

Ю.Р. Валькман, д.т.н., зав. отд., А.В. Ланбина, аспирант, Е.В. Потапчук, аспирант, Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ РЕТАРГЕТИНГА ДЛЯ СБОРА И ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

This article describes the joint work retargeting and banner network to collect and Mining bottom of the users on the Internet. The advantages of joint operation of these systems.

Введение. Данная работа представляет собой краткое изложение использования метода сбора и интеллектуального анализа данных в компьютерных технологиях.

Объектом исследования является сбор и интеллектуальный анализ данных в компьютерных технологиях. Предмет исследования – метод ретаргетинга при сборе и интеллектуальном анализе данных в компьютерных технологиях. Цель исследования – Разработать модель использования ретаргетинга для сбора и интеллектуального анализа данных на примере украинской баннерообменной сети С8. Ожидаемые результаты – практическая модель использования метода ретаргетинга в работе украинской баннерообменной сети С8, выявление преимуществ и недостатков метода.

На сегодняшний день на рынке украинских провайдеров интернет рекламы существует проблема удержания устойчивого интереса пользователя к предлагаемым товарам. Одним из самых основных составляющих этой проблемы является неэффективный, низкотехнологичный, времязатратный сбор и анализ информации о потребителях. Компании – рекламодатели теряют десятки тысяч гривен на нецелевые показы, коэффициент конвертации показов в клики в среднем не превышает 0,3%, а это ведет к снижению спроса на интернет рынок. Сегодня спрос на интеллектуальные методы сбора и анализа информации о пользователях растет с каждым днем. Очевидно, что использование разработанной модели ретаргетинга в работе украинской баннерообменной сети С8 позволит компании удержать имеющихся клиентов и привлечь новых.

1. Понятие баннерной (баннерообменной) сети

Баннерная сеть — система обмена рекламными графическими баннерами или текстовыми блоками, при которой за показ чужих баннеров на своём сайте участник сети получает определённый процент показов собственного баннера на сайтах других участников проекта за вычетом процента комиссии, который использует владелец сети.

Другими словами, баннерная сеть – это единая мультисайтовая система

показа графического или текстового контента, в которой участвуют как владельцы рекламных площадей, так и клиенты, покупающие баннерпоказы.

Баннер — графический файл или текстовый блок, при нажатии на который происходит переход на рекламируемую им страницу.

В задачи баннерной сети входит следующее:

- ✓ Показ рекламных баннеров
- ✓ Ведение статистики просмотров баннеров
- ✓ Ведение статистики переходов (кликов) пользователями по баннерам
- ✓ Сбор и хранение информации о пользователях (куки), которые просмотрели или кликнули по баннерам
- ✓ Расчет соотношения количества кликов по баннеру к количеству показов (CTR), т.е. эффективность каждого баннера в отдельности
- ✓ Таргетинг показов – размещение рекламных баннеров с конкретизацией площадок, географии пользователей, дней недели, времени суток и количества уникальных показов в сутки.

Баннерообменная сеть С8, рассматриваемая в данной работе, обладает аналитическим комплексом функций, которые структурируют полученные данные о пользователях и показах в удобную и информативную базу данных. Аналитический комплекс С8 работает по принципу data mining, самостоятельно формируя новые классы и кластеры.

2. Понятие интеллектуального анализа данных

Интеллектуальный анализ данных (англ. *Data Mining*) — выявление скрытых закономерностей или взаимосвязей между переменными в больших массивах необработанных данных.

Английский термин «Data Mining» не имеет однозначного перевода на русский язык (добыча данных, вскрытие данных, информационная проходка, извлечение данных/информации) поэтому в большинстве случаев используется в оригинале. Наиболее удачным непрямым переводом считается термин «интеллектуальный анализ данных» (ИАД).[1]

ИАД включает методы и модели статистического анализа и машинного обучения, дистанцируясь от них в сторону *автоматического* анализа данных. Инструменты ИАД позволяют проводить анализ данных предметными специалистами (аналитиками), не владеющими соответствующими математическими знаниями.

Задачи, решаемые методами ИАД:

1. Классификация — отнесение входного вектора (объекта, события, наблюдения) к одному из заранее известных классов.
2. Кластеризация — разделение множества входных векторов на группы (кластеры) по степени «похожести» друг на друга.
3. Сокращение описания — для визуализации данных, лаконизма моделей, упрощения счета и интерпретации, сжатия объемов собираемой и хранимой информации.
4. Ассоциация — поиск повторяющихся образцов. Например, поиск

«устойчивых связей в корзине покупателя» (англ. *market basket analysis*) — вместе с пивом часто покупают орешки.

