А.В. Супруненко

Способ экспертной оценки с использованием сети репутации для решения задачи классификации веб-контента

Аннотация. В статье описываются ключевые свойства известных моделей систем репутации, которые могут быть использованы для решения задач классификации. Приводится описание способа экспертной оценки с использованием сети репутации для решения задачи классификации веб-контента. Показана реализация приведенного способа в системе фильтрации веб-контента.

Ключевые слова: классификация контента, фильтрация контента, сеть репутации, распределённая система, облачный сервис.

Введение

Интернет — быстро растущая и изменяющаяся среда. Каждый день появляется и исчезает огромное количество сайтов и страниц, содержащих информацию по различным областям знаний. Глобальная сеть становится источником не только текстового, но и мультимедийного контента. Вместе с тем растет и молодеет Интернет-аудитория, все чаще посетителями сети становятся дети.

В последнее время в разных странах стало появляться все больше ограничительных мер по размещению противоправной информации и доступу к ней, подкрепленных законодательно. Учитывая размер глобальной сети и скорость изменения информации в ней, законодательные меры, обязывающие контент-провайдеров запрещать доступ к противоправной информации того или иного характера, не всегда оказываются своевременными и эффективными. Кроме того, информация, размещаемая на Интернет-ресурсах, не являясь сама по себе противоправной, может, однако, нанести вред здоровью и развитию детей в силу того, что предназначена для более старшей, например, совершеннолетней, категории пользователей. Таким образом, задача классификации веб-контента по нескольким признакам, позволяющим осуществлять разграничение доступа по заданному набору критериев, является актуальной для Интернет-пользователей различных категорий.

В данной статье приводится обзор свойств существующих моделей, использующихся в системах репутации, с акцентом на зависимость значений репутации от определенных свойств модели. Описывается способ экспертной их оценки, особенностью которого является вычисление совокупной оценки Интернет-ресурса на основе множества индивидуальных экспертных оценок, взвешенных относительно показателей репутации каждого из экспертов. Приводится описание реализации данного способа в распределенной системе фильтрации веб-контента.

1.Свойства существующих моделей систем репутации

Далее будут рассмотрены некоторые свойства [11, 13] существующих моделей систем репутации, которые могут быть использованы для решения задач классификации и оказывают влияние на способы перерасчёта значений репутаций агентов.

Прямое и транзитивное доверие. Существуют системы [2, 8], моделирующие только отношения прямого доверия. В таких системах решение о взаимодействии с агентом принимается на основании опыта предыдущих контактов и (возможно) дополнительных характеристик, таких как, например, контекст взаимодействия (ситуативность), взаимности, выполнение ролей и т.д. [5]. При этом на изменение значения репутации влияет только непосредственный опыт взаимодействия между агентами. Системы, в которых помимо прямого доверия учитывается также и транзитивное [3, 4, 6, 9], позволяют формировать мнение об агенте на основе непрямого взаимодействия, например, при помощи «семантического расстояния» [3], «ядра доверия» [4], цепочки или сети доверия [6, 9, 14] или социальной сети [7]. В этом случае значения репутации могут быть относительными или вычисляемыми (в зависимости от выстроенных транзитивных отношений), а изменения этих значений могут происходить и без непосредственного взаимодействия между агентами.

Анонимность агентов сети. В некоторых существующих системах [6, 15] анонимность агентов является одним из свойств, позволяющих повысить устойчивость системы к манипуляциям недобросовестных пользователей. В этом случае исключается или минимизируется возможность агента целенаправленно повышать собственную репутацию и объединяться с другими агентами с этой целью.

Устойчивость, надежность и автономность системы. Устойчивость и надежность обусловлены механизмом изменения репутации агентов и, как, например, в [4], отсутствием стимулов совершения пустых транзакций между агентами [15]. Автономность или саморегулируемость отражает степень зависимости системы от необходимости внешних регулирующих воздействий. Существуют модели, например [6], способные работать полностью автономно.

Выражение значения репутации. Существуют модели, в которых значения репутации выражаются дискретными значениями [3, 4, 12] или числовыми значениями в определенном диапазоне [5-8, 15]. Причем в модели [8] значение находится в фиксированном предопределённом диапазоне, в модели [15] ограничено снизу и сверху. Также возможно выражение репутации в виде совокупности значений, как, например, в [10]. Изменение значения репутации происходит после завершения процесса взаимодействия между агентами сети репутации. Известны модели, в которых решение о взаимодействии принимается с использованием пороговых значений [2, 14]. В некоторых моделях, кроме завершения транзакции, используются и другие условия изменения репутации, например, выполнение определенных действий (ролей) [5].

