

Выявление скрытых закономерностей в реакции общества на бренд: анализ привлекательности названия методами машинного обучения

Е. М. Татур, Е. В. Клименко

Санкт-Петербургский государственный университет

katiandkate@gmail.com, ekaterinaklimenko0700@gmail.com

Аннотация

В статье рассматривается проблема выявления скрытых закономерностей в реакции общества на бренды посредством оценки привлекательности названия методами машинного обучения. Авторы проводят анализ большого объёма данных, включающих названия различных торговых марок и информацию об их популярности среди потребителей, с применением математических методов и алгоритмов машинного обучения. В ходе исследования проведены эксперименты со статическими предсказывающими моделями Word2Vec. Результаты исследований доказывают, что разрабатываемый авторами подход позволяет находить названия, удовлетворяющие требованиям запросам сферы маркетинга.

Ключевые слова: классификация текстов, бренд, нейминг

Библиографическая ссылка: Татур Е. М., Клименко Е. В. Выявление скрытых закономерностей в реакции общества на бренд: анализ привлекательности названия методами машинного обучения // Компьютерная лингвистика и вычислительные онтологии. Выпуск 7 (Труды XXVI Международной объединённой научной конференции «Интернет и современное общество», IMS-2023, Санкт-Петербург, 26–28 июня 2023 г. Сборник научных статей). — СПб: Университет ИТМО, 2024. С. 60–66. DOI: 10.17586/2541-9781-2024-7-60–66

1. Введение

Название бренда является одним из возможных способов привлечения внимания клиента к компании. Коммуникативная эффективность названия оказывает прямое влияние на возможность выхода торговой марки на рынок, а также успешность её продвижения. Наименование бренда представляет собой его базовый, наиболее константный атрибут, смена которого крайне нежелательна, так как порождает множество маркетинговых проблем. Маркетинговая значимость названия торговой марки объясняется большим количеством требований, предъявляемых к результату [1]. Имя бренда должно точно отображать заключённую в него идею и ценности компании, ясно выражать его конкурентные преимущества, а также положительно эмоционально ассоциироваться с объектом нейминга. В теории нейминга существуют различные алгоритмы разработки названия торговой марки.

2. Что такое нейминг?

Нейминг — процесс создания названия для компании, продукта, услуги, бренда или любого другого объекта. Он включает в себя анализ целевой аудитории, конкурентов, маркетинговых стратегий и т.д. Цель нейминга — создать уникальное, запоминающееся и привлекательное название, которое отражает суть объекта и привлекает внимание

потенциальных клиентов [2]. Названия могут иметь эмоциональный и символический аспект. Нейминг также может вызывать определенные ассоциации, чувства и атмосферу. Например, название «Coca-Cola» может вызывать ассоциации с безалкогольными напитками и праздниками.

Лингвистическая природа нейминга также связана с грамматическими и синтаксическими аспектами языка. Названия могут создаваться с помощью различных морфологических и синтаксических процессов словообразования, таких как суффиксация, префиксация, композиция и т.д. Например, название «Photography» образовано путём добавления суффикса «-graphy» к слову «photo», а наименование «ReDesign» появилось путём присоединения префикса «-re» к слову «design».

Лингвистическая природа нейминга определяет разнообразие способов формирования неймов, включая следующие:

- вторичная номинация — создание нейма путём использования уже существующего слова;
- словообразовательная модель — создание нейма путём применения продуктивной словообразовательной модели;
- фонетическая модель — создание нейма путём изменения произношения или ударения известного слова;
- синтаксическая модель — создание нейма путём сочетания слов в определенной синтаксической структуре и т.д.

Таким образом, лингвистическая природа именования включает в себя изучение процессов создания имён, их семантики и символического значения, а также их функции в коммуникации.

Алгоритм разработки названия торговой марки может быть замечен при изучении структуры названия. Например, оно может содержать название сферы и местоположение компании или филиала, что говорит о прямом конструировании названия торговой марки. Подобные действия удлиняют название, что может влиять на его запоминаемость или благозвучие.

