

традиционных методов в пользу цифровых технологий, но позволяет на основе новой инструментальной базы освоить новую смысловую среду, дает гуманитариям возможность внести свой вклад в развитие информационных технологий.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Franco Moretti *Graphs, Maps, Trees: Abstract Models for a Literary History* [Electronic resource]. – Mode of access: https://www.mat.ucsb.edu/~g.legrady/academic/courses/09w259/Moretti_graphs.pdf.

2. Woloch, A. *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel* / A. Woloch. – Oxford : Princeton University Press, 2003. – P. 13–14.

С.И. Коровкин

Беларусь, Гомель, ГГУ имени Ф. Скорины

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДЫ АНАЛИЗА WEB-СТРАНИЦ В СЕТИ ИНТЕРНЕТ

В настоящее время все больше людей предпочитают получать и делиться информацией в сети Интернет: оставляют всевозможные отзывы о товарах, различных продуктах, местах, где побывали, и т.д. Чтобы повысить качество предлагаемых товаров и услуг, компаниям важно знать, как реагируют на их деятельность клиенты. Это очень большое количество информации, которое нужно обработать, структурировать и анализировать. В результате оценки данной информации компании могут делать выводы о дальнейших путях развития и способах привлечения новых клиентов. Таким образом, создание приложения для сбора и анализа различных моделей данных, таких как отзывы о товарах, записи из социальных сетей и др., является актуальной разработкой.

Web Mining – это использование методов интеллектуального анализа данных для автоматического обнаружения веб-документов и услуг, извлечения информации из веб-ресурсов и выявления общих закономерностей в Интернете.

Целью информационного извлечения является превращение коллекции документов в легко усвояемую и проанализированную информацию. Благодаря динамике и разнообразию веб-содержимого, создание ручного режима систем информационного извлечения не представляется возможным. В связи с этим большинство систем по извлечению данных сосредотачивают внимание на конкретные веб-сайты. Другие используют обучающие машины или методы интеллектуального анализа данных и способны извлекать веб-документы в автоматическом или полуавтоматическом режиме. С этой точки зрения Web Mining является частью процесса извлечения информации из Интернета.

Сбор данных – непростая задача, так как используемые алгоритмы могут быть очень сложными, а данные не всегда доступны в одном месте. При разработке приложения следует учесть следующее:

1. Методология получения данных и взаимодействие с пользователем.
2. Проблемы с производительностью.
3. Проблемы с разнородными типами данных.
4. Хранение данных.
5. Анализ данных.
6. Визуализация данных.

Н.С. Королёв, Д.С. Кузьменков
 Беларусь, Гомель, ГГУ имени Ф. Скорины

КЛАССИФИКАЦИЯ И ОБРАБОТКА НЕСТРУКТУРИРОВАННЫХ ДАННЫХ

В современном мире производители продукции (услуг) для конечного пользователя (B2C сектор) заинтересованы в совершенствовании качества обслуживания. Одним из самых быстрых, современных и действенных способов получить информацию о недостатках (или преимуществах) является сбор и анализ пользовательских данных. Использование анализа данных позволяет производителю реализовать двусторонний механизм обмена информацией и слышать голос пользователей, что позволяет позитивно влиять на качество продукта.

Для имплементации этого процесса необходимо собрать данные из публичных источников (и частных), путем интегрирования решения по обработке в пользовательское решение, надежно сохранить их (с возможностью масштабирования) и произвести непосредственную классификацию.

Сбор данных с публичных источников осуществляется посредством использования API и реализован на примере Twitter. Для частных данных добавлен способ загружать данные через Excel и через предоставляемый API.

Для хранения данных и последующей работы с ними был выбран свободный движок для поиска Elasticsearch. Elasticsearch – тиражируемая свободная программная поисковая система, основанная на Lucene. Написана на Java, распространяется по лицензии Apache, в основе использует библиотеку Lucene (как и вторая по популярности поисковая система – Solr), официальные клиенты доступны на Java, .NET (C#), Python, Groovy и ряде других языков. Обеспечивает горизонтально масштабируемый поиск, поддерживает многопоточность.

В качестве классификационных правил можно использовать не только ключевые слова, но и другие типы Lucene запросов. Был разработан механизм препарсинга для возможности внедрения собственных имён. Для удобства пользования добавлена функциональность для показа связанных слов и превью для выбранных классификационных правил.

Пользователю предоставлен интерфейс для построения пользовательских моделей классификации с использованием современных средств и просмотра отчетов по ним с использованием HTML 5 и JavaScript (Angular).

Д.О. Кривецкая, Е.И. Мирская
 Беларусь, Брест, БрГУ имени А.С. Пушкина

ВЫЧИСЛЕНИЕ ПЕРВОГО МОМЕНТА ОСРЕДНЁННОЙ ОЦЕНКИ СПЕКТРАЛЬНОЙ ПЛОТНОСТИ МНОГОМЕРНОГО ВРЕМЕННОГО РЯДА

Рассмотрим действительный стационарный в широком смысле случайный процесс $X(t) = \{X_a(t), a = \overline{1, r}\} t \in Z$, с неизвестной взаимной спектральной плотностью $f_{ab}(\lambda)$, $\lambda \in \Pi = [-\pi; \pi]$, $a, b = \overline{1, r}$.

Пусть $X_a(0), X_a(1), \dots, X_a(T-1)$ – T последовательных наблюдений, полученных через равные промежутки, за составляющей $X_a(t)$ процесса $X(t)$, $t \in Z$, $a = \overline{1, r}$.

