

3. Высокий – низкий – наблюдения имеют относительно высокие собственные значения анализируемого показателя, окружены наблюдениями с относительно низкими значениями анализируемого показателя. Пространственная автокорреляция отрицательная.

4. Низкий – высокий – наблюдения имеют относительно низкие собственные значения анализируемого показателя, окружены наблюдениями с относительно высокими значениями анализируемого показателя. Пространственная автокорреляция отрицательная.

3. Результаты пространственного анализа участников социальной группы «Абзаково» по возрасту

География участников группы отражена на рис. 3. Следует отметить, что более 70% наблюдений – участников группы проживают в районе Южного Урала, то есть в непосредственной близости от горнолыжного курорта. Больше всего участников группы проживают в г. Магнитогорске (1271 наблюдение, или 16,3%), следующим городом является Уфа (952 чел., или 12,2% анализируемых участников).

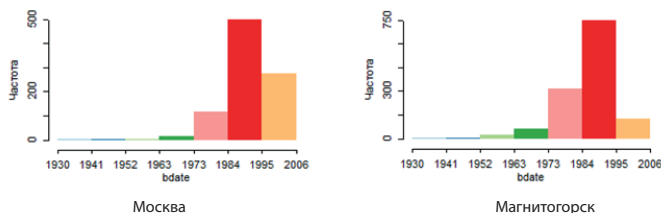
Оценка локального индекса Морана показала, что присутствует определенная (значимая на уровне $p < 0,05$) пространственная связь наблюдений, позволяющая выделить отдельные города (рис. 3). Красным выделены территории, на которых наблюдения с более высокими значениями показателя соседствуют с наблюдениями, имеющими также относительно более высокие значения показателя. Применительно к анализируемому показателю «год рождения» можно говорить о том, что в зонах, выделенных красным цветом, преобладают молодые участники группы, а в зонах, выделенных синим, люди более старшего возраста соседствуют с людьми, средний возраст которых также относительно высок.

Рис. 3. Локальные индексы Морана для участников группы «Абзаково» в России

Fig. 3. Local Moran's I for Abzakovo group members located in Russia



Рис. 4. Гистограммы участников группы «Абзаково» в Москве и Магнитогорске
Fig. 4. Histograms of the Abzakovo group members located in Moscow and Magnitogorsk



В целом эти расчеты можно подтвердить, проанализировав возрастную структуру участников в разрезе каждого города. Например, при равном диапазоне годов рождения участников группы (1930–2006 годы рождения) в Москве для 914 наблюдений среднее значение составляет 1991 год, медиана – 1992 год, а в Магнитогорске для 1271 наблюдений среднее значение – 1986 год, медиана – 1987 год (рис. 4).

Если обратить внимание на локализацию синих точек, то можно отметить их концентрацию в районе южного Урала, в непосредственной близости от горнолыжного курорта «Абзаково». Это, в свою очередь, указывает на то, что близость способствует охвату более взрослых туристов. На юге и юго-западе страны преобладают красные точки, указывающие на то, что на этих территориях наблюдения – участники группы имеют более высокие значения года рождения (то есть относительно всей совокупности наблюдений более молодые люди), соседствуют с аналогично относительно более молодыми участниками группы. Косвенно это говорит о том, что в этих городах, возможно, существуют группы с общими ценностями, включающими интерес к горным лыжам. Объективно люди, общаясь, передают информацию о курорте «из уст в уста», способствуя расширению круга туристов в этой возрастной прослойке. Такое понимание особенностей распределения возможных клиентов горнолыжного курорта определяет стратегии его продвижения на различных территориях.