5. Прогнозирование

6. Анализ отклонений — Например, выявление нетипичной сетевой активности позволяет обнаружить вредоносные программы.

7. Визуализация

В интернет технологиях, где профиль посетителя веб ресурса либо скрыт совсем, либо данные профиля очень скудные, критерии классификации данных о поведении пользователя могут появляться в результате «обучения без учителя» самой системой сбора данных. К таким критериям можно отнести алогичные критерии:

- Цветовое предпочтение (любимые цвета пользователя)
- Ситуативное предпочтение (выбор в зависимости от контекста)

В процессе самообучения системы появляются новые кластеры, характерные для множества профилей пользователей.

Далее, для того, чтобы максимально раскрыть тему функционирования метода ретаргетинга, следует ввести понятие профиля пользователя, или cookies.

3. Понятие cookie

Кúки (слово не склоняется; от англ. *cookie* — печенье) — небольшой фрагмент данных, созданный веб-сервером и хранимый на компьютере пользователя в виде файла, который веб-клиент (обычно веб-браузер) каждый раз пересылает веб-серверу в HTTP-запросе при попытке открыть страницу соответствующего сайта. Применяется для сохранения данных на стороне пользователя, на практике обычно используется для:

- аутентификации пользователя;
- хранения персональных предпочтений и настроек пользователя;
- отслеживания состояния сессии доступа пользователя;
- ведения статистики о пользователях.

Приём куки обозревателями (браузерами) требуют многие сайты с ограничениями доступа, большинство интернет-магазинов. Настройка оформления и поведения многих веб-сайтов по индивидуальным предпочтениям пользователя тоже основана на куки.[2]

Куки используются веб-серверами для различения пользователей и хранения данных о них.

К примеру, если вход на сайт осуществляется при помощи куки, то после ввода пользователем своих данных на странице входа куки позволяют серверу запомнить, что пользователь уже идентифицирован и ему разрешён доступ к соответствующим услугам и операциям.

Многие сайты также используют куки для сохранения настроек пользователя. Эти настройки могут использоваться для персонализации, которая включает в себя выбор оформления и функциональности. Например, Википедия позволяет авторизованным пользователям выбрать дизайн сайта.

Поисковая система Google позволяет пользователям (в том числе и не зарегистрированным в ней) выбрать количество результатов поиска, отображаемых на одной странице.

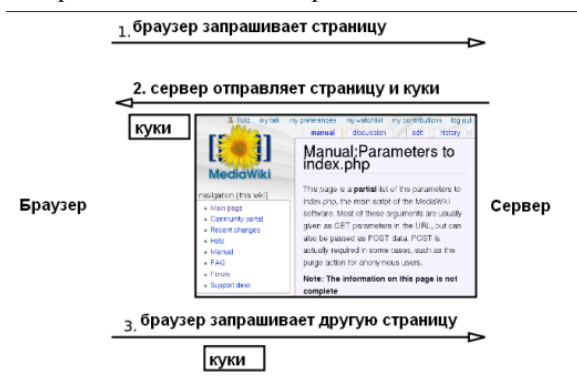


Рис.1. Возможное взаимодействие между браузером и сервером

Куки также используются для отслеживания действий пользователей на сайте.[3] Как правило, это делается с целью сбора статистики, а рекламные компании на основе такой статистики формируют анонимные профили пользователей, для более точного нацеливания рекламы.

Работа куки представлена на рис.1.

Многие авторы и исследователи считают, что благодаря появлению профиля пользователя интеллектуальный анализ данных в принципе стал возможен. Именно с помощью куки можно измерить и классифицировать пользователей веба.

Теперь для того, чтобы построить функциональную практическую модель ретаргетинга для баннерообменной сети, необходимо рассмотреть сам принцип метода ретаргетинга.

4. Понятие и метод работы ретаргетинга

Ретаргетинг (англ. Retargeting – перенацеливание)– это технология, которая предоставляет уникальную возможность персонализировать рекламное обращение к разным категориям аудитории сайта, перенацеливая пользователей на интересующую их информацию.

