

Исходя из полученных зависимостей получаем, что наличие льготного периода увеличивает грант-элемент т.е. является неблагоприятным фактором для кредитора, что в свою очередь выгодно для должника, т.к. сокращает его потери, связанные с выплатами по кредиту.

ЗАКЛЮЧЕНИЕ

В статье приведены новые формулы для расчета условной потери кредитора W (условной субсидии заемщика) льготного займа, погашаемого линейно возрастающими или убывающими платежами, при наличии льготного периода или без. Данная форма погашения кредита, безусловно, более удобна для должника по сравнению с погашением кредита равными платежами, так как позволяет ему выбрать ту форму погашения, которая в большей степени учитывает наличие у него денежных средств в разные моменты времени. Изложенная теория анализа может найти применение в банковском секторе при анализе эффективности инвестирования банками средств в различные проекты.

Литература

1. *Четыркин Е.М.* Методы финансовых и коммерческих расчетов / Е.М. Четыркин. – М.: Дело Лтд, 1995.
2. *Четыркин Е.М.* Финансовый анализ производственных инвестиций / Е.М. Четыркин. – М.: Дело Лтд, 2001.
3. *Кирлица В.П.* Финансовая математика. Руководство к решению задач / В.П. Кирлица. – Мн.: ТетраСистемс, 2005.

КЛАССИФИКАЦИЯ ДОКУМЕНТОВ МЕТОДОМ ОПОРНЫХ ВЕКТОРОВ

А. О. Варивончик

ВВЕДЕНИЕ

В последнее время в мире очень большое внимание уделяется технологиям автоматизации процессов и в частности классификации документов. Сложно представить сферу человеческой деятельности, где бы не могли найти применение эти технологии – особенно в этом заинтересованы поисковые системы, для возможности получения релевантной информации интересующей пользователя. Имея возможность автоматически классифицировать документы, они смогут выдавать более качественные результаты поиска.

Классифицировать документы – значит проводить многоступенчатое, иерархическое и дихотомическое деление их на роды, виды, подвиды,

разновидности. В то же время в распознавании текста есть много нерешенных проблем и проблем, которые в ближайшее время решить не представляется возможным. Так, человек распознает символы на основе многолетнего опыта, сравнивая картинку с множеством виденных ранее по очень сложным алгоритмам. И все равно часто человек не может понять написанное другим. Компьютер же не обладает таким опытом и такой вычислительной мощностью, как человеческий мозг – потому ошибки появляются достаточно часто.

В статье рассматривается проблема разработки приложения, реализующего и демонстрирующего метод опорных векторов для классификации текстовых документов.

РЕАЛИЗОВАННОЕ ПРИЛОЖЕНИЕ

Демонстрационная программа представляет собой Windows-приложение, предназначенное для классификации текстовых документов.

Для реализации интерфейса пользователя использовалась графическая библиотека Java Swing. В качестве библиотеки для реализации самого метода опорных векторов была выбрана `jlsvm` [1].

Реализация предоставляет пользователю графический интерфейс:

- загрузка текстовых документов в программу для проверки;
- настройка параметров метода опорных векторов;
- изучение препроцессинга обработки документа;
- отображение информации, к какому классу принадлежит документ;
- сортировка результатов работы по папкам, где название папки отождествляется с названием класса.

Возможности разработчика шире, потому что часть функций не была включена в графический интерфейс. Он может:

- сохранить результаты обучения на базе данных в файл;
- самостоятельно установить классы для классификации;
- подготовить свою базу для обучения и тестовые данные.

В приложении были исследованы различные ядра для классификации документов:

- линейное;
- логистическое;
- экспоненциальное.

Лучше всего, себя показало линейное ядро. Существует гипотеза, что при достаточно большой размерности обучаемых документов

применение нестандартных ядер понижает точность вычислений, так как, до сих пор неясно в каких случаях, какие ядра оптимальнее применять, и если пытаться применять ядерные преобразования без должного глубокого исследования никаких улучшений не будет получено.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для обучения была выбрана одна из тестовых коллекций текста, которые легко можно найти в свободном доступе в интернете. Для реализации программного комплекса была использована коллекция RCV1 (Reuters Corpus Volume 1). Эта коллекция отсортированных вручную новостей агентства Reuters за период с августа 1996 по август 1997 года. Коллекция содержит около 810000 новостных статей на английском языке [2].

В коллекции кроме самих текстов и результатов их ручного классифицирования использовался документ, который хранил всевозможные слова, встретившиеся в этих документах. Всего около 40000 слов. Для этих слов уже был сделан стемминг (удалено окончание) и из списка слов удалены такие, слова, которые встречаются в практически в любом документе и не могут повысить точность распознавания документа. Например: предлоги.

Из всей группы документов были выбраны 1000 документов для тестов. Также словарь всех слов был урезан до 4000 слов, для того, чтобы время работы программы было хоть сколько-нибудь приемлемым. В противном случае время работы программы становилось слишком большим (порядка десятка минут для обучения). Слова были выбраны из словаря самые часто встречающиеся в документах. После этого 700 документов использовались для обучения, а оставшиеся 300 использовались для тестирования после обучения. Приложение продемонстрировало отличные результаты для линейного ядра.

Всего было проведено обучение по 88 категориям. В каждом из случаев результаты распознавания для конкретного класса не падали меньше 83 %. Общий процент распознавания оказался равен 97 %.

ЗАКЛЮЧЕНИЕ

В работе создано приложение для классификации текстовых документов. Для приложения реализованы базовые и наиболее используемые на сегодняшний день алгоритмы метода опорных векторов. Для максимальной эффективности необходимо, чтобы программа была предвари-

тельно обучена на базе данных предполагаемых текстов для классификации. Для различных сценариев эти базы будут существенно отличаться, также как и словари по которым производится построение документа. Соответственно встаёт задача подбора этих баз данных.

Проведенные исследования, показывают универсальность метода опорных векторов и их способность эффективно решать задачу классификации текста.

Литература

1. Интернет-адрес: <http://dev.davidsoergel.com/trac/jlibsvm/>.
2. Интернет-адрес: http://www.ai.mit.edu/projects/jmlr/papers/volume5/lewis04a/lyrl2004_rcv1v2_README.htm.

ПРИМЕНЕНИЕ МЕТОДА ОПОРНЫХ ВЕКТОРОВ К РЕШЕНИЮ ЗАДАЧИ КЛАССИФИКАЦИИ АУДИОДАННЫХ

Д. В. Волчек

ВВЕДЕНИЕ

В последние десятилетия прогресс в области цифровой музыки вылился в значительный рост наборов музыкальных данных. Такой рост постепенно выработывал большой интерес к проблеме структурной организации огромных музыкальных коллекций и вопросу исследования различных методов поиска цифровых данных.

Во многих научно–технических областях в основе решения прикладных задач лежит классификация сигналов (конечных последовательностей измерений некоторой величины). Классификация музыкальных произведений по жанрам также является хорошо известной проблемой в области исследования музыкальной информации и имеет широкий спектр применений. Но учитывая субъективность решения данной проблемы, такая классификация чаще всего производилась вручную. В связи с этим актуальной является проблема разработки эффективных методов для автоматической классификации сигналов. Один из возможных подходов к ее решению состоит в применении метода опорных векторов.

В данной статье рассматривается способ построения классификатора звуковых данных, который основывается на методе опорных векторов. Также рассматриваются проблемы реализации приложения, демонстрирующего работу такого классификатора.