4. Результаты пространственного анализа рейтинга объектов туристической инфраструктуры в районе горнолыжного курорта «Абзаково»

География объектов инфраструктуры и рейтинги, которые они получили на картах Google, представлены на рис. 5. Видно, что объекты, находящиеся непосредственно на лыжной трассе и в Абзаково, оцениваются туристами хуже. Однако пространственный анализ позволяет оценивать инфраструктуру под ракурсом окружения. Расчет локальных индексов Морана позволил выделить четыре статистически значимые точки. В нижней части рис. 5 четко выделяется синяя точка, указывающая на то, что этот объект имеет низкие значения рейтинга и окружен объектами, также имеющими относительно низкие значения показателя.

В Новоабзаково выделяются три красные точки, для которых характерны относительно высокие значения рейтингов,

и они окружены объектами, также имеющими относительно более высокие оценки. Здесь акцент делается не на рейтинге объекта, а на соотношении его рейтинга и рейтинга окружения, а также концентрация таких объектов в пространстве.

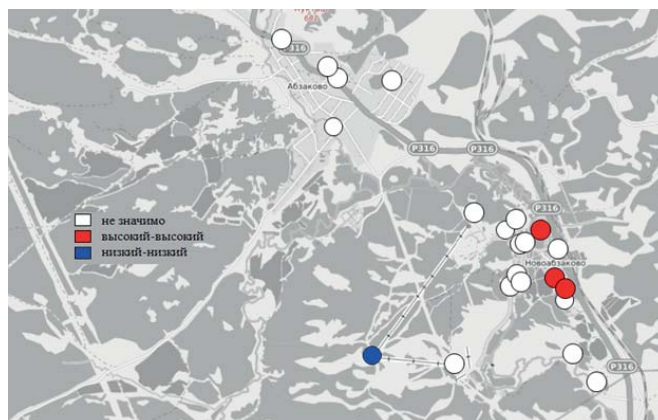
5. Заключение

В настоящее время с развитием информационных технологий спектр данных, учитываемых при принятии стратегических решений, может быть значительно расширен. Появление открытых источников неограниченной информации, которая к тому же географически структурирована, открывает широкий диапазон возможностей. В рамках настоящей работы был представлен небольшой пример таких возможностей, включая применение новых технологий сбора данных, а также их последующего анализа методами пространственной статистики.

Результаты анализа географии участников группы «Абзаково» показали не только их преимущественное проживание в районе Южного Урала, но и позволили выделить статистически значимую пространственную зависимость их возраста, проявляющуюся в соседстве более взрослых участников группы вблизи курорта. Результаты анализа рейтинга объектов туристической инфраструктуры в районе горнолыжного курорта «Абзаково» показали, что более рейтинговые объекты группируются в восточных районах д. Новоабзаково, а в зоне лыжни отмечаются пространственно связанные низкие значения. При анализе объектов инфраструктуры не было выявлено статистически значимых обратных зависимостей. В то же время при анализе именно они вызывают наибольший интерес, акцентируя внимание на объектах, имеющих низкие значения показателей в сравнении с окружением, имеющим высокие значения, и наоборот.

Безусловно, здесь требуется дальнейший анализ причин сложившейся ситуации. Однако в рамках настоящей работы такой задачи не ставилось. Ключевая цель состояла в выявлении возможностей, которые открывают новые инструменты, развивающиеся с появлением больших данных. Необходимо понимать, что в данном случае анализировалось небольшое количество объектов наблюдения. При переходе

Рис. 5. Локальные индексы Морана для объектов инфраструктуры горнолыжного курорта «Абзаково»
Fig. 5. Local Moran's I for infrastructure facilities of the Abzakovo ski area



на большее количество наблюдений крайне важной становится скорость определения ключевых точек в пространстве, требующих различных подходов к анализу, получить которую позволяют новые инструменты сбора и анализа данных. Кроме того, увеличение числа наблюдений расширяет возможности включения в модели большего количества признаков и способствует увеличению степеней свободы в эконометрических моделях, определяющих достоверность статистических оценок. Но необходимо отметить и ряд недостатков, в частности присущих парсингу как способу сбора данных для анализа. В первую очередь это невозможность

проверки достоверности информации, указанной участниками в личных профилях социальных сетей, а также субъективность и искусственная накрутка выставляемых оценок на интерактивных картах. Вторым недостатком является частичное отсутствие сведений по ряду полей, необязательных к заполнению, а также их сокрытие настройками приватности. Также необходимо учитывать, что активность в интернете свойственна не всем людям, соответственно, анализируя контент социальных сетей и рейтинги, нужно понимать, что это мнения людей определенного склада характера, социального типа и т. д.