Пошаговая реализация метода ретаргетинга выглядит следующим образом:

1. Установка на веб-ресурс скрипта, который позволяет «запоминать» куки зашедших пользователей, а так же другую информацию о пользователе, доступную из личного профиля .
2. Наполнение базы данных куками пользователей и информацией о предпочтениях пользователей на основе их поведения.
3. Классификация пользователей согласно их предпочтениям и поведению (активности в сети).
4. Кластеризация пользователей согласно общим для больших групп признакам (поведение, предпочтение, географическое расположение и т.п.)
5. Формирование выборки ЦА (целевой аудитории) из кластеров для

каждой отдельно взятой рекламной кампании.

6. Сценарий рекламной кампании «следует» за пользователем, реализуясь на каждой площадке (сайте), куда бы не зашел пользователь.

В графическом виде ретаргетинг можно представить следующей схемой:

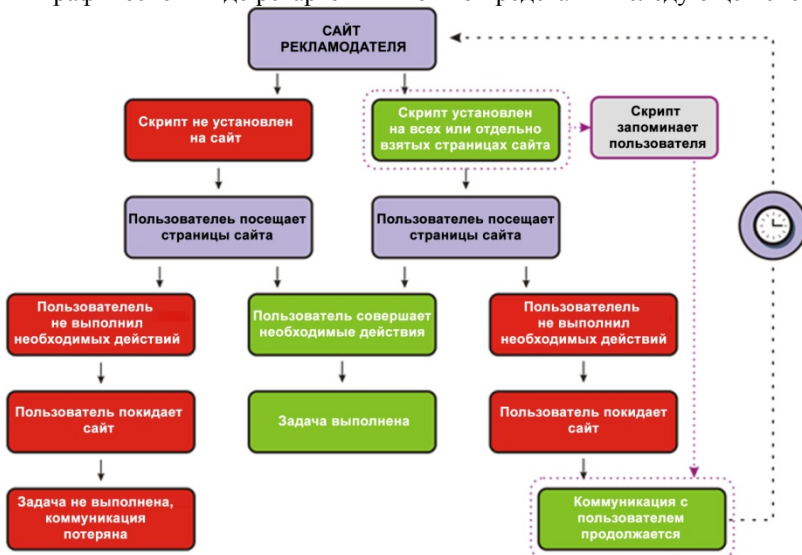


Рис.2 Схема работы ретаргетинга

Примером реализации ретаргетинга может послужить следующий план кампании:

- посетителям, которые еще не были на сайте, показываются баннеры, предлагающие посетить сайт — «приманки». Так как число таких посетителей велико, используется рекламная кампания с обычными баннерами, проходящая на площадках с относительно дешевым трафиком и большой аудиторией;
- посетителям, которые уже были на сайте, но не заинтересовались им (малая глубина просмотра, не прошли регистрацию и т.п.), показываются баннеры, которые сообщают о дополнительных бонусах для зарегистрированных пользователей, об обновлении сайта и т.п. Задача этих баннеров — побудить посетителя вернуться на сайт и потратить время на его более внимательное изучение. Такой баннер может вести непосредственно на страницу регистрации, новый раздел и т.д.;
- посетителям, которые давно не были на сайте, показываются «напоминающие» баннеры, которые могут содержать информацию о наиболее часто посещавшихся ими разделах;
- посетители, которые представляют наибольший интерес — например,

уже делавшие заказы на сайте, подвергаются воздействию наиболее дорогостоящих, но эффективных рекламных носителей.

Реальные сценарии могут быть более детализированы. Количество таких сценариев не ограничено и определяется целями и задачами рекламной кампании. Ретаргетинг значительно повышает степень воздействия рекламных сообщений на аудиторию, реализуя необходимые сценарии и активно мотивируя посетителей выполнять интересующие владельцев проекта действия.

Инновационный подход ретаргетинга заключается в том, что система сама учится формировать классификацию пользователей и создавать новые кластеры, динамично реагируя на изменения в поведении пользователей. Таким образом, у бизнес-аналитиков появляется возможность исследовать факторы, которые влияют на изменение поведения пользователей.

