

Возможности извлечения маркетинговой информации в e-commerce

Михаил Сливинский, Викимарт
mikhail.slivinskiy@wikimart.ru

21.10.2014, конференция “Большие данные в национальной экономике”

1. Доступность аудитории и нюансы масштабирования
2. Высокая цена ошибки
3. Невозможен push-маркетинг
4. Изучение потребностей покупателей
5. Персонализация и рекомендательные системы
6. Возможность измерений на всех этапах выбора, покупки, использования

Big data в маркетинге e-commerce проектов базируется на:

- накоплению большого объема данных
- извлечению неочевидных зависимостей

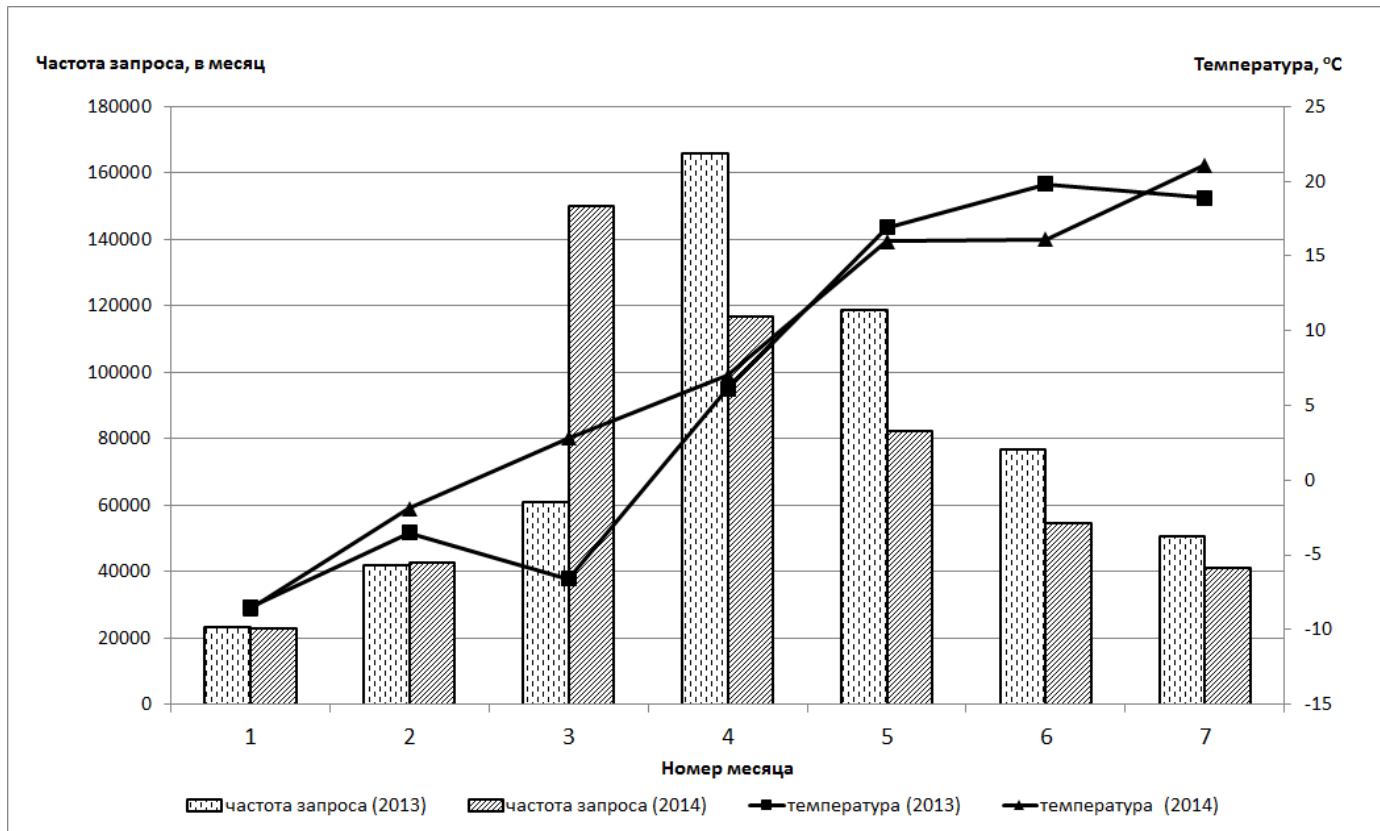
Особенности маркетинга online

Собственные источники:

- источники трафика, поисковые запросы и страницы входа на сайт
- события в сессии на сайте
- история транзакций
- ...

Внешние источники:

- статистика запросов к поисковым системам
- браузерные плагины (neiron, similarweb, alexa, ...)
- результаты поиска в поисковых системах
- корпуса отзывов
- ...



запрос [детские велосипеды],
Яндекс

Корреляция температуры воздуха и спроса

Традиционно рекомендательные системы обычно строят на коллаборативной фильтрации, т.е. принципе: **рекомендуем этому пользователю то, что интересно похожем на него пользователям.**

Основные недостатки:

- для обучения системы нужно много пользователей (= денег) и действий
- обучаясь “с нуля” на действиях пользователей, мы снижаем их лояльность

Нестандартные задачи:

- нужен универсальный механизм измерения сходства любых двух товаров (“**похожие товары**”)
- нужно построить хорошую универсальную товарную рекомендацию для пользователя, о котором ничего не известно (“**отранжированный список товаров в категории**”)

Рекомендательные системы

Из анализа сессий, начавшихся с просмотра конкретного товара, измеряем вероятность изменения первичной потребности (в бренде, свойстве, цене).

