Литература

1. Башарин Г.П. Анализ очередей в вычислительных сетях. Теория и методы расчета / Г.П. Башарин, П.П. Бочаров, Я.А. Коган – М.: Наука. Гл. ред. физ.-мат. лит., 1989. - 336 с.
2. Шелухин О.И. Фрактальные процессы в телекоммуникациях : учеб, пособие / О.И. Шелухин, А.М. Тенякшев, А.В. Осин ; Под ред. Шелухина О.И. М.: Радиотехника, 2003. – 479с.
3. Абрамовиц, М. Справочник по специальным функциям с формулами, графиками и математическими таблицами : пер. с англ. - М.: Наука, 1979. – 835с.
4. Бочаров П.П Теория массового обслуживания : учеб, пособие / П.П Бочаров, А.В. Печинкин М.: Изд-во РУДН, 1995. – 529с.

У статті розглянуто проблему спаму в соціальних мережах, класифікація легітимних і нелегітимних користувачів

Ключові слова: соціальні мережі, класифікація, спам, алгоритм

В статье рассмотрена проблема спама в социальных сетях, классификация легитимных и нелегитимных пользователей

Ключевые слова: социальные сети, классификация, спам, алгоритм

The problem of spam in social networks, legitimate and non-legitimate users' classification is considered in this article

Key words: social networks, classification, spam, algorithm

УДК 001.891:65.011.56

ПОДХОД К КЛАССИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ В СОЦИАЛЬНЫХ СЕТЯХ

А.А. Куликова
Кафедра искусственного интеллекта
Харьковский национальный университет радиоэлектроники
пр. Ленина,14, г. Харьков, Украина, 61166
Контактный тел.: (057) 337-27-53, 093-776-41-20
E-mail: ganna.kulikova@gmail.com

1. Введение

Социальные сети, привлекающие сегодня к себе всеобщее внимание пользователей Интернета, сформировались за очень короткий промежуток времени. Они объединяют в себе блоги (сетевые дневники), сети медиа-ресурсов, сети персональной информации (MySpace, LinkedIn, Facebook, Вконтакте), системы закладок (del.icio.us), wiki-энциклопедии и другие. Данные Web-сайты представляют собой автоматизированную социальную среду для обеспечения коммуникации как отдельных, так и групп пользователей, объединенных общими интересами. Количество пользователей в этих сетях увеличивается с беспрецедентной скоростью, вызывая интерес у представителей науки, бизнеса и IT-индустрии [1]. Такие Web-сайты фактически представляют собой большое хранилище общедоступной информации, в первую очередь, персонального характера.

Однако в то же время развитие Internet, технологий проектирования социальных сетей привело к тому, что одной из основных проблем пользователей стал избыток информации, в том числе и незапрошенной, – спама.

Спам представляет собой масштабную рассылку коммерческой, политической и иной рекламы (информации) или иного вида сообщений лицам, не выражавшим желания их получать. Значительная часть атак основана на методах социальной инженерии (привлечение пользователей недобросовестной рекламой и т.д.), другая – на использовании уязвимостей в механизмах работы социальных сетей. Существует достаточно много видов спама, распространяемого в социальных сетях, но, прежде всего, стоит отметить рекламу, некоторые виды мошенничества, фишинг, распространение вредоносного программного обеспечения.

Технологии рассылки спама в социальных сетях совершенствуются: спаммеры отмечают пользователей социальной сети на фотографиях, видеозаписях, добавляются в друзья, приглашают в группы и так далее, в целом используют все возможности социальной сети в корыстных целях.

Борьба со спаммерами в социальных сетях важна для улучшения сервисов, предоставляемых социальной сетью для участников, уменьшения количества нежелательного и опасного контента, а так же для развития самих социальных сетей.

2. Анти-спам стратегии

Существует несколько анти-спам стратегий в социальных сетях: основанные на идентификации (identification-based), основанные на ранжировании (rank-based) и основанные на интерфейсе и ограничениях (interface-, limit-based), но наиболее эффективные результаты показывает сочетание всех стратегий [2].

Причиной же массовых рассылок спама является, прежде всего, наличие спаммеров и спам-ботов в социальных сетях, поэтому их качественная идентификация является залогом успеха защиты систем, базирующаяся на анализе профилей пользователей, социального поведения, а так же контента, сгенерированного самими пользователями [3]. Эти атрибуты позволяют определить характеристики, которые присущи обычному пользовательскому поведению и, следовательно, могут улучшить распознавание вредоносных спам-ботов и спаммеров. В зависимости от типа социальной сети и возможностям, предоставляемым пользователям, атрибуты могут различаться, качественная идентификация должна учитывать все признаки, которые потенциально могут быть важны для классификации пользователей.

