



УДК 004.822:514

РАСПОЗНАВАНИЕ БОТОВ В СОЦИАЛЬНЫХ СЕТЯХ

Кипаева Е.В., Кириченко М.И., Орлова Ю.А., Заболеева-Зотова А.В.

*Волгоградский государственный технический университет,
г. Волгоград, Россия*

M9iy.aht@gmail.com

Nemozy@list.ru

yulia.orlova@gmail.com

zabzot@gmail.com

В работе изучается описание понятия "Бот", рассматривается применение ботов в различных сферах жизни. Проводится исследование методов защиты от роботизированного внесения информации и общения. В статье также приведены существующие системы распознавания ботов, а также рассматривается бот, как интеллектуальная диалоговая система.

Ключевые слова: бот; обратный тест Тьюринга; методы распознавания; интеллектуальная диалоговая система.

Введение

Интеллектуальные диалоговые системы стали очень распространены в наше время. Данные системы могут использоваться в различных сферах жизни, например, для автоматической проверки знаний людей, для ведения диалогов с пользователями глобальной сети Интернет, для диагностики неисправностей автомобилей, а также для развлечений.

Работа авторов статьи направлена на изучение так называемых виртуальных собеседников или, как их еще часто называют – чат-ботах. В статье описывается понятие сущности бота, его применение и методы распознавания. Также рассматривается бот, как интеллектуальная диалоговая система.

1. Описание понятия бот

1.1. Бот

Бот (англ. bot, сокр. от robot) - программа, автоматически выполняющая действия на компьютере вместо людей. Впервые термин «робот» использовал писатель Карел Чапек в своей пьесе «Россумские универсальные роботы». В семидесятых годах был употреблен впервые термин «бот».

1.2. Применение ботов

Боты могут использоваться в различных сферах жизни, например:

- имитация живого человека при общении в чате;
- автопилот для участия в Интернет-аукционах, скупки билетов и прочих ценностей;
- автоматизация рутинных действий, таких как рассылка разнообразных рекламных акций по электронной почте;
- массовые вычисления, в частности взлом паролей и индексация сетевых ресурсов, получение и предоставление приватной информации, включая пароли и номера кредитных карт.

2. Методы защиты от ботов. Обратный тест Тьюринга

Бороться с современными ботами средствами IT-технологий довольно сложно. Какая бы не была придумана защита, через несколько дней она обходится хакерами, и решения борьбы с данной защитой перепродаётся среди взломщиков.

Для защиты от подобных агентов, системы начинают формировать нечто вроде «иммунной системы», а именно на входе производится тестирование интеллекта, где пользователь должен убедить систему, что представляет собой человека (рисунок 1).

Поскольку в данной системе компьютер верифицирует наличие интеллекта у исследуемого объекта, такой тест называется обратным тестом Тьюринга, поскольку прямой тест Тьюринга производится человеком, для определения наличия

у машины интеллекта. Институт Карнеги Меллон ввел специальный термин CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), что переводится как «Полностью Автоматический Открытый Тест для Отделения Компьютеров от Людей» [Коробкова и др., 2012].

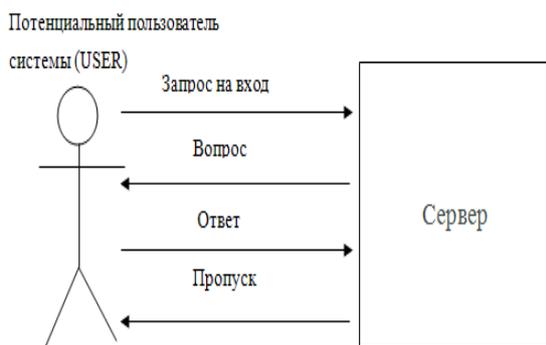


Рисунок 1 – Иммунная система для защиты от интеллектуальных агентов (веб-роботов)

Обратные тесты Тьюринга, могут быть как зрительными так и слуховыми. Типовой пример внешнего вида такого теста на веб-странице показан на рисунке 2.

