Роль Big Data в Обеспечении Безопасности Детей в Интернете

Назакет Меликова
Институт Информационных Технологий НАНА, Баку, Азербайджан
naranara_68@mail.ru

Аннотация — **В данной статье мы рассматриваем виды рисков, ожидающих ребенка в Интернете, определение термина Big Data, а также роль Big Data в предотвращении рисков и обеспечении безопасности детей в Интернете.**

Ключевые слова — безопасность; Big Data; методика анализа больших данных; кибер-риски.

I. ВВЕДЕНИЕ

Много лет назад, когда Интернет только начал входить в жизнь простых людей, даже взрослые люди боялись начать осваивать эту, прежде неисследованную, территорию. Как оказалось, прогресс обогнал взрослых, и все чаще новое поколение помогает сориентироваться в Интернете своим родителям. Конечно, такое положение дел не может не радовать, так как естественно, наше будущее зависит от наших детей и поэтому поступающая к детям информация должна отфильтрована как по возрастным категориям, так и по моральным принципам. В связи с этим, в нынешнюю информационную эру крайне остро поставлен вопрос обеспечения безопасности детей в Интернете.

Число пользователей Интернета в Азербайджане стремительно растет, и стоит отметить, что доля лиц, в возрасте до 18 лет достаточна велика. Не удивительно, что Интернет является информационной средой для подростков и детей, без которой многие юные пользователи не представляют себе жизнь. Ведь в киберпространстве можно найти единомышленников и обсудить с ними насущные вопросы, отыскать нужную информацию для учебных целей, послушать любимого исполнителя или приобрести редкую книгу. Таким образом, Интернет помогает детям как по учебе, так и в организации досуга. Но, кроме этого существуют многочисленные социальные сети, которые сами по себе затягивают ребенка. Онлайн-игры, а также огромное количество «вредных» сайтов, из которых дети получают информацию, совершенно для них не предназначенную, а иногда и опасную для психического, или даже физического здоровья. Выявить из такого большого объема не структурированных данных безопасную для детей информацию поможет технология Big Data.

Во всем мире технология Big Data успешно применяется. Внедрение Big Data во многих областях деятельности обещает увеличение производительности, сокращение расходов и т.д. Однако, чаще всего из виду упускается то, что сам по себе сбор большого количества исходных данных не является полезным или выгодным. Перспектива возможности использования данных есть статистические зависимости, которые можно почерпнуть из их анализа. Это действительно имеет значение [1]. Перспектива того, как вы можете использовать данные — и статистические зависимости, которые можно почерпнуть из их анализа, — вот что действительно имеет значение [1]. В Азербайджане разработка и применения технологий анализа данных началась недавно, причем данные технологии направлены в основном на бизнес стратегии, а не на обеспечение безопасности пользователей, и тем более детей. Однако, мы считаем, что применение Big Data в сфере обеспечения безопасности детей принесет большую пользу как детям, так и нашему государству, и это именно то направление, в котором использование Big Data будет рентабельно, целесообразно и продуктивно.

II. ВИДЫ ИНТЕРНЕТ- РИСКОВ

Статистика показывает, что каждую минуту в Интернете появляется много новых сайтов, на онлайн-шоппинг пользователи тратят $272 070, сервисы электронной почты отправляют 204 млн писем и т.д. [2]. Цифры действительно впечатляют. И таким образом наши дети ежеминутно подвергаются опасности в киберпространстве.

Итак, как же нам защитить своего ребенка в этой полезной, но, не редко опасной среде? Сперва разберем, какие виды риска поджидают нас в Интернет-среде.
Среди предполагаемых опасностей киберпространства можно выделить 4 основные группы: [3]

- ***контентные риски*** (сведения, включающие в себя элементы эротики, насилия, агрессии, ненормативной лексики; данные, зарождающие межрасовую вражду, пропагандирующие наркотики, суицид и прочие материалы, способные спровоцировать причинение вреда здоровью);
- ***коммуникационные риски*** (уголовно наказуемые контакты, а именно, груминг- общение взрослых и установление дружеских отношений с целью изнасилования, киберпреследования (или кибербуллинг)- домогательство, оскорбления по средством личных сообщений и др.);

- ***электронные или кибер-риски*** (хищение персональных данных, спам и вирусная атака (при помощи таких вредоносных или шпионских программ, как черви, вирусы и «троянские кони»), кибер-мошенничество, и т.д.);
- ***потребительские риски*** (приобретение недоброкачественного товара и незапланированные финансовые потери, злоупотребление правами потребителя, различные поделки продукции, продаваемой посредством Интернета, фальсификация изделий, хищение персональной информации с целью онлайн-мошенничества, агрессивная реклама и др.).