3. Постановка задачи

Задача классификации представляет собой задачу отнесения образца к одному из нескольких попарно не пересекающихся множеств.

Пусть X – множество описаний объектов, Y – множество номеров (наименований) классов.

$$y^*: X \to Y.$$

Существует неизвестная целевая зависимость – отображение, значения которой известны только на объектах конечной обучающей выборки X^m, по которой производится настройка (оптимизация параметров) модели.

$$X^m = \{(x_1, y_1),...,(x_m, y_m)\}.$$

Требуется построить алгоритм a, способный классифицировать произвольный объект $x \in X$.

$a: X \to Y$. Пусть дан набор k пользователей U.

$$U = \{u_1, u_2,..., u_k\}.$$

Каждый пользователь u_i имеет профиль p_i и связанный с ним контент и активность пользователя в социальной сети [4]. Профиль представляет собой страницу, которая создается и контролируется пользователем. Каждое сообщество так же имеет свой профиль, но во многом они совпадают.

Задача заключается в определении того, что пользователь u_i является спаммером или легитимным пользователем с помощью классификатора c, когда имеется p_i.

$$c: u_i \to \{\text{spammer}, \text{legitimate\_user}\}.$$

Классификатор с прогнозирует, является ли пользователь ui спаммером. Для построения с необходимо выделить набор m особенностей, характерных признаков F из U, с помощью которых будет производиться классификация.

$$F = \{f_1, f_2,..., f_m\}.$$

Для построения классификатора необходимо определить, какие параметры влияют на принятие решения о том, к какому классу принадлежит образец. Исходные данные обязательно должны быть непротиворечивы. Для этого необходимо иметь достаточно большую размерность пространства признаков (количество компонент входного вектора, соответствующего образцу).

Точность классификатора определяется с помощью тестовой выборки. Оценку качества, сделанную по тестовой выборке, можно применить для выбора наилучшей модели или дальнейшей ее модификации.

4. Методы предварительной обработки профилей

Для проведения классификации необходима предварительная обработка профилей пользователей, которая в первую очередь включает в себя непосредственное выделение информации, признаков и атрибутов с Web-страницы учетной записи [5].

Модель «Множество слов» (Bag of words) – упрощающее предположение, использующееся в обработке естественного языка и поиске информации (information retrieval). В этой модели, текст (предложение или документ) представляется как неупорядоченный набор слов, без учета грамматики и порядка слов частности. Для этого необходимо убрать HTML-тэги, знаки препинания, стоп-слова, все слова перевести в нижний регистр [6].

Стоп-словами являются слова, не несущие самостоятельной смысловой нагрузки. Как правило, к ним относятся предлоги, союзы, частицы, местоимения, вводные слова, междометия, предикативы.

Далее, для выделенной из профиля пользователя информации, необходим стемминг – процесс нахождения основы слова для заданного исходного слова.

Одним из наиболее популярных и эффективных алгоритмов стемминга является стеммер Портера, опубликованный Мартином Портером в 1980 году. Основное достоинство стеммера Портера заключается в том, что не используются словари, и выделение основы осуществляется путем преобразования слова согласно определенным правилам [7]. Алгоритм не использует баз основ слов, а лишь, применяя последовательно ряд правил, отсекает окончания и суффиксы, основываясь на особенностях языка, в связи с чем работает быстро, но не всегда безошибочно.

Обработанные таким образом профили полностью удовлетворяют требованиям дальнейшей классификации.

5. Разработка общего алгоритма классификации

В результате анализа было выделено два типа данных, которые содержатся в профиле пользователя

социальной сети: категориальные и свободно заполняемые (free-form data).

Категориальными являются поля, которые могут принимать только ограниченное число значений, например, пол, возраст, семейное положение.

Данные, которые вводятся непосредственно пользователями, в большинстве случаев – текстовые: «Обо мне», «Интересы», – свободно заполняемые.

Это различие позволяет использовать соответствующие алгоритмы машинного обучения, которые так же могут зависеть от контекста, предоставляемого категориальными полями.

Однако, в зависимости от задачи и контекста, некоторые атрибуты могут быть не релевантными. Если классификаторы был обучены в нескольких контекстах, то все обнаруженные условные зависимости, усредненные по всем условиям, не могут гарантировать высокую точность классификации применимо к конкретному случаю.