3. Процесс проведения исследования

В рамках исследования рассматриваются следующие алгоритмы: название, основанное на ключевых словах, связанных с брендом и его ценностями; поиск синонимов, ассоциаций и метафор для расширения ключевых слов; генераторы названий брендов. Также существуют возможности выбрать пустые название, выбранное лишь из принципа не нарушения авторских прав или торговых марок других компаний. Такие примеры рассматриваются отдельно на основе результатов исследования.

Для проведения данного исследования был собран эмпирический материал — названия брендов в количестве 910 с разделением их на сферы бизнеса — 70 брендов в каждой сфере. Разделение на сферы бизнеса сделано с выделением смежных сфер: кафе и рестораны, санатории, базы отдыха, отели и так далее (табл. 1). В ходе исследование сравниваются оценки в смежных сферах.

В рамках исследования наиболее подходящим методом тестирования была выбрана оценочная шкала, поскольку она предоставляла респондентам возможность выставить оценку по различным критериям, что в свою очередь помогало полно и всесторонне оценить названия брендов. Оценка производилась по конкретно заданным показателям, покрывающим основные характеристики, требующие анализа. Были выделены основные критерии, которые выбирались респондентами без зависимости между значениями: соответствие, запоминаемость, благозвучие, понятность, оригинальность, привлекательность, выразительность. Каждый из критериев отвечал за узнаваемость названия бренда. Так, например, соответствие позволяет человеку понять, к какой сфере бизнеса относится компания без дополнительной информации (логотип, слоган и другие).

Также в ходе опросов были собраны информация о респондентах: возраст, пол, сфера деятельности.

Таблица 1. Соответствие каждой сфере смежных к ней сфер с учётом синонимичности и вида деятельности

№	Сфера бизнеса	Наличие смежной сферы
1	Косметология	Фармацевтика, Салоны красоты
2	Фармацевтика	Косметология, Взрослые клиники
3	Рекламные агентства	Туристические агентства
4	Транспортные компании	-
5	Салоны красоты (парикмахерские)	Косметология
6	Санатории и базы отдыха	Отели
7	Рестораны	Кафе
8	Отели	Санатории и базы отдыха
9	Взрослые клиники	Фармацевтика
10	Магазины одежды	-
11	Продуктовые магазины	Кафе, Рестораны
12	Туристические агентства	Рекламные агентства
13	Кафе	Рестораны

В процессе исследования использовалась оценочная шкала от «0» до «5», где «0» — несоответствие критерию, а «5» — полное соответствие критерию. В ходе исследования оценки объединялись в три группы специально для того, чтобы учесть погрешность. Оценки «0» и «1» относятся к первой группе, «2» и «3» ко второй и «4» и «5» к третьей. Важно отметить, что высокое соответствие оценок позволяет неймингу выполнять свою основную задачу отражения уникальных отличий, которые попадут в сознание потребителей и создадут нужные ассоциации в необходимые моменты. Поиск таких зависимостей позволит оценивать нейминги на соответствие задачам, которые перед маркетологами ставят компании.

Для проверки принадлежности нейминга или его части к синонимам, ассоциациям и метафорам, связанными со сферой бизнеса, используются статические предсказывающие модели Word2Vec [3; 4; 5]. Далее сравниваются оценки неймингов и пути их создания в смежных сферах. В процессе работы дополнительно использовалась библиотека Gensim. Были разделены все части речи, удалены знаки препинания и стоп-слова. В таблице 2 представлено описание корпусов, на основе которых используются статические предсказывающие модели Word2Vec. В данном случае, под стоп-словами подразумеваются слова, которые использовались для стилизации названия или широкого обозначения местности. Под это обозначение в предложенном контексте попали части наименований, дублирующие названия сайтов компаний: www, ru, ru и так далее. Названия разбивались на части вручную, также сокращённые названия дополнялись.

