

Разработка и исследование технологии геомаркетинга на основе транспортных факторов и нелинейной регрессионной модели

И.А. Тахтаров^а, А.В. Сергеев^а

^а Самарский национальный исследовательский университет имени академика С.П. Королёва, 443001, ул. Молодогвардейская, 151, Самара, Россия

Аннотация

Работа посвящена решению одной из задач геомаркетинга о выборе наиболее выгодных мест для расположения торговых объектов. Предложена технология, использующая нелинейную регрессионную модель на основе пространственных факторов для оценки потенциальной выгоды. Для испытания технологии реализовано экспериментальное программное обеспечение для сбора данных, расчета факторов, построения и применения модели. Проведен эксперимент, в котором рассматриваются транспортные, экономические и демографические факторы и продовольственные торговые объекты по городу Самара.

Ключевые слова: геомаркетинг; регрессия; геоинформатика; ГИС; пространственные факторы

1. Введение

Геомаркетинг - это дисциплина, применяющая геоинформационные инструменты и методы для решения маркетинговых задач. Основной целью геомаркетинга является повышение эффективности предпринимательской деятельности на основе анализа разнородных пространственных данных. Одна из частых задач геомаркетинга – задача об оптимальном размещении нового объекта торговли.

В городской среде торговые объекты зачастую размещаются там, где прибыль может не достигать максимальных показателей, по причине того, что лица, принимающие управленческие решения, не владеют сведениями об окружающей территории и методами обоснованного выбора места. При этом современные геоинформационные технологии способны обеспечить сбор, хранение и обработку разнообразных сведений о городской среде. В настоящей работе изучается взаимосвязь между транспортными, экономическими и демографическими пространственными факторами и эффективностью продовольственных магазинов.

Методом исследования влияния факторов на востребованность торговой площадки является построение математической модели. Учитывая прошлый опыт с моделью линейной регрессии, с которую не удалось достичь приемлемой погрешности, и предполагая, что причины неудач – в более сложном характере взаимосвязей, в настоящем исследовании решено рассматривать нелинейную регрессию с полиномиальной функцией над факторами.

Полученные в результате исследования математическая модель и программное обеспечение могут применяться для выбора одного из нескольких возможных мест размещения нового торгового объекта в интересах населения, предпринимателей и городской администрации и способствовать развитию потребительского рынка города, а именно, применяться для задач:

- создания условий для внедрения новых форматов мелкой розничной торговли;
- повышения инвестиционной привлекательности периферийных территорий города;
- совершенствования и развития сферы торгового обслуживания населения в границах пешеходной доступности от крупных мест приложения труда.

2. Расчет пространственных факторов и обучающей выборки

Для исследования было выбрано пять пространственных факторов. У каждого из них значение определяется в произвольной точке городской среды, при этом значение характеризует всю окрестность заданного радиуса вокруг точки. Факторы окрестного населения, доступности общественного транспорта, интенсивности движения личного транспорта и доступности автомобильных парковок положительно влияют на эффективность торговой точки, фактор влияния конкурентов – наоборот. На рисунке 1 изображена карта города Самары с пространственными объектами, по которым вычисляются факторы. Способ вычисления каждого вида фактора объясняется далее.

Фактор окрестного населения показывает, какое количество горожан проживает в пешей доступности от указанной точки – принято считать их потенциальными покупателями. Для расчета необходим пространственный слой зданий со сведениями о количестве жителей каждого из них. Пространственным запросом выбираются здания, попавшие в круг заданного радиуса от точки, и затем производится суммирование количества жителей с весом, зависящим от удаленности здания:

$$\varphi_{\text{населения}} = \sum_{i=1}^N \begin{cases} \left(1 - \frac{r_{i \text{ здания}}}{r_{\text{шаговая}}}\right) * c_{i \text{ жителей}}, & r_{i \text{ здания}} \in [0; r_{\text{шаговая}}] \\ 0, & \text{иначе} \end{cases}$$

где N – общее количество домов,

$r_{i \text{ здания}}$ – расстояние между исследуемой точкой и i-м зданием,

$r_{\text{шаговая}}$ – радиус шаговой доступности, в исследовании взято значение в 200 м,

C_i жителей – количество человек, проживающих в i -м здании.



Рис. 1. Карта города Самары с пространственными объектами.