Литература

1. Булгаков Д.С. (2018). Использование пакета NLTK и модели Вадера для оценки туристической привлекательности страны на примере Российской Федерации. *Web of Scholar*, 2(20): 4–7.
2. Бутенко И.В. (2015). Тенденции развития туристической отрасли в России. *Статистика и экономика*, 5: 90–93.
3. Днепровская Н.В. (2020). Метод исследования компетенций субъектов цифровой экономики. *Открытое образование*, 24(1): 4–12.
4. Домбровская В.Е. (2020). Аспекты туристского брендинга территории на примере Тверской области. *Вестник ТвГУ. Сер. «География и геоэкология»*, 3(31): 94–101.
5. Пигарева Е.Ю., Шевелев И.А. (2018). Данные пользователей Instagram как основа для маркетингового анализа в событийном туризме. *Туризм и рекреация: инновации и ГИС-технологии: Материалы X Международной научно-практической конференции* (г. Астрахань, 26–27 мая 2018 г.) / сост. И.С. Шарова, М.М. Иолин. Астрахань. Новая линия, 6–12.
6. Плотников А.В. (2016). Анализ геодезических запросов в поисковой системе Яндекс для привлечения абитуриентов в вузы. *Практический маркетинг*, 1(239): 12–16.
7. Селиверстов Я.А., Никитин К.В., Шаталова Н.В., Киселев А.А. (2019). Оценка состояния транспортных магистралей Северо-Западного федерального округа с использованием анализа тональности отзывов пользователей сети Интернет. *Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление*, 12(3): 7–24.
8. Тархов С.А. (2019). Пространственный подход в географии туризма. *Географический вестник*, 4(51): 172–178.
9. Трифонов Н.Ю., Ливинская В.А., Коржуков В.В. (2020). Регрессионная модель оценки автомобилей на основе парсинга интернет-данных. *Системный анализ и прикладная информатика*, 2: 4–9.
10. Цыганова М.С., Буреш С.В., Чернушенко Д.А. (2020). Разработка инструментария анализа данных о государственных закупках по федеральным законам № 44-ФЗ и № 223-ФЗ (на базе аналитической платформы Loginom). *Вестник Череповецкого государственного университета*, 6(99): 59–72.
11. Шведов Д.И. (2017). Парсинг и матчинг данных для оперативного анализа динамических ситуаций. *Горный информационно-аналитический бюллетень*, 6: 200–204.
12. Adhinugroho Y., Putra A.P., Luqman M., Ermawan G.Y., Takdir, Mariyah S., Pramana S. (2020). Development of online travel Web scraping for tourism statistics in Indonesia. *Information Research*, 25(4): 885. URL: <https://doi.org/10.47989/irpaper885>.
13. Alaei A.R., Becken S., Stantic B. (2017). Sentiment analysis in tourism: Capitalizing on Big data. *Journal of Travel Research*, 004728751774775.
14. Anselin L., Syabri I., Smirnov O. (2002). Visualizing multivariate spatial correlation with dynamically linked Windows. *Computing Science and Statistics*. URL: <https://www.semanticscholar.org/paper/Visualizing-Multivariate-Spatial-Correlation-with-Mirnov/4e34bd70317377971ba8df7259288b972ad6a239>.
15. Chang Y.-C., Ku C.-H., Chen C.-H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management*, 80, 104–129.
16. Ćurlin T., Jaković B., Miloloža I. (2019). Twitter usage in Tourism: Literature review. *Business Systems Research*, 10(1): 102–119.
17. Lee Y., Pennington-Gray L., Kim J. (2019). Does location matter? Exploring the spatial patterns of food safety in a tourism destination. *Tourism Management*, 71: 18–33.
18. Li D., Zhou X., Wang M. (2018). Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities*, 74: 249–258.
19. Moran P. The interpretation of statistical maps (1948). *Journal of the Royal Statistical Society*, ser. B, 10: 243–251.
20. Piccoli G. (2001). Web-site marketing for the tourism industry: Another view. *The Cornell Hotel and Restaurant Administration Quarterly*, 42(6): 63–65.
21. Salas-Olmedo M.H., Moya-Gómez B., García-Palomares J.C., Gutiérrez J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66: 13–25.
22. Zhang S., Zhang G., Ju H. (2020). The spatial pattern and influencing factors of tourism development in the Yellow River Basin of China. *PLoS ONE*, 15(11): e0242029. URL: <https://doi.org/10.1371/journal.pone.0242029>.