2.Описание способа экспертной оценки интернет-ресурсов

В предложенном способе агенты сети репутации не взаимодействуют друг с другом напрямую. В качестве показателя репутации используется не мнение об агенте других участников сети [1], а степень доверия системы данному агенту. В этом случае является неизменным контекст взаимодействия, а значит, репутация может быть выражена единственным значением. Значение репутации агента используется при расчете совокупной оценки Интернет-ресурса. При этом каждый агент оказывает влияние на совокупную оценку пропорционально его репутации. Могут использоваться многофакторные оценки. Изменения значений репутации каждого из агентов происходит на основе опыта его участия в оценке того или иного Интернет-ресурса.

Корректировка значения репутации каждого агента сети описывается формулами:

$$r_i^* = r_i + \frac{R^- \cdot r_i}{n \cdot R}$$

для агентов, чьи оценки совпали с совокупной оценкой;

$$r_i^* = r_i - \frac{R^+ \cdot r_i}{n \cdot R}$$

для агентов, чьи оценки не совпали с совокупной оценкой,

где r_i^* - скорректированное значение репутации эксперта; r_i - текущее значение репутации эксперта; R^- - сумма значений репутаций агентов, чьи оценки не совпали с совокупной оценкой; R^+ - сумма значений репутаций агентов, чьи оценки совпали с совокупной оценкой; n - количество агентов, оставивших оценки; R - сумма значений репутаций агентов, оставивших оценки.

Предложенная модель обладает следующими свойствами:

1. Анонимность пользователей. Агенты взаимодействуют с системой, указывая свой идентификатор, в соответствие которому поставлено значение репутации. Значение репутации корректируется на основе опыта участия агента в формировании оценки Интернет-ресурса.

2. Независимость пользователей. Агенты не взаимодействуют друг с другом. В системе отсутствует требование минимального числа агентов. При расчете совокупной оценки Интернет-ресурса отсутствует пороговое значение, определяющее достаточное количество оценок (как, например, в модели, описанной в [2]).

3. Отсутствие транзитивных отношений. Отношения доверия существуют только между системой и каждым из агентов, основанное на значениях их репутации. Учитывая отсутствие непосредственного взаимодействия между агентами, также отсутствует необходимость в транзитивном доверии между агентами.

4. Отсутствие привилегий для новых участников сети. Первоначальное значение репутации агента позволяет ему увеличить свое значение репутации в случае совпадения его мнения с рассчитанной оценкой каждого Интернет-ресурса. Уменьшение значения репутации в случае несовпадения его мнения с рассчитанной оценкой Интернет-ресурса происходит пропорционально текущему значению репутации.

5. Надежность. Способ перерасчета значения репутации учитывает возможность «манипуляции» системой множеством недобросовестных агентов.

6. Саморегулируемость и автономность. Репутационная масса системы увеличивается по мере регистрации новых агентов, уменьшая тем самым значимость каждого из них в системе.

3.Реализация способа

Здесь приведен вариант реализации способа в системе фильтрации веб-контента. Система включает в себя (Рис. 1):

1. Одно или несколько клиентских устройств пользователя, с которых предпринимаются попытки доступа к Интернет-ресурсам.