Введите символы, показанные на картинке.



Рисунок 2 – Внешний вид обратного теста Тьюринга

В данном примере, сложность теста заключается в искажении символов, затрудняющем применение технологии OCR (Optical Characters Recognition). Однако, как показывают исследования Microsoft, подобные тесты могут быть решены без участия человека с применением нейросетевых алгоритмов [Розалиев и др., 2010].

Более сложные обратные тесты требуют смыслового анализа изображений, фраз, решения задач (рисунок 3).



Рисунок 3 – Обратный тест Тьюринга, требующий от пользователя указать три изображения, представляющих собой «Природу»

Такие тесты не сводятся к решению «усложненной» проблемы OCR, а требуют совершенно иного уровня искусственного интеллекта, для их взлома, пока еще не созданного. К проблемам подобных тестов можно отнести громоздкость, ограниченный запас терминов, ограниченный запас изображений.

3. Существующие системы распознавания ботов

В наше время существует достаточное количество систем для распознавания ботов. Примером такой системы может являться программный продукт Akismet, который автоматически обнаруживает спам (рассылка коммерческой и иной рекламы или иных видов сообщений лицам, не выразившим желания их получать). Данная система постоянно развивается и обучается, что позволяет обнаруживать новых сложных ботов, с которыми другие сервисы не могут справиться. Для личного пользования разработчики предоставляют ключ. Объясняется бесплатное использование тем, что сервису надо обучаться, чтобы более крупные компании и коммерческие проекты могли уверенно использовать на платной основе и не беспокоиться за спам. Положительными сторонами является то, что боты распознаются абсолютно на любых социальных сетях или сервисах общения. Всё зависит от клиентов, от того, как они будут использовать доступ к этому сервису. Минусами для русских разработчиков является то, что разработчик-иностранец и цены на использование сервиса очень высокие [Котенко и др., 2012].

Также примером такой системы является программный комплекс HumanPresent, который блокирует действия, выполняемые автоматизированными системами. По словам главы Pramana Санджая Сегала, компания планирует запустить HumanPresent как в виде SaaS-сервиса, так и в виде программы, предотвращающей активность ботов внутри веб-приложений. Программный комплекс может быть использован для защиты таких онлайн-форм, как бланки регистрации почтовых аккаунтов и транзакций интернет-магазинов, а также для обнаружения автоматизированных кликов по рекламным баннерам. Сегал сообщил о том, что Pramana использует 32 параметра для выявления активности бота на веб-странице. В их числе – особенности и частота нажатия на клавиши мыши и клавиатуры, а также некоторые другие вещи. Например, если человек нажмет на ссылку и ничего не произойдет, он, скорее всего, попытается нажать на нее снова, тогда как бот не станет этого делать.

«Vkontakte Antispam» - система распознавания ботов, предназначенная для социальной сети ВКонтакте. С его помощью можно сократить в разы количество спама в группах и публичных страницах, которыми вы управляете. На данный момент в системе реализован только функционал

выявления и добавления ботов в чёрный список. Программный продукт был создан в 2012 году Дмитрием Москниным. Достоинствами данной системы является усиленная направленность на обнаружение спама с дальнейшей ликвидацией и отправки автора в чёрный список. Хочется отметить, что программа распространяется бесплатно и является довольно эффективной. К недостаткам данной разработки можно отнести то, что она не может проверять ссылки, содержащие пробелы.

4. Критерии для распознавания ботов в социальной сети "ВКонтакте"

Методы распознавания ботов разделены на два вида - на проверку получаемой информации во время присутствия пользователя на ресурсе (онлайн) и на проверку информации, которая уже была ранее залита на онлайн-сервис (оффлайн). Во втором случае информацией могут являться фотографии, которые пользователь выкладывал с последней встречи выпускников, публичные сообщения, информация о зарегистрированном человеке, например, дата рождения, место проживания. По всем этим данным можно определить реальный ли это человек. Сто процентного результата, конечно, получить не сможет даже сам человек, не говоря уже про то, что анализировать данную информацию, может компьютер.