5. Использование ретаргетинга в баннерообменной сети

Для того, чтобы сформировать удобную базу посетителей площадок, участвующих в баннерообменной сети С8, необходимо использовать возможности ретаргетинга. Ретаргетинг дает возможность получить наиболее полную информацию о поведении пользователя в сети, т.к. он отслеживает пребывание каждого пользователя на всех площадках и формирует историю участия пользователя в рекламных кампаниях. [4] При этом сбор статистики включает в себя комплекс данных, доступных только при взаимодействии ретаргетинга и баннерообменной сети:

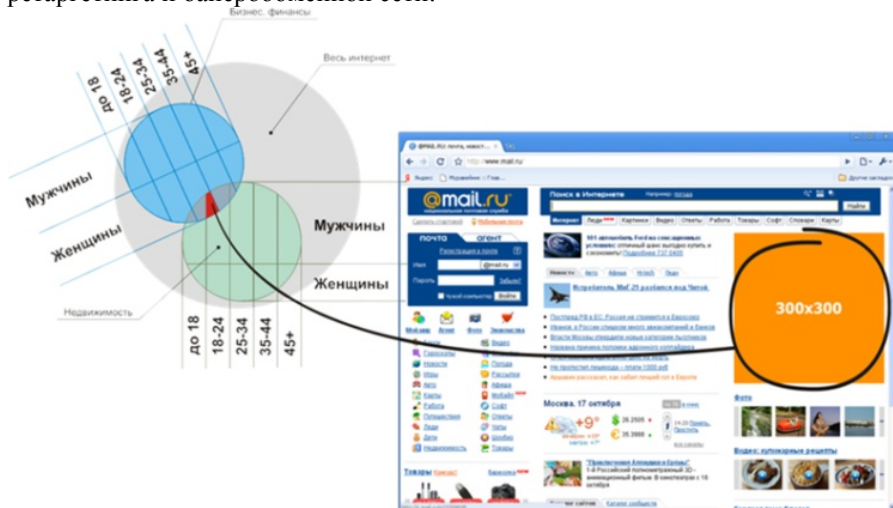


Рис.3 Отчет об активности пользователей на конкретном баннерном месте

1. На каких площадках пользователь наиболее часто кликает на медиа баннеры (предпочтения по контексту)
2. В какое время пользователь посещает ту или иную площадку (день

недели, время суток)

3. На каких площадках, участвующих в отдельно взятой рекламной кампании, пользователь проявляет наибольшую активность.
4. За какой период времени рекламная кампания достигает цели (через какое время пользователь совершает действие, назначенное как результат)

На основании полученных данных можно формировать аналитические отчеты по каждой отдельно взятой рекламной кампании (Рис.3), а так же по тенденциям поведения пользователей в целом.

Для баннерообменной сети C8, которая охватывает около 75% всех рекламных площадок Украины, 45% площадок России и 30% русскоязычных площадок Европы, модель ретаргетинга может быть представлена в таком виде:

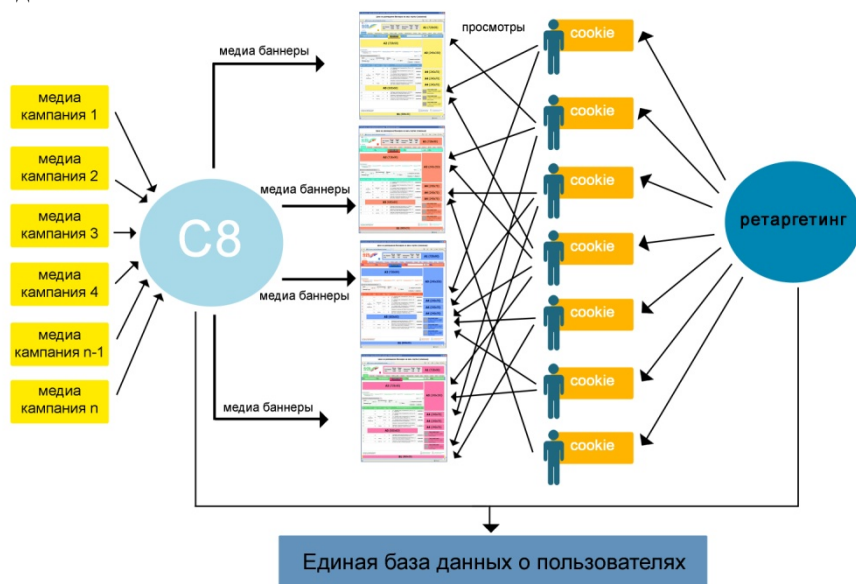


Рис.4 Модель ретаргетинга

При использовании куки от ретаргетинга и баннерообменной сети получаемая информация о предпочтениях пользователя более полная и точная. Также становятся доступными данные, напрямую не получаемые ни одной из этих систем, например информация о контексте, что можно видеть в таблице.