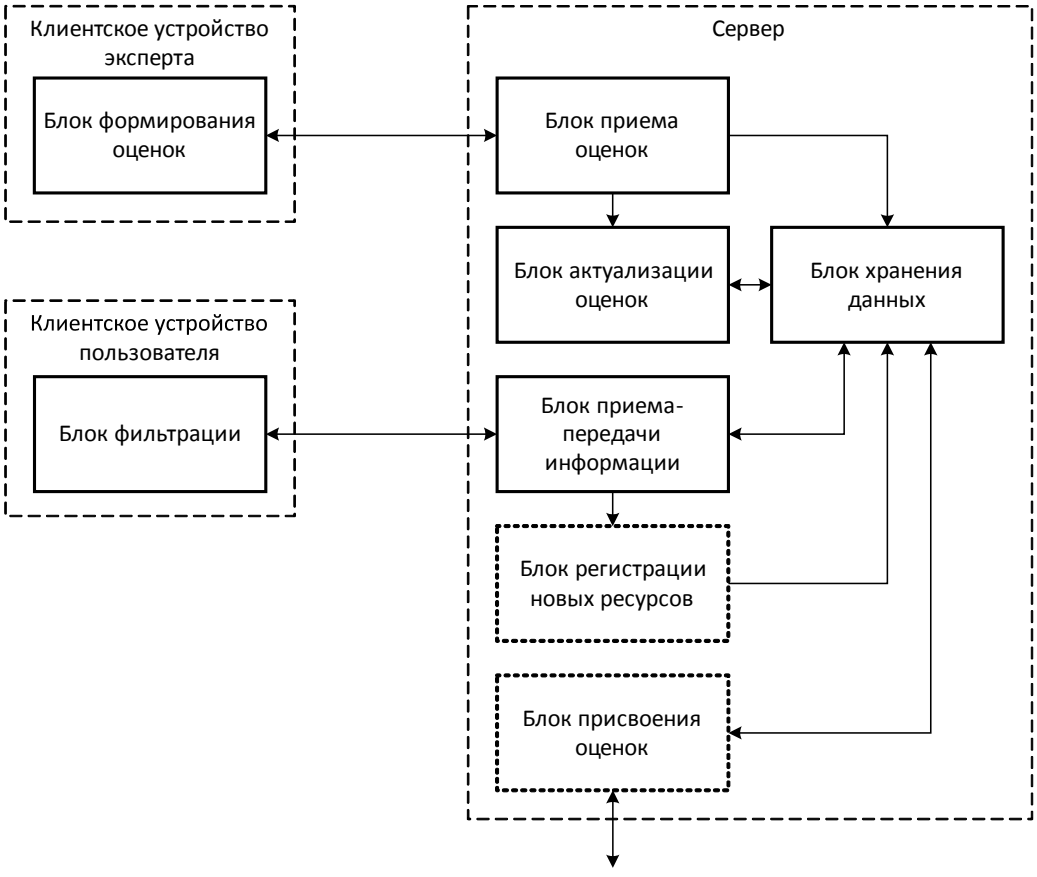

Рис. 1. Структурная схема системы фильтрации контента

В качестве такого устройства могут выступать настольные и/или мобильные устройства, оснащенные браузером или другими средствами доступа к сети. На устройстве устанавливается фильтрующее приложение, осуществляющее предоставление или запрет доступа. Приложение может быть реализовано в виде плагина к браузеру, либо как служба в системе.

2. Одно или несколько клиентских устройств эксперта (агента сети репутации). В качестве такого устройства могут использоваться настольные и/или мобильные устройства, на которых установлено приложение, позволяющее отправлять в систему экспертные оценки Интернет-ресурсов. Приложение может быть реализовано в виде плагина к браузеру или как веб-служба.

3. Серверная часть в составе блоков приема и актуализации оценок, сервиса приема запросов пользователей и отправки ответов, базы данных, а также блоков регистрации новых ресурсов и присвоения оценок.

Предварительно на клиентском устройстве пользователя, посредством которого предполагается осуществлять доступ к Интернет-ресурсам, например, на персональном компьютере, производится настройка профиля доступа пользователя к веб-контенту с учетом его возраста, тематических предпочтений, способа ограничения доступа («мягкий» — разрешено все, что не запрещено явно, или «жесткий» — запрещено все, кроме того, что разрешено явно) и других критериев.

С учетом настроенного профиля доступа пользователя осуществляется доступ к веб-контенту с одновременной фильтрацией веб-контента, при которой реализуются следующие подэтапы:

1. Формируется пользовательский запрос на получение целевого веб-контента в виде запроса по протоколу HTTP(S) или другого принятого для данной сети. В качестве целевого веб-контента может выступать, например, страница сайта с текстовыми и/или мультимедийными материалами.

2. Извлекается идентификатор целевого веб-контента из сформированного запроса в виде сетевого адреса URL (Uniform Resource Locator) или другого принятого для данной сети идентификатора.

3. Осуществляется получение совокупной оценки целевого веб-контента из базы данных. Для этого производится запрос в локальную или удаленную базу данных, содержащую список идентификаторов веб-контента и соответствующие им совокупные оценки. В случае неполучения совокупной оценки целевого веб-контента из базы данных оценок после ее запроса, производится запрос во внешние базы данных, доступные в сети, например, в сети Интернет, на получение оценок целевого веб-контента и, в случае получения, сохранение их в базе данных так же, как и оценки, полученные от экспертов. В случае неполучения оценок из внешних баз данных целевой веб-контент считается неоцененным и решение о разрешении или запрете доступа к веб-контенту принимается на основании критериев, описанных для такого случая в профиле доступа пользователя.

4. Принимается решение о доступе или отказе в доступе пользователя к целевому веб-контенту на основании установленного пользователем профиля доступа и полученной совокупной оценки. После этого на основании принятого решения пользователю предоставляется или запрещается доступ к целевому веб-контенту. При этом пользователю может быть направлено сообщение, содержащее причину отказа в доступе.

Одновременно с этим или после этого проводится актуализация и формирование совокупных оценок веб-контента, при которых реализуют следующие подэтапы:

1. Осуществляется выбор веб-контента для его оценки и производится его экспертная оценка, включающая классификацию выбранного веб-контента по нескольким признакам, например, по принадлежности к возрастной и тематической категории, наличию обсценной лексики и т. п.

2. Производится сохранение присвоенной оценки веб-контента в базу данных, где в соответствии с оценкой находятся идентификационные данные эксперта, присвоившего оценку и идентификатор веб-контента.