Рисунок 4 – Главное окно при авторизации в социальной сети "ВКонтакте".

Чтобы помочь определить бот это или нет, на сегодняшний день выделяют ряд критериев, которые, как уже писалось ранее, не дают сто процентного результата, но совокупность этих критериев повышает шанс обнаружить бота. Чтобы привести пример такого критерия рассмотрим социальную сеть "ВКонтакте" (Рисунок 4). Если на аватарке (графическое представление пользователя, двумерное изображение (иконка) или трёхмерная модель) стоит изображение с явной рекламой стороннего контента и другие фотографии отсутствуют, то, скорее всего, это будет бот, созданный для рекламных акций какого-либо продукта.

В своей работе мы предлагаем следующий ряд критериев, выделенный для распознавания ботов в социальной сети "ВКонтакте".

Таблица 1 – Список критериев выявления ботов

Критерий	Комментарий
Заблокирован ли пользователь.	Блокировка пользователей в «ВКонтакте» в 99 из 100 случаев происходит из-за вмешательства сторонних программ, которые осуществляют спам-деятельность.
Количество подписчиков.	Количество людей, которые подписаны на пользователя, не должно быть больше 300.
Полнота заполнения полей при регистрации.	Когда регистрируют ботов, то обычно стараются заполнить минимум данных, чтобы избежать лишних затрат времени. Если ботов регистрируют с помощью программных средств, то поля заполняются по максимуму.
Количество комментариев от друзей.	Если большинство комментариев написано не от друзей то, возможно, это бот.
Орфография в комментариях.	Человеку свойственно делать орфографические ошибки. Если за все время существования страницы все сообщения написаны без единой ошибки – это признак автоматизированной системы.
Ведения диалога	Современные автоматизированные системы еще не обучены поддерживать беседу на заданную тему. Поэтому если при общении с пользователем разговор перерастает в форму "вопрос – ответ" то, скорее всего это бот.

5. Бот, как интеллектуальная диалоговая система

Одним из примеров ботов являются чат-боты. Это специальные виртуальные собеседники, созданные для общения в различных Интернет-каналах. В научных исследованиях такие боты называются интеллектуальными диалоговыми системами [Заболеева-Зотова и др., 2010].

Интеллектуальные диалоговые системы работают по общей схеме: воспринимают сообщение пользователя, как запрос и формирует соответствующий ответ на основе имеющихся знаний. Систем, учитывающих в совокупности морфологию, синтаксис и семантику запросов на естественном языке - нет. Несмотря на большой объем работ, проведенных в области обработки естественно-языковых текстов, к настоящему

моменту не предложено механизмов практического их применения для поиска информации [Möller, 2005].

В России довольно много разработано интеллектуальных диалоговых систем, но подавляющее большинство очень примитивны. Данные системы могут только поддерживать беседу на заданную тему по типу вопрос-ответ. Умение поддерживать беседу в таких разработках не реализовано.

В данной статье хотелось бы предложить определенный минимум, которым должна обладать интеллектуальная диалоговая система. Такая система должна обладать следующими функциональными возможностями:

- умение поддерживать беседу на заданную тему, то есть система должна не просто отвечать на вопросы, но и задавать их, спорить, отстаивать свою точку зрения;
- механизм самообучения;
- механизм "запоминания" и "узнавания", то есть система должна запоминать собеседника, чтобы в новом диалоге с ним не пришлось начинать беседу сначала;
- обладать индивидуальностью и эмоциональностью;
- умения выполнять какие-либо действия, например, если собеседник скидывает какую-либо ссылку, то система должна ее открыть.