Контекст в данном случае представляется как некоторый набор атрибутов, которые не являются значимыми для классификации непосредственно, но влияющие на релевантность заведомо более значимых атрибутов, имеющих большую силу дискриминации [8].

Таким образом, для проведения классификации были выделены 3 типа атрибутов.

1. Целевые атрибуты. Это атрибуты, принадлежность к которым вычисляется.

2. Прогнозирующие атрибуты (predictive attributes). Это атрибуты, значения которых наблюдаются и которые влияют на принадлежность пользователя определенному классу. Дальнейшее обучение классификатора и сама классификация будет производиться именно на основе прогнозирующих атрибутов.

3. Контекстные атрибуты (contextual attributes). Это атрибуты, которые не имеют четко выраженный видимый эффект на целевые атрибуты, но влияют на релевантность интеллектуальных атрибутов. Контекстный атрибут может условно зависеть от других контекстных атрибутов. По контекстным атрибутам выбирается соответствующий классификатор.

К наиболее часто используемым и универсальным контекстным атрибутам относятся пол, возраст, семейное положение, образование, географическое положение: такие атрибуты пользователей наиболее распространены в социальных сетях разного типа.

Атрибут возраст может классифицироваться согласно возрастной группе, к которой принадлежит пользователь: ребенок, подросток, юноша и т.д. Семейное положение имеет такие значения, как женат/замужем (married), в отношениях(in a relationship), не в отношениях (single). Образование же может быть средним, профессиональным, высшим, пользователь может иметь научное звание и научную степень.

Количество классов и, соответственно, значения контекстных атрибутов могут варьироваться в зависимости от типа и возможностей социальной сети, а так же от типа поставленной задачи.

В свою очередь, к прогнозирующим атрибутам могут относиться такие признаки, как информация о себе, интересы, количество друзей, участие в группах, число отправленных сообщений за определенный промежуток времени (1 день), число заявок, вступлений в группы за определенный промежуток времени (1 день) и т.д.

Общий подход к классификации пользователей на спаммеров и легитимных пользователей и обучение классификаторов проходить по следующему алгоритму.

1. Обучающая выборка TrainingSet разбивается на n классов искусственно по выбранным экспертами контекстным атрибутам, характерным для рассматриваемой социальной сети.

$$K = \{k_1 ... k_n\}.$$

K – множество контекстных атрибутов.

$k_1 ... k_n$ – классы контекстных атрибутов пользователей социальной сети.

Каждый класс контекстных атрибутов k_i так же разбивается на m подклассов атрибутов, выбранных экспертом для определенной социальной сети.

$$K = \{a_1 ... a_m\}.$$

$a_1 ... a_m$ – подклассы классов контекстных атрибутов.

2. По количеству подклассов m, на которые разбиты контекстные атрибуты, строится гиперкуб – многомерный массив данных.

Таким образом, выборка профилей TrainingSet разбивается на M подмножеств.

$$M = \prod_{i=1}^{m} a_i.$$

3. Для каждого из сформированных подмножеств множества M обучающей выборки TS строится и обучается свой классификатор c_i, его эффективность проверяется на тестовой выборке TestSet с помощью рассмотренных спам-метрик. Классификатор обучается только на прогнозирующих атрибутах, контекстные атрибуты в данном случае уже не учитываются.

Последующая классификация новых профилей будет проходить следующим образом.

1. Согласно контекстной информации в профиле p пользователя u выбирается один из классификаторов c_i.

2. Для непосредственной классификации будут использоваться только прогнозирующие атрибуты.

3. Профили, отнесенные к классу спаммеров, могут так же инспектироваться модераторами.

Таким образом, целесообразной является классификация с помощью нескольких классификаторов. Базовые классификаторы могут выбираться в зависимости от типа социальной сети, возможностям и сервисам, предоставляемым пользователям.

Использование комбинации классификаторов позволяет повысить точность классификации при решении практических задач. Теоретические и эмпирические результаты показывают, что результат комбинации классификаторов наиболее эффективен, когда классификаторы являются независимыми. Для построения независимых классификаторов наиболее эффективным методом является обучение отдельных членов на различающихся подмножествах признаков:

построение набора классификаторов на основе декомпозиции исходного набора признаков, описывающих объекты данных, в большинстве случаев имеет преимущества.

6. Разработка архитектуры системы детектирования нелегитимных пользователей

Для проведения классификации пользователей различных социальных сетей на легитимных и нелегитимных была разработана архитектура системы, представленная на рис. 1, состоящая из слоев, каждый из которых выполняет соответствующую функцию.