3. Осуществляется расчет совокупной оценки веб-контента на основе присвоенной экспертом оценки, значения его репутации, а также присвоенных оценок и значений репутаций экспертов, ранее оценивших данный веб-контент, а именно — как средневзвешенное относительно значений репутаций экспертов значение всех присвоенных данному веб-контенту оценок. При этом оценки, сделанные эксперта-

ми с бóльшими значениями репутации, учитываются при расчете в большей мере, а оценки, сделанные экспертами с меньшими значениями репутации, учитываются в меньшей степени.

4. Рассчитанная совокупная оценка веб-контента сохраняется в базе данных, где ей в соответствие ставится идентификатор веб-контента.

5. Производится корректировка значения репутации каждого эксперта, оценившего данный веб-контент, на основе рассчитанной совокупной оценки веб-контента и присвоенной данным экспертом оценки, а также сохранение скорректированного значения репутации в базу данных.

Корректировка значения репутации каждого эксперта осуществляется следующим образом:

1. Значение репутации повышается пропорционально ее текущему значению в случае совпадения присвоенной экспертом оценки с рассчитанной совокупной оценкой веб-контента и одновременно понижается пропорционально текущим значениям репутаций экспертов, ранее оценивших данный веб-контент, значения оценок которых отличаются от рассчитанной совокупной оценки.

2. Значение репутации понижается пропорционально текущему ее значению в случае несовпадения присвоенной экспертом оценки с рассчитанной совокупной оценкой веб-контента и одновременно повышается пропорционально текущим значениям репутаций экспертов, ранее оценивших данный веб-контент, оценки которых совпадают с рассчитанной совокупной оценкой.

Рассмотренная модель может быть использована и в других системах.

Заключение

Представленный способ реализован в распределенной системе фильтрации контента «Этикум». Данные о пользователях системы (агентах) хранятся в облаке. В качестве фильтрующего приложения используется расширение соответствующего браузера или системное приложение, анализирующее исходящие сетевые запросы. Данные о классифицированных Интернет-ресурсах хранятся в облачной базе данных и доступны через веб-сервис, в том числе и для сторонних пользователей. Набор функций для передачи экспертных оценок в систему является частью фильтрующего приложения для пользователей, зарегистрированных в системе в качестве экспертов. Перерасчет показателей репутации пользователей осуществляется централизованно с сохранением полученных значений в облаке.

Литература

1. Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка // http://www.ozhegov.org/words/30637.shtml .
2. Супруненко А. В. Модель открытой распределенной системы фильтрации веб-контента // Системы управления и информационные технологии, 2011, № 1 (43), с. 90–95.
3. Abdul-Rahman A., Hailes S. Supporting trust in virtual communities // In: Proc. of Hawaii International Conference on System Sciences, 2000.
4. Advogato Trust Metric // http://www.advogato.org/trust-metric.html
5. Carter J. Reputation Formalization for an Information-Sharing Multi-Agent System // Computational Intelligence, vol. 18 (2), p. 515–534.
6. Kamvar S.D., Schlosser M.T., Garcia Molina H. The EigenTrust Algorithm for Reputation Management in P2P Networks // Proceedings of the 12th international conference on World Wide Web, 2003, p. 640–651.
7. Golbeck J., Parsia B., Hendler J. Trust Networks on the Semantic Web // Cooperative Information Agents VII, 2003, p. 238–249.
8. Marsh S. Formalising Trust as a Computational Concept // PhD dissertation, University of Stirling, 1994.
9. Richardson M., Agrawal R., Domingos P. Trust management for the semantic web // International Semantic Web Conference, 2003, p. 351–368.
10. Sabater J., Sierra C. Reputation and social network analysis in multi-agent systems // Proceedings of the first international joint conference on Autonomous agents and multiagent systems, 2002, p. 475–482.
11. Sanger J., Richthammer C., Pernul G. Reusable components for online reputation systems // Journal of Trust Management 2015, 2:5, 2015.
12. Schillo M., Funk P., Rovatsos M. Using trust for detecting deceitful agents in artificial societies // Applied Artificial Intelligence, 14, 2000, p. 825–848.
13. Sherchan W., Nepal S., Paris C. A Survey of Trust in Social Networks // ACM Computing Surveys, Vol. 45, No. 4, Article 47, August 2013.
14. Yu B., Singh M. P. An evidential model of distributed reputation management // Proceedings of the first international joint conference on Autonomous agents and multiagent systems, 2002, p. 294–301.
15. Zacharia G. Trust management through reputation mechanisms // Applied Artificial Intelligence, 2000, vol. 14, p. 881–907.

Александр Владимирович Супруненко. Нижегородский государственный технический университет им. Р.Е. Алексеева. Институт радиоэлектроники и информационных технологий. Кафедра «Вычислительные системы и технологии». Старший преподаватель.