Столкновение с любым из этих рисков может тяжело повлиять на психическое или физическое состояние ребенка.

III. МЕТОДИКИ АНАЛИЗА BIG DATA

Как известно, Big Data- это совокупность технологий, которые должны быстро и результативно обрабатывать большие объемы структурированных и плохо структурированных данных параллельно в разных аспектах, причём, этих данных не просто много, их объемы постоянно растут.

Методики анализа Big Data довольно разнообразны, причем исследователи продолжают трудиться над разработкой современных методик и усовершенствованием имеющихся.

Наиболее широко распространены следующие виды методики: [4]

1) Поочередное сравнение контрольной выборки с другими, также широко применяется при работе с данными меньших объемов (A/B testing);
2) выявление взаимосвязей (Association rule learning);
3) классификация на основе выявления ранее не заданных признаков (Cluster analysis);
4) сбор данных из большого количества источников (Crowdsourcing);
5) обнаружение данных в ранее не рассматриваемых для данной цели знаниях, используемых для принятия решения (Data mining, в нем применяются следующие методы: association rule learning, classification, cluster analysis, regression);
6) использование множества предикативных моделей для повышения точности прогнозов (Ensemble learning);
7) представление решений в виде так называемых ‘хромосом`, которые аналогично могут комбинироваться и мутировать; за основу принимается решение, которое наиболее «приспособлено» (Genetic algorithms);
8) генерирование self-learning алгоритмов на основе анализа и оценки эмпирических данных (Machine learning);
9) совокупность присущих информатики и лингвистики методик, направленных на распознавание естественного человеческого языка (Natural language processing);
10) анализ и выявление взаимосвязи узлов в сетях (Network analysis);
11) Набор численных методов для оптимизации систем и улучшения результата, стремящегося к идеальному (Optimization);
12) совокупность заданных шаблонов с элементами обучения «без учителя» для предсказания поведенческой модели пользователей (Pattern recognition);
13) математическая модель предсказания событий (Predictive modeling);
14) статистические методы для анализа влияния одной или нескольких независимых переменных на зависимую, причем соответствующая зависимость математическая (Regression);
15) метод выявления нужной информации из общего потока, основанный на распознавании естественного человеческого языка (Sentiment analysis);
16) отчасти заимствованный набор методик анализа пространственных данных (Spatial analysis);
17) наука о сборе, организации и интерпретации данных, включая разработку опросников и проведение экспериментов (Statistics);
18) выявление функциональной взаимосвязи на основе машинного обучения (Supervised learning);
19) моделирование поведения сложных систем часто используется для прогнозирования, предсказания и проработки различных сценариев при планировании (Simulation);
20) анализ повторяющихся с течением времени последовательностей данных (Time series analysis);
21) совокупность методик, схожих с Cluster Analysis и основанных на Machine Learning, позволяющих выявить скрытые функциональные взаимосвязи (Unsupervised learning).

IV. BIG DATA В БОРЬБЕ С ИНТЕРНЕТ- РИСКАМИ

A. Контентные риски

Как мы видим, технология Big Data –это революционное решение в области анализа и обработки данных, и эффективно применяется во многих областях, например, в сфере Интернет-рекламы. С таким же успехом

мы можем также использовать данную технологию для обеспечения безопасности детей.

Анализ запросов поисковой системы и правильно построенный алгоритм могут помочь в этом деле. Существующие ныне детские браузеры слишком ограничены, так как сортировка ссылок для поисковой системы происходит вручную. Это - своего рода "сокращенный вариант" Интернета, куда вошли лишь те ресурсы, которые составители проекта посчитали безопасными для детей. Например, в браузере «Гогуль», направленном на русскоязычную аудиторию, по умолчанию доступ разрешён более чем к 7000 развивающих и обучающих ресурсов на русском языке, отобранных и одобренных детскими психологами и педагогами [5].

Естественно, в Интернет-пространстве в разы больше полезной для детей информации, и применение технологии BigData значительно разнообразило бы досуг детей, пользующихся детским браузером, и повысило безопасность тех несовершеннолетних пользователей, которые игнорируют детские поисковые системы.

B. Коммуникационные риски

Коммуникационные риски также можно предотвратить, проанализировав в социальных сетях, онлайн-мессенджерах, чатах, форумах входящие или отправляемые ребенком в личной переписке данные на предмет угрозы со стороны взрослых или сверстников. Big Data аналитика позволяет вычислить из потока данных те самые доли процента информации, необходимых для принятия решение об опасности данной переписки.