Нижним слоем являются некоторые социальные сети, для которых требуется провести классификацию, независимо от выполняемых ими функций, количества пользователей и т.д. Атрибуты, необходимые для непосредственной классификации, или информация, необходимая для вычисления определенных атрибутов, выделяется из многопользовательской системы с помощью API, доступного для многих популярных социальных сетей, таких как Twitter, Facebook, Vkontakte и т.д.

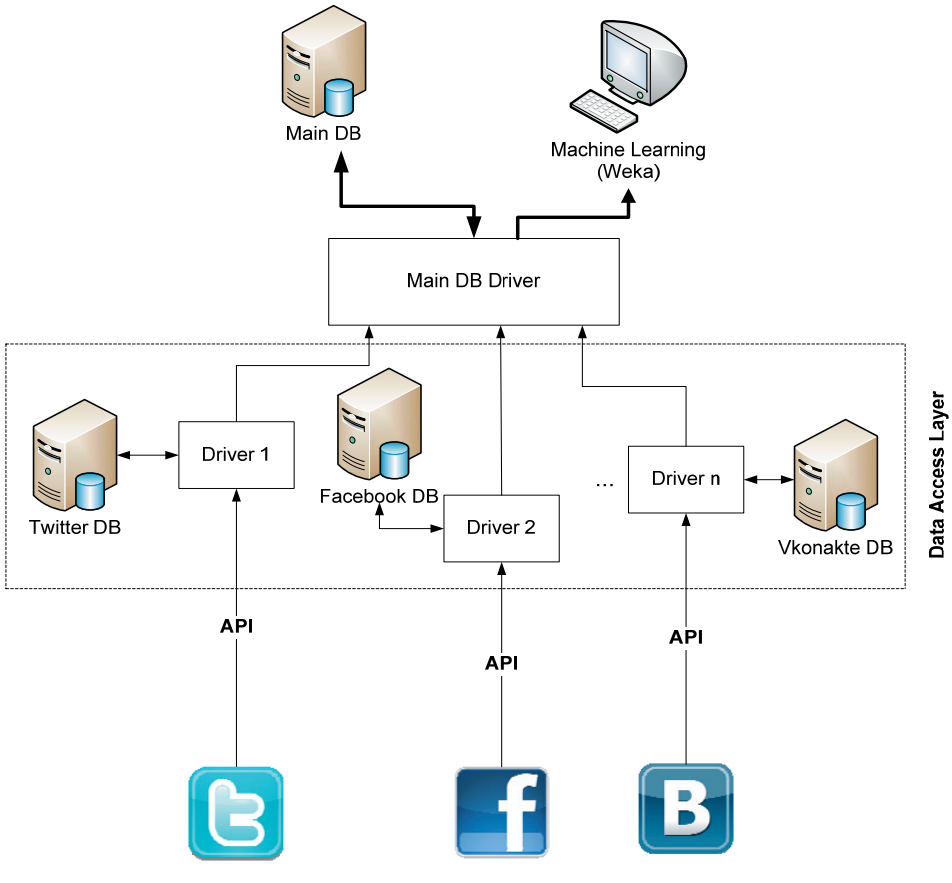

Рис. 1. Архитектура системы классификации

Для обработки и сохранения данных необходим Data Access Layer, состоящий из баз данных, разработанных для каждой социальной сети, и набора программ-драйверов, которые производят непосредственную обработку данных и приведение их к единому виду.

Драйвер(Driver) разрабатывается для конкретной базы данных, содержит функции манипуляции с БД самым быстрым способом. Программа запрашивает и получает только те данные, которые запрошены: транслируются уже отобранные данные. Для каждой социальной сети определенный драйвер взаимодействует с соответствующим API, осуществляя передачу, сохранение и обновление данных в соответствующей базе данных.

Данные из всех социальных сетей, для которых необходима классификация, обрабатываются следующим слоем – Main DB Driver. Этот слой, состоящий из нескольких программ, позволяет обработать данные, полученные из социальных сетей: вычислить некоторые атрибуты, привести к виду, необходимому для хранения в главной базе данных, и непосредственно осуществить передачу данных в главную базу данных.

