

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ В ИНТЕРНЕТ-МАРКЕТИНГЕ

В данной статье рассматриваются вопросы, позволяющие развивать интернет-бизнеса с применением интернет-технологий. Изучены показатели продвижения сайта. Анализ запросов проводится с использованием многомерных методов, а именно методов кластерного анализа, однофакторного и двухфакторного дисперсионных анализов. Для кластеризации ключевых слов по приоритетным группам использовался статистический пакет SPSS.

Ключевые слова: многомерный статистический анализ, метод к-средних, интернет-маркетинг, продвижение сайта.

Современная политика большинства компаний предполагает постоянное отслеживание эффективности развития основных показателей с учетом широкого внедрения интернет-технологий ведения бизнеса. Это новое направление маркетинга бурно развивается. В современной научной литературе обсуждаются различные аспекты интернет-маркетинга. Аналитика разрабатывается по широкому спектру вопросов: от финансовых до социально-экономических. Интернет-деятельность как показатель ведения бизнеса в настоящее время играет ключевую роль. В работе в качестве примера решения важной стратегической задачи деятельности одной из компаний рассматривается специфика поискового продвижения сайтов в сети Интернет, а также различные методологии для увеличения конверсии с сайтов. Современные бизнес-технологии активно базируются на использовании интернет-технологий, однако любая аналитическая деятельность требует использования математического инструментария. Широкий спектр задач решается многомерными статистическими методами. Начало аналитической деятельности предполагает грамотный сбор информации и последующую обработку. Современные бизнес-технологии оценки прогноза показателей развития компании с необходимостью требуют широкого внедрения аналитической компоненты, заложенной в программе. Именно многомерные статистические методы в настоящее время являются актуальными и практически значимыми, так как позволяют комбинированно анализировать большие массивы данных.

Таким образом, в статье представлена методология использования многомерных статистических методов в интернет-маркетинге. Комбинирование методов однофакторного и двухфакторного дисперсионного анализов, корреляционного анализа и кластерного анализа в сочетании с использованием информационных продуктов Microsoft Excel, статистических пакетов SPSS, Statistics и сервиса сбора интернет-статистики «Яндекс Метрика» позволяет повысить показатели качества работы сайта.

Современная научная литература к основным элементам online-маркетинга относит известные показатели, такие как *цена, продукт, место продаж, продвижение продукта на рынке*, последнее предполагает развитие комплекса мер по продвижению сайта и самого товара в поисковых сетях. Включает в себя огромный арсенал инструментов (поисковое продвижение, контекстную рекламу, баннерную рекламу, e-mail-маркетинг, аффилиативный маркетинг, вирусный маркетинг, скрытый маркетинг, интерактивную рекламу, работу с блогами и т. д.). Кроме того, в литературе хорошо описаны методологии для решения проблемы продвижения продукта на рынке: контекстная реклама, медийная реклама, e-mail-маркетинг, вирусный маркетинг и многие другие методы, которые способ-

* © Трусова А.Ю., 2018

Трусова Алла Юрьевна (a_yu_ssu@mail.ru), кафедра математики и бизнес-информатики, Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Российская Федерация, г. Самара, Московское шоссе, 34.

ствуют развитию рекламы. Особо выделяется в настоящее время методология Seo как поисковое продвижение сайта в сети Интернет. Определение Seo – «комплекс мер по внутренней и внешней оптимизации сайта для повышения позиций ресурса по определенным поисковым запросам в результатах поисковой выдачи с целью увеличения целевого трафика на сайт и последующего получения дохода от этого трафика».

Основным в методологии Seo в настоящее время становится структуризация сайта и запросов по нему, что определяется оптимизацией группы запросов, которые повысят индекс посещения сайта и выполнения на нем заказа.

Метод качественных проб или ручная обработка в настоящее время не позволяет повысить быстро эффективность сайта. В исследовании использовались данные из средств сбора статистики «Яндекс Метрика» и показатели из онлайн сервисов «Яндекс» Wordstat, Seopult, MegaIndex. Данные сервисы представляют различные статистические показатели, например,

- частотность поискового запроса,
- наличие или отсутствие сезонности,
- стоимость клика в директе,
- позицию сайта по конкретному запросу в поисковой выдаче,
- бюджет на продвижение запроса в ссылочных агрегаторах,
- количество документов по этому запросу в выдаче «Яндекса»,
- источники переходов,
- поведенческий фактор и многие другие показатели.