Интеллектуальная диалоговая система может быть внедрена в различные социальные сети для рекламных акций. Если система будет обладать функциональными возможностями, описанными выше, то и распознать ее будет гораздо тяжелее.

Заключение

Обработка информации и принятие решений при человеко-компьютерном взаимодействии является актуальными задачами, требующими пристального внимания. [Бобков и др., 2011] На современном этапе развития информационных технологий разработка методов автоматического распознавания бота является актуальной задачей, позволяющей решить ряд экономических, социальных, бытовых проблем и играющей важную роль в вопросах безопасности.

Разрабатываемые системы распознавания ботов могут применяться в бизнес-компаниях, где требуется работать с клиентами через Интернет. Использование системы позволит увеличить скорость определения ботов, тем самым, повышая эффективность работы аналитиков. Интерес к применению в своих целях может найти не только организация, но и частное лицо. Например, при общении с пользователем в Интернете, можно в любой момент запустить систему распознавания ботов и убедиться, что ведешь беседу с реальным человеком. С помощью данных программ можно проводить социальные исследования, например, на

количество ботов в той или иной социальной сети. Также такие системы могут применяться в сфере обеспечения безопасности для повышения эффективности работы соответствующих служб на крупных объектах.

Работа частично поддержана Российским фондом фундаментальных исследований (проекты 12-07-00266, 12-07-00270, 13-07-00459, 13-07-97042).

Библиографический список

- [Бобков и др., 2011] Развитие системы автоматизированного определения эмоций и возможные сферы применения / Бобков А.С., Заболева-Зотова А.В., Орлова Ю.А., Розалиев В.Л. // Открытое образование. - 2011. - № 2. - С. 59-62.
- [Заболева-Зотова и др., 2010] Автоматизация семантического анализа текста технического задания: монография. / Заболева-Зотова А.В., Орлова Ю.А. - Волгоград, ИУНЛ. 2010. - 155 с.
- [Коробкова и др., 2012] Обратный тест Тьюринга - противостояние естественного и искусственного интеллекта / Коробкова С. В. [и др.] // Нейрокомпьютерная обработка сигналов и изображений. - 2012. - 8 с.
- [Котенко и др., 2012] Исследование бот-сетей и механизмов противодействия им на основе имитационного моделирования / Котенко И. В. [и др.] // СПИИРАН. - 2012. - 26 с.
- [Розалиев и др., 2010] В.Л. Розалиев, А.С. Бобков, О.С. Федоров Применение нейронных сетей и грануляции при построении автоматизированной системы определения эмоциональных реакций человека / Изв. ВолгГТУ. Серия «Актуальные проблемы управления, вычислительной техники и информатики в технических системах»: межвуз. сб. науч. ст. / ВолгГТУ. - Волгоград, 2010. - Вып. 9, № 11. - С. 63-68.
- [Möller, 2005] Möller, S. Quality of Telephone-based Spoken Dialogue Systems / S. Möller, Springer Science + Business Media, Inc. Boston, 2005

RECOGNITION BOTS IN SOCIAL NETWORKS

Kipaeva E.V., Kirichenko M.I., Orlova Yu.A.,
Zaboleeva-Zotova A.V.

*Volgograd State Technical University,
Volgograd, Russia*

M9i.aht@gmail.com

Nemozy@list.ru

yulia.orlova@gmail.com

zabzot@gmail.com

In this articles the description of the concept of "bot", considered the use of bots in various spheres of life. A study of methods of protection against bots. The article also shows the existing recognition systems bots and bot regarded as intellectual dialogue system.

Intelligent dialogue systems have become very common nowadays. These systems can be used in various spheres of life, such as automatic test people's knowledge, for dialogue with the users of the Internet global network for fault diagnosis of vehicles, as well as for entertainment. Work the authors focused on the study of the so-called virtual companions, or, as they are often called - chat bots. The article describes the concept of the essence bot, its use and recognition methods. Also considered bot as intellectual dialogue system.