Main Database хранит информацию обо всех пользователях социальных сетей, полученную на нижних уровнях, приведенную к единому формату, позволяющему осуществлять классификацию, анализ, дискретизацию и другие действия над данными, которые могут быть необходимы. Все атрибуты пользователей должны иметь числовой формат, что позволяет использовать различные алгоритмы классификации с высокой точностью, так как существует достаточно много способов манипулирования текстом, применяемых спаммерами: использование разных языков, шрифтов, символов, пробелов между символами и т.д. Такие манипуляции производятся с целью уменьшить эффективность классификации.

Важной частью является блок ML(Machine learning) – блок, в котором представлены различные алгоритмы машинного обучения, дискретизации и т.д., которые могут работать с любыми данными из главной базы данных: для одного и того же набора данных могут применяться различные алгоритмы классификации, можно оценить их точность и проанализировать результаты. В качестве такого блока можно использовать Weka (Waikato Environment for Knowledge Analysis) – свободное программное обеспечение для анализа данных, написанное

на Java в университете Уайкато(Новая Зеландия), распространяющееся по лицензии GNU GPL. Weka представляет собой набор средств визуализации и алгоритмов для анализа данных и решения задач прогнозирования, вместе с графической пользовательской оболочкой для доступа к ним.

С помощью такого подхода достигается универсальность: в главной базе данных хранится информация о пользователях (атрибуты) из любой из возможных социальных сетей, представленная в таком виде, что с ней могут любые алгоритмы ML, находящиеся в соответствующем блоке.

7. Выводы

Для решения проблемы спама в социальных сетях были рассмотрены основные анти-спам стратегии. Важной частью является идентификация спаммеров, основанная, прежде всего, на основе анализа профилей пользователей, социального поведения, а так же некоторых видов контента, сгенерированного самими пользователями. Необходимой для дальнейшей классификации является предварительная обработка профилей пользователей, методы которой рассмотрены.

Алгоритмы классификации являются наиболее важной частью системы детектирования нелегитимных пользователей. В работе рассмотрен поход, учитывающий контекст, присущий различным сетям, рассмотрены основные контекстные и прогнозирующие атрибуты.

Представленный подход к классификации пользователей универсален, позволяет проводить классификацию пользователей в социальных сетях любого типа.

В статье рассмотрена архитектура системы исследования и классификации пользователей, которая не только получает информацию из социальных сетей любого типа с помощью соответствующего API, но и обрабатывает данные, приводит к единому виду, хранит в единой для всех социальных сетей базе данных в формате, с которым могут работать любые алгоритмы машинного обучения, анализа данных и т.д., находящиеся в соответствующем блоке.

Таким образом, достигается универсальность системы: в главной базе данных хранится информация о пользователях (атрибуты) из любой из возможных социальных сетей, представленная в таком виде, что к этим данным могут применяться любые методы, алгоритмы.

Литература

1. Boyd, D. Social Network Sites: Definition, History, and Scholarship [Text] / D. Boyd, N. Ellison // Journal of Computer-Mediated Communication. – 2007. – Vol. 13, № 1. – P. 210–230.
2. Graham, P. A Plan For Spam [Электронный ресурс] / P. Graham. – Режим доступа : www. URL: http://www.paulgraham.com/spam.html.
3. Hayati, P. HoneySpam 2.0: Profiling Web Spambot Behaviour [Text] / P. Hayati, K. Chai, V. Potdar, A. Talevski // Lecture Notes inComputer Science. – 2009. – Vol. 5925. – P. 335–344.
4. Webb, S. Honeypots: Making Friends With A Spammer Near You [Text] / S. Webb , J. Caverlee, C. Pu // Proceedings of the Fifth Conference on Email and Anti-Spam (CEAS 2008), 21 August, 2008. – Mountain View, 2008.
5. Heymann, P. Fighting Spam on Social Web-sites: A Survey of Approaches and Future Challenges [Text] / P. Heymann, G. Koutrika, P. Garcia-Molina // IEEE Internet Computing. – 2007. – Vol. 11, № 6. – P. 36–45.
6. Webb, S. Uncovering Social Spammers: Social Honeypots+Machine Learning [Text] / S. Webb , J. Caverlee, K. Lee // Proceedings of the 33rd Annual ACM SIGIR Conference (SIGIR 2010), 19–23 July 2010, Geneva, Switzerland. – Geneva, 2010.
7. Porter, M. An algorithm for suffix stripping [Text] / M. Porter // Program. – 1980. – Vol. 14, № 3. – P. 130–137.
8. Terziyan, V. A Bayesian Metanetwork [Text] / V. Terziyan // International Journal of Artificial Intelligence Tools. – 2005. – Vol. 14, № 3. – P. 371–384.