References

1. Bulgakov D.S. (2018). Using the NLTK package and the Vader model to assess the tourist attractiveness of a country on the example of the Russian Federation. *Web of Scholar*, 2(20): 4-7. (In Russ.)
2. Butenko I.V. (2015). Trends in the development of the tourism industry in Russia. *Statistics and Economics*, 5: 90-93. (In Russ.)
3. Dneprovskaya N.V. (2020). The method to study the competencies of the subjects of the digital economy. *Open Education*, 24(1): 4-12. (In Russ.)
4. Dombrovskaya V.E. (2020). Aspects of tourist branding of the territory on the example of the Tver region. *Herald of Tver State University. Ser: Geography and Geoecology*, 3(31): 94-101. (In Russ.)
5. Pigareva E.Yu., Shevelev I.A. (2018). Instagram user data as a basis for marketing analysis in event tourism. In: Sharova I.S., Iolin M.M. (comp.). *Tourism and recreation: Innovations and GIS technologies: Materials of the X International scientific and practical conference* (Astrakhan, May 26-27, 2018). Astrakhan, Novaya liniya, 6-12. (In Russ.)
6. Plotnikov A.V. (2016). Analysis of geo-dependent requests in Yandex search engine to attract enrollees. *Practical Marketing*, 1(239): 12-16. (In Russ.)
7. Seliverstov Ya.A., Nikitin K.V., Shatalova N.V., Kiselev A.A. (2019). Road pavement assessment of the North-West Federal District using sentiment analysis of the Internet user reviews. *St. Petersburg Polytechnical University Journal. Computing, Telecommunication and Control*, 12(3): 7-24. (In Russ.)
8. Tarkhov S.A. (2019). Spatial approach to tourism geography. *Geographical Bulletin*, 4(51): 172-178. (In Russ.)
9. Trifonov N.Yu., Livinskaya V.A., Korzhukov V.V. (2020). Regression model for car valuation based on internet data parsing. *System analysis and applied information science*, 2: 4-9. (In Russ.)
10. Tsyganova M.S., Buresh S.V., Chernushenko D.A. (2020). Data analysis instrumentation development for public procurements under 44 and 223 federal laws (based on loginom analytical platform). *Cherepovets State University Bulletin*, 6(99): 59-72. (In Russ.)
11. Shvedov D.I. (2017). Parsing and matching of data for on-line analysis of dynamic situations. *Mining Informational and Analytical Bulletin*, 6: 200-204. (In Russ.)
12. Adhinugroho Y., Putra A.P., Luqman M., Ermawan G.Y., Takdir, Mariyah S., Pramana S. (2020). Development of online travel Web scraping for tourism statistics in Indonesia. *Information Research*, 25(4): 885. URL: <https://doi.org/10.47989/irpaper885>.
13. Alaei A.R., Becken S., Stantic B. (2017). Sentiment analysis in tourism: Capitalizing on Big data. *Journal of Travel Research*, 004728751774775.
14. Anselin L., Syabri I., Smirnov O. (2002). Visualizing multivariate spatial correlation with dynamically linked Windows. *Computing Science and Statistics*. URL: <https://www.semanticscholar.org/paper/Visualizing-Multivariate-Spatial-Correlation-with-Mirnov/4e34bd70317377971ba8df7259288b972ad6a239>.
15. Chang Y.-C., Ku C.-H., Chen C.-H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management*, 80: 104-129.
16. Ćurlin T., Jaković B., Miloloža I. (2019). Twitter usage in Tourism: Literature review. *Business Systems Research*, 10(1): 102-119.
17. Lee Y., Pennington-Gray L., Kim J. (2019). Does location matter? Exploring the spatial patterns of food safety in a tourism destination. *Tourism Management*, 71: 18-33.
18. Li D., Zhou X., Wang M. (2018). Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities*, 74: 249-258.
19. Moran P. The interpretation of statistical maps (1948). *Journal of the Royal Statistical Society*, ser. B, 10: 243-251.
20. Piccoli G. (2001). Web-site marketing for the tourism industry: Another view. *The Cornell Hotel and Restaurant Administration Quarterly*, 42(6): 63-65.
21. Salas-Olmedo M.H., Moya-Gómez B., García-Palomares J.C., Gutiérrez J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66: 13-25.
22. Zhang S., Zhang G., Ju H. (2020). The spatial pattern and influencing factors of tourism development in the Yellow River Basin of China. *PLoS ONE*, 15(11): e0242029. URL: <https://doi.org/10.1371/journal.pone.0242029>.