Вопросы влияния различных параметров интернет-оценки на конверсию сайта, позиции сайта в поисковой выдаче, уровень продаж с сайта и т. п. являются актуальными при ведении интернет-бизнеса. Для эффективного решения этой задачи с успехом реализуется подход многофакторного анализа, классификации с обучением и без обучения.

Специалисты выделяют в качестве ключевого такой показатель, как поведенческий фактор, который влияет на поисковую выдачу «Яндекса». К одной из функций поведенческого фактора относится корректировка результатов поисковой выдачи на основании анализа поведения пользователя при работе на сайте. Уровень поведенческого фактора определяется как «высокий» в случае удовлетворения пользователем всех запросов, и его можно оценить, используя глубину просмотра страниц, показатель отказов, показатель времени, проведенного на сайте. Рассчитанные значения поведенческого фактора находятся в интервале от 3 до 8, что свидетельствует о среднем уровне эффективности работы сайта и необходимости выявления степени влияния компонент аналитики друг на друга. Следующим важным показателем являются источники переходов как параметр сбора статистики с сайта. В этой связи изучение влияния источников перехода на конверсию сайта является актуальным и практически значимым. Таким образом, актуальность исследования обосновывается возможностью определять, какой отложенный спрос формируют те или иные источники трафика. В современной научной литературе методы выявления степени влияния источников перехода на конверсию изучены слабо. В исследовании рассматривается одна из возможных ситуаций анализа поисковой системой источников перехода, влияющих на конверсию сайта. В настоящее время существует семь категорий трафика, которые объединены общим понятием – «источники перехода». В литературе детально сформулированы следующие источники трафика: переходы по рекламе, переходы из поисковых систем, внутренние переходы, прямые заходы, переходы по ссылкам на сайтах. В статье рассматривается изучение влияния источников перехода на сайт средствами дисперсионного анализа. Данный математический метод направлен на поиск зависимостей в экспериментальных данных при исследовании значимости различий в средних значениях.

В таблице представлены данные сервиса сбора статистики «Яндекс Метрика» по видам переходов за 2017 год.

Таблица

Статистические данные по видам переходов

2017 год	Классификация переходов				
	из поисковых систем	Прямые заходы	по ссылкам на сайтах	по рекламе	Внутренние
январь	6394	7166	1026	8543	1890
февраль	6325	7278	1933	8934	3020
март	6635	7260	1848	6245	3320
апрель	6729	6940	1598	7557	3043
май	4984	6677	1498	8590	1805
июнь	4743	6301	1497	8686	1858
июль	4907	6412	1470	8250	2872
август	4740	6370	1420	5961	2681
сентябрь	4679	6640	1441	8161	2817
октябрь	4517	6435	1452	8792	3758
ноябрь	4176	6385	1445	8958	3602
декабрь	3961	6038	332	7888	2518

В таблице представлено количество конверсионных переходов на сайт в месяц из разных источников. По результатам однофакторного дисперсионного анализа установлено, что влияние фактора источника перехода на изучаемые показатели конверсии является статистически значимым на уровне 0,05. Кроме того, результаты свидетельствуют об отсутствии статистически значимого влияния фактора времени (месяца) на изучаемые показатели конверсии. Применение двухфакторного дисперсионного анализа предполагает изучение введения двух факторов, а именно – источников перехода трафика на сайт и фактора времени (по месяцам). Установлено, что источники перехода на сайт являются причиной изменчивости случайной величины. Выборочное значение критерия Фишера для фактора времени не попадает в критическую область, так как $2,58 < 184,863$. Таким образом, фактор времени является причиной изменчивости случайной величины. Изучение комплексного влияния двух факторов: времени и источника перехода свидетельствует об их статистически значимом влиянии на конверсию, что подтверждается выводами статистики двухфакторного дисперсионного анализа.

Таким образом, существенные различия между средними значениями по группам, а именно: между источниками переходов на сайт: из поиска, из рекламы, внутренних переходов, по ссылкам и так далее – подтверждаются результатами факторного анализа.

Известно, что чем большее количество целевых запросов охватывает сайт, тем выше его позиции в рейтинге поисковых систем и тем большее количество клиентов он может привести. Продвижение сайта по нескольким высокочастотным запросам не даст того результата, который поможет окупить затраты на продвижение ресурса. Самыми распространенными источниками сбора семантики являются: Yandex Wordstat, подсказки в поиске, счетчики сбора статистики и всевозможные базы запросов.