В качестве оценки взаимной спектральной плотности $f_{ab}(\lambda)$, $\lambda \in \Pi$, $a, b = \overline{1, r}$, в работе исследована статистика

$$\widehat{f_{ab}^{(r)}}(\lambda) = \frac{1}{L} \sum_{l=0}^{L-1} I_{ab}^{lN}(\lambda), \quad (1)$$

$\lambda \in \Pi$, $a, b = \overline{1, r}$, которая построена путем осреднения расширенных периодограмм по L непересекающимся интервалам наблюдений.

Доказана

Теорема. Математическое ожидание оценки взаимной спектральной плотности $\widehat{f_{ab}^{(r)}}(\lambda)$ заданной соотношением (1), имеет вид

$$M \widehat{f_{ab}^{(r)}}(\lambda) = \int f_{ab}(x) \Phi_N(x - \lambda) dx, \\ \lambda \in \Pi, a, b = \overline{1, r},$$

где функция $\Phi_N(x)$, $x \in \Pi$, задается равенством

$$\Phi_T(x) = \left[2\pi \sum_{t=0}^{T-1} h_t^2(t) \right]^{-1} |\varphi_T(x)|^2,$$

а $\varphi_T(x)$, $x \in \Pi$, задается выражением

$$\varphi_T(x) = \sum_{t=0}^{T-1} h_t(t) e^{ixt},$$

где $x \in \Pi$.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Труш, Н. Н. Асимптотические методы статистического анализа временных рядов / Н. Н. Труш. – Минск : БГУ, 1999. – 218 с.

Е.И. Кульгун, А.П. Худяков

Беларусь, Брест, БрГУ имени А.С. Пушкина

РЕШЕНИЕ КРАЕВОЙ ЗАДАЧИ ДУФФИНГА

Решение краевой задачи Дуффинга проведено конечно-разностным методом с использованием метода неопределенных коэффициентов многоточечной аппроксимации производных. По полученному набору данных восстанавливалось решение при помощи различных алгоритмов аппроксимации. Восстановленная функция подставлялась в исходную задачу для того, чтобы узнать общую погрешность решения.

Точность решения проверялась по значению $\|R\|_{L_2}$:

$$\|R\|_{L_2} \cong \sqrt{\frac{b-a}{N} \sum_{i=1}^N R_i^2(x_i)}, \quad (1)$$

где N – количество точек разбиения отрезка интегрирования, $R(x)$ – невязка на отрезке $[x_{i-1}; x_i]$.

Результаты эксперимента для неперiodической и периодической задачи Дурфинга приведены в таблицах 1 и 2 соответственно, в ячейках которых записана точность при соответствующих алгоритмах аппроксимации. Случай неперiodической задачи рассматривался при крайних условиях I рода.

Таблица 1 – Точность при решении неперiodической задачи Дурфинга

Вид аппроксимации	Число точек разбиения отрезка				
	256	300	384	456	512
Отрезок ряда Фурье по полиномам Чебышева I рода					
	1.16406E-7	2.06484E-8	2.34147E-8	4.30735E-8	1.81517E-8
Сплайн 3-й степени					
	1.32515205	1.22457378	1.08289812	0.99399641	0.93820214
Неперiodический сплайн 5-й степени					
	0.00015704	8.220455E-5	3.021524E-5	1.510246E-5	9.473211E-6
Модификация эрмитовской аппроксимации					
	2.478211E-7	1.400339E-8	3.52402E-10	1.63685E-11	1.60925E-12

Таблица 2 – Точность при решении периодической задачи Дурфинга

Вид аппроксимации	Число точек разбиения отрезка				
	256	300	384	456	512
Отрезок ряда Фурье по полиномам Чебышева I рода					
	2.744785E-6	6.068492E-6	3.491175E-7	1.087275E-8	4.799831E-8
Сплайн 3-й степени					
	1.32403148	1.22353316	1.08197146	0.99314219	0.93739378
Периодический сплайн 5-й степени					
	0.00015686	8.210539E-5	3.017897E-5	1.508436E-5	9.461872E-6
Отрезок тригонометрического ряда Фурье					
	4.54369E-13	5.24434E-13	8.70482E-13	2.28253E-12	2.21896E-12
Метод наименьших квадратов на базе тригонометрических функций					
	3.38544E-13	1.20728E-14	3.74246E-15	3.98330E-15	5.44381E-15
Модификация эрмитовской аппроксимации					
	8.01354E-11	2.25109E-11	3.12853E-12	7.90877E-13	3.13389E-13

Из таблицы 1 видно, что наилучшим алгоритмом восстановления сеточного решения неперiodической задачи Дурфинга является модификация эрмитовской аппроксимации. При малом количестве точек разбиения отрезка аппроксимация отрезка ряда Фурье по полиномам Чебышева I рода может оказаться точнее, но это количество точек, как правило, не является оптимальным, поэтому предпочтительнее использовать сеточное решение неперiodической задачи модификацией эрмитовской аппроксимации. К тому же аппроксимация по полиномам Чебышева налагает ограничения на сетку. А на чебышевской сетке итерационные процессы плохо сходятся при большом числе точек аппроксимации производных, и разболтка в них наступает на 1 порядок раньше, чем на равномерной сетке.

Из таблицы 2 видно, что наилучшим алгоритмом восстановления сеточного решения периодической задачи Дурфинга является метод наименьших квадратов на