

нального компьютера с возможностью цифровой обработки сигналов. Оцифровывание выбранных каналов происходит последовательной коммутацией ключей с использованием одного АЦП. В модуле используется АЦП последовательного приближения. Модуль АЦП предназначен для оцифровки сигналов постоянного уровня и переменного напряжения. К одному компьютеру можно подключать несколько различных модулей АЦП-ЦАП, что позволяет создавать мобильные измерительные комплексы.

В результате исследования создана система управления ИЕП, которая может быть использована для задания различных режимов работы устройства и воздействия на него через организованные обратные связи на основе аппаратного и программного обеспечения ZETLab и ZETView. Это позволяет осуществлять контроль и управление резонансными токами и напряжениями для того, чтобы устройство оставалось работоспособным в режимах холостого хода нагрузки, вычислять текущие значения емкости и индуктивности, при которых в цепи поддерживается резонанс, задавать новые параметры (индуктивности и емкости) в случае необходимости регулирования выходного тока устройства.

1. Конесев С.Г., Хазиева Р.Т., Анализ динамики патентования индуктивно-емкостных преобразователей для систем стабилизации тока, Электротехнические и информационные комплексы и системы, 4, (2016).

ИССЛЕДОВАНИЕ ПОДСИСТЕМЫ КОНТЕНТНОЙ ФИЛЬТРАЦИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Кононова Н.В., Самокаева Т.А., Андрусенко Ю.А.

ФГАОУ ВО «Северо-Кавказский федеральный университет»,
г. Ставрополь, Россия

*E-mail: knv_fm@mail.ru

THE STUDY OF SUBSYSTEMS OF CONTENT FILTERING USING MACHINE LEARNING METHODS

Kononova N.V., Samokaeva T.A, Andrusenko U.A.

North Caucasus Federal University, Stavropol, Russia

Annotation. The Internet is a rapidly growing area that should not get out of control. Content filtering is widely used in various countries, but it is mostly unreliable methods.

При моделировании реальных процессов часто условия классической линейной модели регрессии оказываются нарушенными. Класс допустимых функций от классификаторов должен быть не слишком большим, чтобы избежать переобучения. Чаще всего используют линейные комбинации распознавателей, и даже не

все линейные комбинации, а только выпуклые. Для исследования подсистемы контентной фильтрации с использованием методов машинного обучения необходимо производить классификацию произвольных текстовых документов произвольным числом иерархически связанных категорий. Процесс реализации фильтрации в рамках настоящего исследования, схож с процессом разработки любой другой программы на основе методов машинного обучения. Методы машинного обучения с учителем требуют наличия обучающей выборки. После того, как обучающая выборка сформирована, каждый экземпляр из выборки выражается как вектор признаков. Далее, определяется категория для каждого элемента выборки – классификатор будет обучаться в дальнейшем. После формирования векторов, реализуется какой-либо алгоритм классификации. Полученный классификатор далее обучается на основе изначальной выборки и становится пригодным для осуществления классификации новых текстов.

Преобразование документов, имеющих вид последовательности символов, к виду, пригодному для алгоритмов машинного обучения в соответствии с задачей классификации. Обычно алгоритмы машинного обучения имеют дело с векторами в пространстве. Отображение документов в пространство признаков также используется и методами, основанными на знаниях. Для улучшения качества будущего классификатора на данном этапе желательно уменьшить размер пространства признаков. Учет всех слов приводит к слишком большой размерности пространства, хотя многие слова слабо влияют на результаты рубрицирования. Высокая размерность пространства признаков может приводить к высокой вычислительной погрешности и низкой скорости работы алгоритмов обучения. Отображение нескольких близких по значению слов в одну координату может улучшить результаты рубрицирования.

Построение классифицирующей функции при помощи обучения на примерах. Качество рубрицирования зависит и от того, как документы будут преобразованы в векторное представление, и от алгоритма, который будет применен на втором этапе. Методы преобразования текста в вектор специфичны для задачи классификации текстов и могут зависеть от коллекции документов, типа текста и языка документа.

Анализ обобщающей способности построенной классифицирующей функции сводится к получению некоторых оценок качества классификации, которые можно будет использовать для сравнения различных методов и оптимизации параметров метода.

Эти оценки качества нельзя получить, проверяя метод на коллекции документов. Иначе можно получить слишком завышенные оценки качества классификации. Такой алгоритм просто «запоминает» все полученные в процессе обучения документы вместе с соответствующими рубриками и сравнивает документы для рубрицирования с запомненными.