Фактор влияния конкурентов показывает, насколько доступны в указанной точке предложения конкурентов. Если фактор окрестного населения характеризует потенциальный спрос, то фактор влияния конкурентов оценивает, насколько этот спрос уже удовлетворен. Значение фактора рассчитывается с помощью пространственного слоя существующих магазинов, в котором каждый объект снабжен условной характеристикой величины. Крупные гипермаркеты имеют характеристику 10, средние сетевые супермаркеты – от 7 до 4, более мелкие магазины – от 3 до 0.

Из слоя выбираются объекты, попавшие в круг заданного радиуса от точки, и затем производится суммирование их характеристик с весом, зависящим от удаленности магазина:

$$\Phi_{\text{конкурентов}} = \sum_{i=1}^N \begin{cases} \left(1 - \frac{r_{i \text{ магазина}}}{r_{\text{макс}}}\right) * W_{i \text{ магазина}}, & r_{i \text{ магазина}} \in [0; r_{\text{макс}}] \\ 0, & \text{иначе} \end{cases}$$

где N – общее количество магазинов,

$r_{i \text{ магазина}}$ – расстояние между исследуемой точкой и i -м магазином,

$r_{\text{макс}}$ – радиус окрестности, в котором допускается влияние конкурентов, в исследовании взято такое же значение, как для радиуса пешей доступности,

$W_{i \text{ магазина}}$ – условная характеристикой величины i -ого магазина.

Фактор интенсивности движения личного транспорта характеризует, насколько много водителей посещает рассматриваемый район в течение суток и могут стать посетителями магазина. Для расчета фактора используется транспортная модель Самары, в которой есть пространственный слой улиц и автодорог, и каждому сегменту дорожной сети сопоставлено количество автомобилей, проезжающих за сутки в прямом и обратном направлении (трафик).

Для расчета значения фактора пространственным запросом выбираются сегменты дорог, находящихся в окрестности заданной величины от точки, и производится суммирование трафика с весом, зависящим от длины сегмента и удаленности от точки.

$$\Phi_{\text{интенсивность движения}} = \sum_{i=1}^N \begin{cases} \left(1 - \frac{r_0}{r_{\text{макс}}}\right) * k_i, & r_{\text{макс}} \in [0; r_{\text{шаговая}}] \\ 0, & \text{иначе} \end{cases}$$

где N – общее количество сегментов дорог,

r_0 – расстояние между исследуемой точкой и ближайшей точкой на i -м сегменте дороги,

$r_{\text{макс}}$ – радиус окрестности, в котором допускается влияние интенсивности движения личного транспорта, в исследовании взято значение 300 м,

k_i – среднее количество автомобилей, проезжающих по i -му сегменту дороги в обоих направлениях за сутки.

Фактор доступности общественного транспорта характеризует, насколько много граждан, не пользующихся личным транспортом, посещает рассматриваемый район. Для расчета фактора используется статистика из информационной системы Транспортного оператора Самары [1], измеряющего относительный пассажиропоток через остановки. Статистика представлена в виде пространственного слоя остановок в виде точечных объектов, с каждым из которых

хранится количество пассажиров, садящихся на транспорт в данной остановке. Для расчета значения фактора суммируется количество пассажиров по остановкам, находящимся в радиусе пешей доступности от точки.

$$\varphi_{\text{доступности общественного транспорта}} = \sum_{i=1}^N \left\{ \left(1 - \frac{r_{\text{остановка}}}{r_{\text{макс}}} \right) * c_i \text{ пассажиров}, r_{i \text{ остановка}} \in [0; r_{\text{макс}}] \right. \\ \left. 0, \text{ иначе} \right.$$

где N – общее количество остановок общественного транспорта,

$r_{\text{остановка}}$ – расстояние между исследуемой точкой и i-й остановкой,

$r_{\text{макс}}$ – радиус окрестности, в котором допускается влияние остановок общественного транспорта, в исследовании взято значение 300 м,

$c_i \text{ пассажиров}$ – количество пассажиров, осуществляющих посадку на транспорт на i-й остановке.

Фактор доступности автомобильных парковок дополняет фактор интенсивности движения личного транспорта и характеризует, настолько легко сможет остановиться водитель для посещения магазина. Способ расчета фактора совпадает с расчетом доступности общественного транспорта, но при этом используется пространственный слой парковок с характеристикой количества мест машин.