Таблица 2. Описание лингвистических корпусов

№	Постоянный идентификатор	Корпус	Размер корпуса (слов)	Размерность вектора
1	ruwikiruscorpورا_upos_cbow_300_10_2021	НКРЯ и Википедия за ноябрь 2021	1.2 млрд	300
2	ruscorpора_upos_cbow_300_20_2019	НКРЯ	270 млн	300

Синонимы с помощью модели Word2Vec находились и внутри названий самих сфер. В таблице 1 представлены смежные сферы и синонимы. Стоит отметить, что некоторые сферы в дальнейшем можно разделить и изучить их оценки отдельно. Например, базы отдыха и санатории - учреждения этих сфер различаются в своей специфике, что влияет на целевую аудиторию. Санатории подразумевают длительный отдых под присмотром

медицинского персонала, а базы отдыха могут использовать в более короткий срок и с иными целями. Также важно отметить, что смежные сферы у базы отдыха и санаториев различны.

4. Результаты исследования

Рассмотрим пример из раздела аптек: Башфармация заменяется в Башкортостанская фармация. В названиях, состоящих из нескольких слов, производился поиск ближайшего синонима сферы. В результате слова с наличием близкого синонима всегда имели оценки по соответствию из 1 группы. Далее к группе синонимов требуется добавить ассоциации и метафоры, поскольку в неймингах часто используется не синонимичная близость, а культурные ассоциации или транслитерация синонимичных слов.

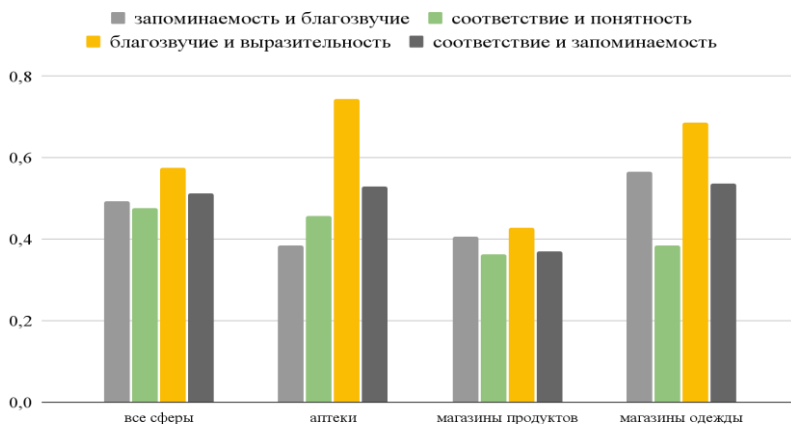


Рис. 1. Сравнение средних значений соответствия оценок запоминаемость и благозвучие, соответствие и понятность, благозвучие и выразительность, соответствие и запоминаемость среди всех сфер, аптек, магазинов продуктов и магазинов одежды

При дальнейшем рассмотрении соответствия значений оценок было выявлено, что для каждой сферы среднее значение двух параметров уникально. Эту тенденцию можно наблюдать на рисунке 1, представляющей соответствия различных оценок как в трех несвязанных сферах, так и среди всех представленных сфер. Рассмотрим среднее значение соответствия оценок запоминаемости и благозвучия после группировки в сферах «Аптеки», «Санатории» (в которые входят санатории и базы отдыха), и «Взрослые поликлиники». Ни одна из них не имеет значение выше среднего, а сфера аптек имеет наименьшее соответствие оценок запоминаемости и благозвучия.

При изучении причин полученного факта было замечено, что сфера аптек имела наибольшее количество стоп-слов. Множество неймингов были составлены из сокращений, которые влияют на благозвучие названий. Среди трех сфер, представленных к сравнению, именно аптеки по среднему показателю имеют большую длину — 14,8 символов, так как некоторые состоят из соединения сокращений, обозначающих регион, в котором работает компания, и ассоциативных слов, связанных со сферой здравоохранения. Самые короткие названия из трех сфер у взрослых поликлиник — средняя длина 10,1 символ.