Информация об авторах

Ирина Александровна Лакман

Кандидат технических наук, доцент, заведующий лабораторией исследования социально-экономических проблем регионов, Башкирский государственный университет (Уфа, Россия). ORCID: <https://orcid.org/0000-0001-9876-9202>, ResearcherID: K-6878-2017, Scopus Author ID: 57192164952.

Область научных интересов: методы статистического (эконометрического) моделирования, предиктивная аналитика, социально-экономические и биологические системы.

Lackmania@mail.ru

Венера Маратовна Тимирьянова

Кандидат экономических наук, доцент, старший научный сотрудник, заместитель заведующего лабораторией исследования социально-экономических проблем регионов, Башкирский государственный университет (Уфа, Россия). ORCID: <https://orcid.org/0000-0002-1004-0722>, ResearcherID: N-3449-2015, Scopus Author ID: 57194428883.

Область научных интересов: методы статистического (эконометрического) моделирования, потребительское поведение, пространственная экономика.

79174073127@mail.ru

About the authors

Irina A. Lakman

Candidate of technical sciences, associate professor, head of the Laboratory for the Study of Socio-Economic Problems of the Regions of the Bashkir State University (Ufa, Russia). ORCID: <https://orcid.org/0000-0001-9876-9202>, ResearcherID: K-6878-2017, Scopus Author ID: 57192164952.

Research interests: methods of statistical (econometric) modeling, predictive analytics, socio-economic and biological systems.

Lackmania@mail.ru

Venera M. Timiryanova

Candidate of economic sciences, associate professor, senior researcher, deputy head of the Laboratories for the Study of Socio-Economic Problems of the Regions of the Bashkir State University (Ufa, Russia). ORCID: <https://orcid.org/0000-0002-1004-0722>, ResearcherID: N-3449-2015, Scopus Author ID: 57194428883.

Research interests: methods of statistical (econometric) modeling, consumer behavior, spatial economics.

79174073127@mail.ru

Статья поступила в редакцию 1.06.2021; после рецензирования 9.09.2021 принята к публикации 27.09.2021. Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 1.06.2021; revised on 9.09.2021 and accepted for publication on 27.09.2021. The authors read and approved the final version of the manuscript.