Первоначальная обработка собранных запросов предполагает первичную сортировку по ключевым словам. Цель данной сортировки – исключить слова с нулевой чистой частотностью и некоммерческим смыслом. Сформированный таким образом список запросов разделяется на группы, согласно которым проводится оптимизация страниц под эти запросы. Применение пакета SPSS позволяет в исследовании сократить время обработки и тем самым обеспечить время для аналитической обработки. Следующий шаг оптимизации требует провести качественный прогноз позиции отобранных ключевых слов, а затем проводится классификация запросов из семантического ядра. Таким образом планируется позиция, которую должны занять в конце этапа выдвижения сайта. Для расчетов

были выбраны такие показатели, как: примерное количество переходов в месяц (по выбранной позиции), прогноз CTR (по выбранной позиции), средняя установленная цена клика в рублях, примерный бюджет в рублях, количество конкурентов в выдаче, частотность и т. п. Другими словами, существует зависимость между прогнозной позицией запроса и каждым из показателей, к примеру, количеством конкурентов в поисковой выдаче. Чем больше конкурентов представлено в поиске, тем сложнее занять одну из первых десяти позиций по данному запросу и так далее. На рис. 2 показан фрагмент таблицы с перечисленными выше параметрами и используемыми значениями.

Процедура проведения кластеризация запросов по эффективности может быть осуществлена как с обучением, так и без обучения. Важное преимущество данного метода – не учитывается мнение маркетолога, тем самым исключается влияние данного мнения на результат кластеризации. Таким образом, результат, полученный после кластеризации позволяет решать множество задач сеопродвижения и формирования стратегии интернет-маркетинга ресурса. В статье применялся итеративный метод кластеризации – метод k-средних.

В анализе используются нормированные значения показателей параметров оценки по каждому ключевому запросу. Целью анализа является получение кластеров или групп запросов, которые наиболее приоритетны к продвижению.

Преимущества данного метода в том, что с ним удобно работать при большом числе объектов исследования.

Результаты анализа представляются в виде нескольких сводных таблиц. Таблица «Принадлежность к кластерам» позволяет определить, к какому кластеру относится определенный запрос. На рис. 1 представлен скриншот таблицы «Принадлежность к кластерам».

Принадлежность к кластерам			
Номер наблюдения	VAR00001	Кластер	Расстояние
1	аппарат для напыления пеноп	6	34,354
2	емкости для компонентов ппу	4	30,424
3	жесткий пенополиуретан	7	54,641
4	жесткий ппу компоненты	7	36,243
5	жидкая резина для гидроизол	2	54,411
6	жидкая резина для гидроизол	2	60,475
7	жидкая резина для гидроизол	2	94,515
8	закрытаячешуйчатый ППУ	1	29,739
9	заливка колодезной кладки ПП	3	45,055
10	заливка межстенного прост	3	51,537
11	заливка пенополиуретана ппу	3	36,263
12	заливка полостей ппу	3	48,111

Рис. 1. Результат кластерного анализа. Таблица «Принадлежность к кластерам»