При детальном рассмотрении на рисунке 2 соответствия высоких оценок запоминаемости и благозвучия, было замечено, что в сфере аптек несмотря на наименьший показатель соответствия двух оценок, большинство из них являются наивысшими, из чего следует зависимость этих показателей в данной сфере. Было также отмечено, что нейминги с соответствующими высокими оценками имеют всего два названия со стоп-словами, среднее значение длины меньше на 1,5 символа — 13,3 символов против 14,8 символов общей средней длины среди неймингов аптек. Среднее значение длины неймингов с

высокими оценками в сферах взрослых поликлиник и санаториев ниже — 11,0 и 9,4 — соответственно.

В сферах, связанных с питанием, — продуктовые магазины, рестораны и кафе, видна обратная зависимость — короткие нейминги чаще имеют высокие оценки запоминаемости и благозвучия. В сфере аптек замечено повышенное использование заимствованных слов из латыни — каждое 5 название содержит транслитерацию, а в сфере санаториев и баз отдыха ниже среднего наличие ближайшего синонима — 3 нейминга из 70. В сфере санаториев также используются синонимы к слову природа — 52 нейминга из 70. На примерах видно, что нейминги строятся не только на основе синонимов — используются разные алгоритмы создания названия бренда.

Также нейминги с высокими оценками запоминаемости и благозвучия были рассмотрены на наличие синонимов. В сферах взрослых поликлиник, аптек и санаториев 72,2% рассматриваемых неймингов имеют близкие синонимы в названиях. В сфере питания всего 16,8% рассматриваемых неймингов имеют близкие синонимы в названиях.

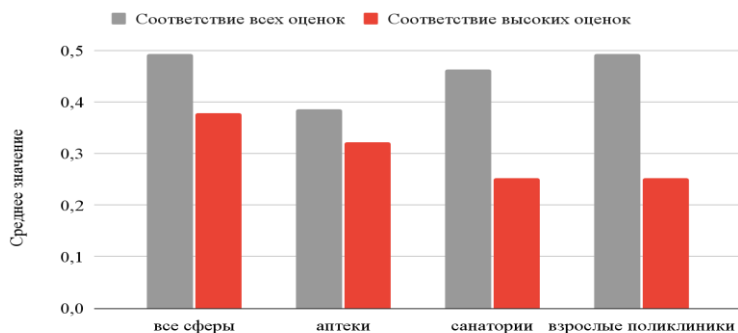


Рис. 2. Сравнение средних значений соответствий оценок запоминаемость и благозвучие во всех сферах, в сфере аптеки, санатории и взрослые поликлиники

Сфера продуктовых магазинов отличается от других сфер усреднёнными низкими показателями, которые демонстрируются на рисунке 2. Средняя длина нейминга магазинов продуктов наименьшая среди всех показателей — 7,8 символов. Среди неймингов с большинством оценок из 3 группы присутствуют лишь названия, состоящие из одного слова. По результатам опроса в сфере продуктовых магазинов большинство неймингов чаще имеют высокую оценку понятности и реже остальных имеют высокую оценку оригинальности.

5. Заключение

В результате эксперимента произведено сравнение значений оценки соответствия нейминга бренда и наличия синонимов в названии, также было проведено сравнение неймингов смежных сфер бизнеса. Рассмотрены основные алгоритмы составления неймингов, выявлены алгоритмы с низкой эффективностью по оценкам запоминаемости и соответствия. Замечено понижение оценок при удлинении названия бренда или внесении нескольких сокращений в сферах продажи продуктов, и повышение оценок при использовании ближайших синонимов в сферах здравоохранения и отдыха.

В дальнейших планах по развитию проекта — создание модели, проверяющей привлекательность и запоминаемость неймингов. Планируется увеличение размера корпуса с 910 слов минимум до 20 тысяч названий с выбранными оценками. Для создания модели также необходимо будет увеличить количество исследуемых сфер, расширить опросы, а также привлечь большее количество респондентов.