Отдельной задачей является получение сведений о магазинах, существующих на территории города, с хорошей актуальностью и полнотой. Существуют открытые интернет-сервисы, собирающие сведения об объектах городской инфраструктуры и предоставляющих их для справочных целей. Среди них выделяется качеством «Яндекс.Справочник» [2], имеющий программный интерфейс (API) для доступа к данным. Для исследования была создана утилита, загружающая из этого источника сведения по всем объектам из категории «Продовольственные магазины» в составе названия и координат и создающая из них пространственный слой. К сожалению, данная справочная система не предоставляет сведений о величине торговой площадки или количестве покупателей, поэтому, ранжирование объектов по величине произошло на основании собственных представлений об этих магазинах. По названиям были выделены крупные гипермаркеты и сетевые супермаркеты, им присвоены более высокие характеристики, чем остальным магазинам.

Таким способом был составлен пространственный слой из 643 продовольственных магазинов. Из него случайным образом были выбраны 30 объектов в качестве контрольной выборки, остальные составили обучающую выборку. Тот же слой в полном составе используется для расчета фактора влияния конкурентов.

Расчет пространственных факторов, согласно описанным выше процедурам, является типичной задачей для геоинформационной системы (ГИС). От ГИС требуется обеспечить хранение пространственных слоев, предоставлять базовый набор пространственных операторов (геокодирование, поиск по пространственным условиям, определение геометрических характеристик) и предоставлять программный интерфейс для реализации собственных функций. В исследовании в качестве базовой ГИС использован Геопортал Самарской области [3], общедоступный геоинформационный интернет-сервис, удовлетворяющим всем этим требованиям.

Для расчета значений факторов была реализована утилита на JavaScript, загружающая виджет с базовой картой из геопортала Самарской области и проводящая процедуру расчета каждого фактора через API виджета. Результаты расчета утилита сохраняет в табличной форме, пример результатов приведен на рисунке 2.

	A	B	C	D	E	F	G	H	I
1	ID	name	W	nas_w	nas	bus	cars	park	rivals
2	13007144	Пятёрочка	5	1286.801847	2976	0	9704.187044	0	4.777319299
3	13007148	Сластена	2	86.16047314	413	531.5789	9916.621787	0.687034911	0
4	13007149	Ноэлл	2	797.7483	2226	2784.354	10052.42597	0	4.449616977
5	13007151	Пятёрочка	5	1187.948907	2913	3078.591	13046.13498	0.201063605	1.458439452
6	13007153	Шик	2	1153.962172	2722	1812.619	8399.942122	0	6.897747363
7	13007154	Универсам	5	675.5872651	2623	791.0387	4760.472749	0	1.949783896
8	13007161	Магнит	2	1351.615989	3174	0	1800.160763	0	1.864815015
9	13007164	Classic продукт	2	951.8013422	2918	1172.098	3156.885148	0.552172165	3.524428916
10	13007166	Эврика	2	734.3643837	1699	0	0	0.363567602	0
11	13007167	Магазин Продпроттовары № 55	2	117.3645408	379	0	431.9595602	0	0
12	13007170	Кондитерские изделия	2	481.4936243	1646	4617.925	0	9.70E-02	0.160967335
13	13007170	Магазин № 1	2	31.4055717	134	0	1435.422434	0	2.675297224
14	13007183	Торговое предприятие СамараПродТо	2	381.1589372	1117	7574.415	12613.16462	0	4.635677424
15	13007184	У Пальча	2	480.8578827	878	895.8788	871.3008873	0	1.065616557

Рис. 2. Результаты расчета значений факторов.

3. Построение математической модели

Как указывалось выше, для исследования была выбрана нелинейная регрессионная модель.

Регрессия обозначает возвращение от множества фактов к порождающему их свойству. Под словом «модель» понимают совокупность формул, уравнений, которые количественно выражают связь между величинами, т.е. закон функционирования экономической системы. Регрессионная модель связывает в виде одного равенства основную

результатирующую переменную, называемую объясняемой переменной, с одной или несколькими объясняющими переменными [4].

Была выбрана полиномиальная модель над 5 независимыми переменными (факторами) со степенями не выше 2, таким образом, всего в модели 21 параметр, относительно которых модель линейна. Зависимой переменной выступает условная характеристика величины магазина. Значения факторов перед расчетом нормировались по значениям, то есть приводились к диапазону [-1, 1]. Общий вид примененной модели, следующий:

$$Y = f_0 + \varphi_n * (f_{11} + f_{12} * \varphi_n + f_{13} * \varphi_d + f_{14} * \varphi_u + f_{15} * \varphi_n + f_{16} * \varphi_k) + \varphi_d * (f_{21} + f_{22} * \varphi_d + f_{23} * \varphi_u + f_{24} * \varphi_n + f_{25} * \varphi_k) + \varphi_u * (f_{31} + f_{32} * \varphi_u + f_{33} * \varphi_n + f_{34} * \varphi_k) + \varphi_n * (f_{41} + f_{42} * \varphi_n + f_{43} * \varphi_k) + \varphi_k * (f_{51} + f_{52} * \varphi_k)$$

где Y – критерий эффективности магазина,

φ_n – фактор окрестного населения,

φ_d – фактор доступности общественного транспорта,

φ_u – фактор интенсивности движения личного транспорта,

φ_n – фактор доступности автомобильных парковок,

φ_k – фактор влияния конкурентов.