Если с помощью технологий Big Data возможно проконтролировать социальные настроения, выявить информацию о готовящихся терактах, волнениях, то почему же не направить анализ данных на обеспечение безопасности детей? На основе уже известных алгоритмов, с помощью машинного обучения можно отследить тематику разговора и своевременно определить попытки кибер-буллинга или груминга и при помощи правоохранительных органов предотвратить возможное преступление и уголовно наказать виновных.

Также, анализ поисковых запросов помогает сформировать понимание о направлении работы, круге интересов. Например, Google имеет возможность сохранять поисковые запросы в привязке к аккаунту пользователя [6]. Предположительно, это помогает выявить социально- опасного человека. И близкое общение ребенка с ним может насторожить.

C. Электронные риски

Возможность столкновения с хищением, незаконным использованием личной информации, заражением компьютера вирусом, спам-атакой называется кибер-риском. Произойти это может при помощи распространения и проникновения вредоносного программного обеспечения, не только посредством электронных носителей, но и через Интернет, при принятии спама или закачки вредоносных файлов.

Справится с этим также поможет технология анализа Big Data. Сегодня ни одна крупная компания не обходится без технологии Big Data для защиты компьютеров при разработке программного обеспечения. Но, с каждым днем появляются новые, способные проникнуть в сеть или компьютер с большим успехом вредоносные программы, количество, скорость и сложность производимых атак возрастает, а необходимый для их обнаружения объем вычислений уже невозможно выполнять на устройствах пользователя своевременно.

Облачные сервисы, применяемые в большинстве антивирусов, способны объединить данные от разных пользователей, и затем, проанализировав, сопоставить их между собой. Таким образом можно выявить источники распространения вредоносных программ, потенциально опасные сайты. В связи с этим, для защиты от электронных рисков, огромные усилия сосредоточены на разработке и усовершенствовании облачных решений (таких как Kaspersky Security Network, Dr.WebCloud и т.д.), основанных на технологии анализа BigData.

D. Потребительские риски

В киберпростанстве широко распространено мошенничество, являющиеся одним из самых распространённых видов потребительских рисков. Чаще всего кибермошенничество осуществляется с помощью различных технических средств и разнообразных программ, следовательно, некоторые его виды могут быть отнесены к группе электронных рисков. Также потребительские риски плотно связаны с коммуникационными, поскольку установление более близкого контакта с жертвой с помощью личных сообщений- один из залогов успеха реализации преступной схемы. По словам руководителя аналитическим подразделением Microsoft по борьбе с преступлениями в сфере высоких технологий DigitalCrimesUnit (DCU) Брайана Хёрда, единственный способ противостоять росту киберпреступности — развивать технологии анализа Big Data [7].

Мошенничество посредством Интернета принимает чудовищные масштабы, зачастую обкрадывают целые страны, не говоря уже об отдельных пользователях, и тем более доверчивых детях. Противостоять кибермошенничеству могут только столь же масштабные системы, основой которых является технология Big Data, т.к. используя очередную схему, мошенники оставляют цифровые следы. И если взятые по отдельности эти изменения ничтожны, то в совокупности, при помощи Big Data, можно выявить характерную структуру преступной деятельности.

ЗАКЛЮЧЕНИЕ

Таким образом, полное обеспечение безопасности детей в Интернете посредством ныне существующих методик анализа Big Data вполне реально. Эффективным

решением для реализации является внедрение известных на сегодняшний день алгоритмов в данный процесс и адаптация их для борьбы с Интернет - рисками. Для каждой группы рисков следуют по отдельности разработать самостоятельную схему защиты, применить ее исключительно на уровне определенной группы рисков, и только после успешного завершения данного этапа, использовать Big Data комплексно для защиты от всех рисков.

ЛИТЕРАТУРА

[1] Д. Райдер, Революционный потенциал больших данных, 2015, http://22century.ru/docs/big-data.

[2] İngate, 75 фактов об интернет-маркетинге, 2012, http://www.slideshare.net/zhomart/75-ingate, стр.7.

[3] Г. Солдатова, В. Серегина, П. Волкова, «Неотложка» в киберпространстве // Журнал Дети в информационном обществе, 2011, стр.56-62.

[4] М.А. Назаренко Технологии и методы анализа больших данных // Международный журнал экспериментального образования, 2015, № 11, стр. 40-41.

[5] Новое Поколение, Федеральная программа безопасного детского интернета «Гогуль», 2009, http://gogul.tv/press/5/.

[6] Google, Политика конфиденциальности, 2015, https://www.google.com/intl/ru/policies/privacy/

[7] А. Васильков. Microsoft :”большие данные “ключ к борьбе с преступностью, 2014, http://www.computerra.ru/96630/big-data-is-key-technology-for-digital-crime-investigations/.