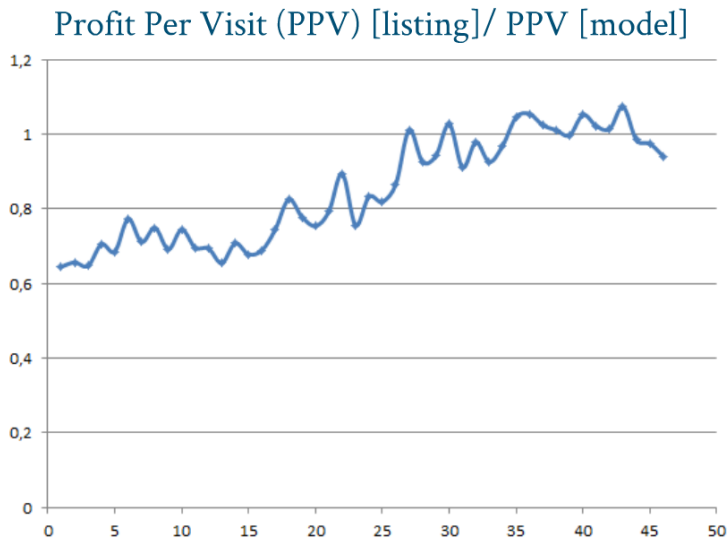
Если покупатель пришел за 40'' ЖК телевизором Samsung, интересуют ли его:

- 40'' телевизоры Panasonic?
- 36'' и 42'' телевизоры Samsung?
- на 7% более дорогая модель?

Теперь мы умеем отвечать на эти вопросы!

Ищем похожие товары

$$Rank = [Спрос]^{\alpha} * [Конверсия]^{\beta} * [Маржа]^{\gamma} * \dots$$



Нужно учесть:

- товар может быть новым и не иметь продаж
- необходимо честное тестирование эффективности алгоритмов

Результат:

+40% отнормированного PPV

Подробнее: “Покупатель — лучший эксперт. Ранжируем товары умно!”, Аркадий Итенберг, Михаил Сливинский, YaC/m 2013, <https://tech.yandex.ru/events/yac/m/talks/824/>

Ранжирование товаров

Устойчивые n-граммы из отзывов:

- **Функциональность:**
регулятор крепости кофе, долго держит заряд, лоток для овощей
- **Сочетаемость:** есть все разъемы, со всеми форматами, со старыми играми
- **«Коммуникабельность»:**
инструкция для сборки, на русском языке, интуитивно понятный интерфейс
- **Условия использования:**
на мокром льду, за МКАДом, при недостаточном освещении
- **Опыт использования:**
лежит в руке, приятный на ощупь, совсем не шумный, крепится к стеклу
- **Личное:** это мой первый, по сравнению с, до этого был, на мой взгляд
- **Экономическое:** за эти деньги, соотношение цена-качество
- **Эмоциональное:** я очень доволен, танцы с бубном, довольна как слон

Подробнее: "Лексическая статистика в оценке качества коммерческих текстов", Ирина Борисова, Wikimart.
<http://www.ashmanov.com/arc/aiconf2012/16-borisova-ai-conf2012.pdf>

Извлечение пользовательских предпочтений

Получение трафика из поисковых систем базируется на связках “запрос-документ”. Для крупного проекта речь идет о сотнях тысяч и миллионах поисковых запросов.

Очевидные источники запросов:

- статистика поисковых систем
- поисковые подсказки

Неочевидные источники поисковых запросов:

- сайты профильных конкурентов
- маркеры в поисковой выдаче

Возможные способа кластеризации запросов:

- через частотные словари
- через сходство поисковой выдачи



{ 439 тыс. активных пользователей
33 млн. запросов в месяц
4 терабайта хранилище }

Извлечение запросов и кластеризация

1	query	requests	results with key	total results	results with key/total results
29397	бокалы на свадьбу	12	218	240	90,83%
29815	бонбоньерки	10	143	250	57,20%
31225	букет невесты	56	1057	1470	71,90%
31279	букет невесты фото	12	198	240	82,50%
34070	вера вонг	17	181	360	50,28%
35298	вечерние платья на свадьбу	15	156	300	52,00%
36832	видеосъемка свадеб	32	453	760	59,61%
36841	видеосъемка свадеб ставрополь	39	270	480	56,25%
40610	все для свадьбы	23	980	1180	83,05%
47460	годовщина свадьбы по годам	23	283	460	61,52%
47475	годовщины свадеб	12	163	290	56,21%
47478	годовщины свадеб по годам	10	137	200	68,50%
47489	годовщины свадьбы названия	14	153	300	51,00%
47491	годовщины свадьбы по годам	13	147	260	56,54%
48522	горько.ру	18	266	360	73,89%
49517	греческие платья	13	168	320	52,50%
54966	диадема	13	148	281	52,67%
76543	какого размера рушник	32	341	640	53,28%

Извлечение запросов через маркеры

Критерий: $\frac{\text{популярность запроса со словом "отзывы"}}{\text{общая популярность запроса}}$

Отзывы:

↑
радионяни
эпиляторы
радар-детекторы
...
ноутбук
шлифовальная машина
...
стремянки
мангалы

Видео:

дровокол
интерактивные животные
йо-йо
...
сушилка для фруктов
стиральная машина
...
кроссовки
электрокотлы

Оценка востребованности типов контента

- ❑ Прогнозирование вероятности того, что пользователь сделает заказ через колл-центр
- ❑ Прогнозирование вероятности отказа пользователя от созданного заказа
- ❑ Прогнозирование вероятности ухода посетителя с сайта

Вопросы?

Михаил Сливинский, Викимарт



mikhail.slivinskiy@wikimart.ru
facebook.com/iseoexpert

Спасибо за внимание!