Литература

- [1] Александрова И. Ю. Проблема нарушения системной реализации психосемантических и психолингвистических требований к названию торговой марки // Вестник университета (Государственный университет управления). 2019. № 2. С. 150–156. DOI: 10.26425/1816-4277-2019-2-150-156.
- [2] Елистратов В. С., Пименов П. А. Нейминг: искусство называть. М.: Издательство «Омега-Л», 2014. С. 8–25.
- [3] RusVectores [сайт]. URL: <https://rusvectors.org/ru/> (дата обращения 11.06.2023).
- [4] Тимофеева М. К. Типология семантических отношений, выявляемых посредством инструмента RusVectores // Научный диалог. 2018. № 8. С. 74–87.
- [5] Radim Řehurek, models.word2vec — Word2vec embeddings // Gensim, 2019. — URL: <https://sysblok.ru/knowhow/word2vec-pokazhi-mne-svoj-kontekst-i-ja-skazhu-kto-ty/> (дата обращения 10.06.2023).
- [6] Radim Řehurek, Word2vec Tutorial, 2014 // Rare technologies — URL: <https://rare-technologies.com/word2vec-tutorial/> (дата обращения 10.06.2023).

Identification of Hidden Patterns in the Public Reaction to a Brand: Analysis of the Attractiveness of Names Using Machine Learning Methods

Ekaterina M. Tatur, Ekaterina V. Klimenko

Saint-Petersburg State University

The article deals with the problem of identifying hidden patterns in the reaction of society to brands by assessing the attractiveness of the name using machine learning methods. The authors analyze a large amount of data, including the names of various brands and information about their popularity among consumers, using mathematical methods and machine learning algorithms. In the course of the study, experiments were carried out with static predictive Word2Vec models. The results of the research prove that the approach developed by the authors allows finding names that meet the requirements of the marketing sphere.

Keywords: classification of texts, branding, naming

Reference for citation: Tatur E. M., Klimenko E. V. Identification of Hidden Patterns in the Public Reaction to a Brand: Analysis of the Attractiveness of Names Using Machine Learning Methods // Computational Linguistics and Computational Ontologies. Vol. 7 (Proceedings of the XXVI International Joint Scientific Conference «Internet and Modern Society», IMS-2023, St. Petersburg, June 26–28, 2023). — St. Petersburg: ITMO University, 2024. P. 60–66. DOI: 10.17586/2541-9781-2024-7-60-66

Reference

- [1] Aleksandrova I. Yu. Problema narusheniya sistemnoj realizacii psihosemanticheskikh i psiholingvisticheskikh trebovanij k nazvaniyu torgovoj marki // Vestnik universiteta (Gosudarstvennyj universitet upravleniya). 2019. № 2. Pp. 150–156. DOI:10.26425/1816-4277-2019-2-150-156. (in Russian)
- [2] Elistratov V. S., Pimenov P. A. Nejting: iskusstvo nazyvat'. M.: Izdatel'stvo «Omega-L», 2014. Pp. 8–25. (in Russian)
- [3] RusVectores. URL: <https://rusvectors.org/ru/> (access date: 10.06.2023).

- [4] Timofeeva M. K. Tipologiya semanticheskikh otnoshenij, vyyavlyaemyh posredstvom instrumenta RusVectors // Nauchnyj dialog. 2018. № 8. Pp. 74–87. (in Russian)
- [5] Radim Řehurek, models.word2vec — Word2vec embeddings // Gensim— 2019. — URL: <https://sysblok.ru/knowhow/word2vec-pokazhi-mne-svoj-kontekst-i-ja-skazhu-kto-ty/> (access date: 10.06.2023).
- [6] Radim Řehurek, Word2vec Tutorial, 2014 // Rare technologies — URL: <https://rare-technologies.com/word2vec-tutorial/> (access date: 10.06.2023).