Запрос	A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	
																		Примерное количество
1	Запрос																	
2	аппарат для напыления пенополиуретана	1	1	1	1	1	1	1	1	2,00	50,00	26,20	0,60	1,20	41,00	21000000,00	5,00	
3	емкость для компонентов ппу	1	1	1	1	1	1	1	1	2,00	40,00	29,10	0,50	1,00	23,00	21000000,00	4,00	
4	квартый пенополиуретан	1	1	1	1	1	1	1	1	3,00	4,41	68,40	4,70	14,10	44,00	21000000,00	4,00	
5	квартый ппу компоненты	1	1	1	1	1	1	1	1	3,00	4,41	68,40	4,70	14,10	8,00	40000000,00	1,00	
6	кардас сезона для прорезки	1	1	1	1	1	1	1	1	8,00	1,32	196,30	21,40	171,20	0,00	21000000,00	5,00	
7	кардас сезона для прорезки	1	1	1	1	1	1	1	1	3,00	3,33	106,10	9,60	26,80	50,00	83000000,00	5,00	
8	кардас сезона для прорезки	1	1	1	1	1	1	1	1	4,00	3,68	100,00	12,10	48,40	3,00	65000000,00	2,00	
9	запись по запросу ППУ	1	1	1	1	1	1	1	1	2,00	33,33	31,00	0,60	1,20	0,00	24000000,00	0,00	
10	запись по запросу ППУ	1	1	1	1	1	1	1	1	2,00	33,33	31,00	0,60	1,20	30,00	76000000,00	2,00	
11	запись по запросу ППУ	1	1	1	1	1	1	1	1	2,00	33,33	116,30	0,80	1,60	56,00	14000000,00	4,00	
12	запись по запросу ППУ	1	1	1	1	1	1	1	1	2,00	33,33	116,30	0,80	1,60	6,00	23000000,00	1,00	
13	запись по запросу ППУ	1	1	1	1	1	1	1	1	2,00	33,33	116,30	0,80	1,60	29,00	89000000,00	4,00	
14	запись ппу	1	1	1	1	1	1	1	1	3,00	4,17	79,50	8,10	24,30	4,00	77000000,00	4,00	
15	запись ппу в полости	1	1	1	1	1	1	1	1	2,00	28,57	174,60	0,30	0,60	4,00	70000000,00	4,00	
16	запись ппу в стене	1	1	1	1	1	1	1	1	2,00	14,29	23,20	0,80	1,60	32,00	10000000,00	4,00	
17	запись ппу в форму	1	1	1	1	1	1	1	1	2,00	40,00	19,40	0,10	0,20	69,00	77000000,00	2,00	
18	запись ппу цена	1	1	1	1	1	1	1	1	2,00	40,00	68,10	0,70	1,40	1,00	47000000,00	5,00	
19	запись ппу	1	1	1	1	1	1	1	1	2,00	14,29	85,00	1,40	2,80	25,00	11000000,00	2,00	
20	запись ппу	1	1	1	1	1	1	1	1	2,00	66,67	29,50	0,50	1,00	0,00	54000000,00	2,00	
21	записки оборудование ппу	1	1	1	1	1	1	1	1	2,00	40,00	35,00	0,40	0,80	6,00	11000000,00	0,00	
22	записки установка ппу	1	1	1	1	1	1	1	1	0,00	0,00	0,00	0,00	0,00	31,00	18000000,00	1,00	

Рис. 2. Параметры для оценки ключевых запросов

В соответствии с полученными результатами была проведена группировка запросов по кластерам. Результат группировки представлен на рис. 3. Данная таблица является руководством к действию для seo-оптимизатора. Важно отметить, что запросы из одного кластера будут более приоритетными для продвижения и, наоборот, запросы из другого кластера будут являться менее приоритетными к продвижению. Кроме разделения по кластерам напротив каждого запроса представлены его эффективность и балл приоритета. После расчета среднего значения по каждому параметру был получен средний балл эффективности и приоритета для каждого кластера. На основании результатов делается вывод о преимуществе продвижения запросов из одного кластера над другими.

Запросы	Позиция	Эффективность	Приоритет	Частотность
+ 6	27,71	79,09	45,12	219,86
оборудование для напыления ппу	24,00	80,89	121,77	515,00
оборудование пенополиуретан	42,00	83,51	101,78	294,00
пенополиуретан оборудование	32,00	83,98	43,83	294,00
оборудование для полимочевины	20,00	77,48	12,83	129,00
оборудование по производству плит	29,00	78,73	13,10	118,00
оборудование для напыления пенополи	24,00	75,87	11,80	108,00
цена на оборудование для ППУ	23,00	75,42	10,93	83,00
+ 4	9,20	78,47	22,19	115,60
компоненты ппу	7,00	78,99	10,12	429,00
купить компоненты для пенополиуретан	13,00	82,85	34,29	98,00
расход компонентов ппу	12,00	80,43	44,09	19,00
компоненты ппу для заливки купить	8,00	80,97	12,87	18,00
компоненты ппу от производителя	8,00	71,13	9,80	14,00
+ 7	33,67	61,95	17,55	41,67
жесткий пенополиуретан	32,00	58,34	9,80	117,00
жесткий ппу компоненты	45,00	68,27	24,34	4,00
компоненты для жестких ППУ	24,00	81,23	18,70	4,00
+ 2	28,25	60,72	12,14	492,75
жидкая резина для гидроизоляции	28,00	62,09	10,81	1542,00
жидкая резина для гидроизоляции цена	43,00	63,71	12,48	233,00
жидкая резина для гидроизоляции купить	32,00	58,35	13,18	170,00
оборудование для гидроизоляции жидк	12,00	48,90	14,05	28,00
+ 1	36,15	57,13	11,15	278,59
+ 3	32,19	45,15	10,43	127,60
+ 5	44,57	34,24	9,60	354,16
+ 8	58,50	28,70	9,60	72,15