Для расчета модели был применен программный пакет IBM SPSS Statistics, а именно, модуль построения нелинейной регрессии по произвольному выражению [5]. Результатом расчета являются значения параметров модели с доверительными интервалами:

Таблица 1. Оценки параметров

Параметр	Оценка	Стандартная ошибка	Нижняя граница	Верхняя граница
f0	2,279	0,565	1,170	3,388
f11	-0,916	0,762	-2,411	0,580
f12	0,128	0,352	-0,563	0,819
f13	-1,731	0,937	-3,571	0,109
f14	-0,207	0,682	-1,546	1,132
f15	0,608	0,548	-0,468	1,684
f16	-0,636	0,448	-1,516	0,244
f21	-0,876	0,887	-2,619	0,866
f22	-0,442	0,530	-1,483	0,599
f23	0,425	0,885	-1,313	2,162
f24	0,062	0,640	-1,196	1,320
f25	0,163	0,691	-1,195	1,520
f31	-0,677	0,696	-2,044	0,689
f32	0,212	0,497	-0,763	1,188
f33	-0,630	0,545	-1,700	0,439
f34	-0,647	0,508	-1,644	0,350
f41	0,255	0,535	-0,795	1,304
f42	0,125	0,449	-0,756	1,006
f43	-0,242	0,418	-1,062	0,579
f51	-1,225	0,538	-2,281	-0,168
f52	-0,776	0,288	-1,341	-0,210

Для верхней и нижней границ доверительный интервал составляет 95%. С помощью контрольной выборки было рассчитано среднее квадратическое отклонение, которое составило 1.226.

4. Заключение

В работе была сделана попытка решить задачу геомаркетинга с помощью построения математической модели. Погрешность построенной модели оказалась достаточно велика, она составляет порядка 15% от диапазона значений зависимой переменной. Однако, для данной предметной области это нормальное явление, и полученные результаты могут применяться на практике для оценивания перспективности размещения торговых объектов в том или ином месте.

Недостатки модели могут объясняться несколькими причинами. Во-первых, функционирование потребительского рынка не всегда подчиняется рациональным соображениям, магазины размещаются не в лучших условиях, и поэтому срок деятельности заметной доли объектов не превышает года [6, 7]. Во-вторых, при построении модели не

использовались достоверные экономические показатели работы магазинов, вместо них выбирались приближенные значения. В-третьих, характер зависимости может быть сложнее и зависеть от факторов, не вошедших в настоящее исследование. При устранении перечисленных недостатков можно получить более точную модель, чему и будут посвящены дальнейшие исследования.

Литература

- [1] Транспортный оператор Самары. Программные интерфейсы (API) [Электронный ресурс]. – <http://tosamara.ru/api/> (дата обращения 11.01.2017 г.).
- [2] Яндекс.Справочник. Документация о сервисе поиска по организациям [Электронный ресурс]. – <https://tech.yandex.ru/maps/doc/geosearch/concepts/about-docpage/> (дата обращения 11.01.2017 г.).
- [3] Геоортал электронного правительства Самарской области [Электронный ресурс]. – <http://geoportal.samregion.ru/about/> (дата обращения 11.01.2017 г.).
- [4] Нелинейная регрессия [Электронный ресурс]. – http://www.machinelearning.ru/wiki/index.php?title=Нелинейная_регрессия (дата обращения 13.01.2017 г.).
- [5] Улучшение точности предсказаний с помощью современных методов регрессии [Электронный ресурс]. – <http://www-03.ibm.com/software/products/ru/spss-regression> (дата обращения 13.01.2017 г.).
- [6] Схема оптимизации размещения нестационарных торговых объектов г.о. Самара. Научное обоснование [Электронный ресурс]. – <http://archive.samadm.ru/node/22321> (дата обращения 11.01.2017 г.).
- [7] О переходе к саморегулированию оценочной деятельности [Электронный ресурс]. – http://dpr.ru/journal/journal_16_20.htm (дата обращения 19.01.2017 г.).