В настоящей работе мы рассмотрели методы сбора информации о потребителях. К преимуществам системы бесспорно относится то, что при интеграции систем сбора данных баннерообменной сети и куки на сайтах с помощью ретаргетинга существенно повышается скорость, точность, а также количество информации о предпочтениях пользователей, для принятия

решения о демонстрации баннерной медийной рекламы.

Информация	Баннерообменная сеть	ретаргетинг	Совокупная система
Показы	+		
Клики	+		
Географию	+	+	
Время суток, недели, месяца, года	+	+	
Профиль пользователя	+	+	
Частота контакта с площадкой	+	+	
Время сессии		+	
Глубина просмотра		+	
Сфера интересов		+	
Информация о контексте			+

К недостаткам системы относится потенциальная возможность кражи персональных данных пользователя (куки). Автоматическая выдача браузером профиля пользователя на запрос от сайта во многих странах запрещена законом (как например в Бельгии).

Заключение. Детальное рассмотрение работы ретаргетинга и баннерообменной сети С8 позволило выявить ряд полезных функций, получение которых невозможно при одиночном использовании каждой из систем. Информация о контексте, т.е. том информационном пространстве (тематика сайта), в котором пользователю показывается баннерная реклама, позволит единой системе ретаргетинг + баннерообменная сеть формировать абсолютно новые классы и кластеры пользователей. Основные признаки формирования новых классов и кластеров основываются на данных скорее психологического характера, нежели на простой статистике:

1. Коэффициент эффективности рекламной кампании:
 - ✓ по показу на наиболее близких по контексту площадках
 - ✓ по показу на площадках с внушительной посещаемостью
 - ✓ по показу в разное время года, месяца, недели, дня
 - ✓ по показу в событийном контексте
2. Предпочтения пользователей:
 - ✓ Сезонные
 - ✓ Событийные

Сегодня, благодаря такой эффективно развивающейся технологии, как ретаргетинг, системные аналитики и маркетологи получили возможность перейти от простого количественного анализа к качественному. Авторы статьи смогли убедиться в эффективной работе ретаргетинга в собственной профессиональной деятельности. Метод значительно облегчил и ускорил работу с большой базой информации о пользователях баннерообменной сети С8.

1. Журавлѐв Ю.И., Рязанов В.В., Сенько О.В. "РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения",— М.: Изд. «Фазис», 2006. — 176 с. — ISBN 5-7036-0106-8
2. Андрей Аликберов Что такое cookies и как с ними работать (1998)
3. Роджер Кларк Куки (англ.) (1 июня 1998)
4. О. Парфенова Современные маркетинговые технологии в Интернете (2009)

Поступила 1.03.2010р.

УДК 358.3.528

М.М.Степанов

ВИЗНАЧЕННЯ ОБСЯГІВ ПРИ РОБОТІ З ЕЛЕКТРОННИМИ КАРТАМИ

З урахуванням можливостей електронних засобів, за допомогою яких формуються електронні карти, пропонується новий метод виміру обсягів.

Одним із видів інформаційного забезпечення є геоінформаційне забезпечення, для збору, обробки та прийняття рішення з часом тільки зростає. Геоінформаційне забезпечення виступає в якості основи інформаційного середовища, що зв'язує в єдиний контекст різноманітні дані в інформаційних системах. Основою цього забезпечення є електронні карти місцевості, які використовують і для виробництва паперових карт та в зв'язку з цим мають ряд їх недоліків.

Аналіз літератури [1,2] показав, що існуючі методи за звичай ґрунтуються на приблизному розрахунку окремих ділянок, а наприкінці отриманні результати підсумовуються. При цьому, похибки можуть складати декілька десятків метрів. При возведенні дамби, при усуненні наслідків повені, враховуючи при цьому сучасний стан збройних сил та економічне становище в державі, це приведе до використання зайвої кількості паливо-мастильних речовин та одиниць техніки.

Тому метою статті є представлення нового підходу до визначення об'ємів, який би задовольнив потреби у точному відтворенні місцевості на електронних картах.

При роботі на електронних (і не тільки) картах задача визначення обсягів, наприклад обсягу води природних і штучних водойм, обсягу різних форм рельєфу суші, обсягів штучно створених споруд є важливою так само, як визначення відстаней і площ. Методи її рішення стосовно до неелектронних карт відомі [1]. В статті розглядається новий, у порівнянні з відомими, підхід, необхідність якого обумовлена наступними причинами.