Рис. 3. Результат группировки поисковых запросов по кластерам

Из рис. 3 видно, что наиболее оптимальными для продвижения являются запросы из кластеров 6, 4, 7 и 2. Причем ключевые слова из кластера 6 необходимо продвигать в первую очередь. В кластере 6 средняя позиция запроса - 27, при этом эффективность - 79, что является значительным показателем. Приоритет в данном кластере 45,12 – это максимальный приоритет среди всех кластеров. При

последующем анализе известные кластеры запросов используются в качестве обучающих выборок, что также позволяет проводить перегруппировку запросов по основным показателям.

Таким образом, развитие интернет-бизнеса в настоящее время требует высокой квалификации в вопросах анализа используемых методологий, а также применения информационных технологий при использовании многомерных статистических методов. Комплексное сочетание методологии математического инструментария с интернет-методиками позволяет решать широкий спектр задач различного уровня, в частности при улучшении качества конверсии с сайтов и анализе ее параметров. Кроме того, грамотная seo-оптимизация сайта на сегодняшний день является одним из ключевых инструментов привлечения клиентов через Интернет. Продвижение сайта в рейтинге поисковых систем напрямую зависит от множества показателей. Ключевыми среди них являются такие, как значение поведенческого фактора, значение источников переходов на сайт, примерное количество переходов в месяц (по выбранной позиции), прогноз CTR (по выбранной позиции), средняя установленная цена клика в рублях, примерный бюджет в рублях, количество конкурентов в выдаче, частотность и другие.

Библиографический список

1. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов. М., 2010. 543 с.
2. Информационный портал uaweb.ua. Персональный SEO блог. URL: <http://uaweb.ua/publication/conversion.html>.
3. Информационный портал yandex.ru. Коллективный IT блог. URL: <https://yandex.ru/blog/metrika-club/4790>.
4. Информационный портал getclientsnow.ru. Персональный SEO блог. URL: <http://getclientsnow.ru/kak-privlech-posetitelej-na-sajt-1>.
5. Информационный портал www.raskruty.ru. Сервис для продвижения сайтов. URL: <http://www.raskruty.ru/povedencheskie-factory.html>.

References

1. Kremer N.Sh. *Teoriya veroyatnostei i matematicheskaya statistika: uchebnik dlya vuzov* [Probability theory and mathematical statistics: textbook for universities]. M., 2010, 543 p. [in Russian].
2. *Informatsionnyi portal uaweb.ua. Personalnyi SEO blog* [Information portal uaweb.ua. Personal SEO blog]. Available at: <http://uaweb.ua/publication/conversion.html> [in Russian].
3. *Informatsionnyi portal yandex.ru. Kollektivnyi IT blog* [Information portal yandex.ru. Collective IT blog]. Available at: <https://yandex.ru/blog/metrika-club/4790> [in Russian].
4. *Informatsionnyi portal getclientsnow.ru. Personal'nyi SEO blog* [Information portal getclientsnow.ru. Personal SEO blog]. Available at: <http://getclientsnow.ru/kak-privlech-posetitelej-na-sajt-1> [in Russian].
5. *Informatsionnyi portal www.raskruty.ru. Servis dlya prodvizheniya saitov* [Information portal www.raskruty.ru. Service for website promotion]. Available at: <http://www.raskruty.ru/povedencheskie-factory.html> [in Russian].

MULTI-DIMENSIONAL STATISTICAL ANALYSIS IN INTERNET MARKETING

This paper discusses issues that allow to develop Internet business with the use of Internet technologies. The parameters of website promotion are studied. The analysis of requests is carried out using multidimensional methods, namely methods of cluster analysis, one-factor and two-factor analysis of variance. The SPSS statistical package was used to cluster keywords by priority groups.

Key words: multivariate statistical analysis, K-means method, Internet marketing, website promotion

Статья поступила в редакцию 17/1/2018.
The article received 17/1/2018.

* *Trusova Alla Yurievna* (a_yu_ssu@mail.ru), Department of Mathematics and Business Informatics, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russian Federation.