

Механизмы вычислительного интеллекта при решении задачи автоматизации прогнозирования электроэнергии

Р.С.Кузнецов, Ю.В.Тимофеев, Н.А. Смирнов, М.С. Тютяев, А.П.Черкис, Н.Л.Щербакова

Волгоградский государственный технический университет, Волгоград

Введение

В данной статье рассматривается решение проблемы автоматического прогнозирования потребления электроэнергии. Такая постановка задачи обусловлена следующими тенденциями: 1) удешевление процесса сбора информации о потреблении электроэнергии; 2) невозможности решения задач для большого количества объектов с привлечением экспертов в заданные (относительно короткие) сроки. Все это привело к необходимости разработки механизмов интеллектуальной обработки большого количества информации [1]. В статье предлагается подход основанный на инвариантном процессе интеллектуальной обработки данных (CrossIndustryStandardProcessforDataMining, CRISP DM) [2, 3]. Для реализации подходов разработана концепция интеллектуальной системы управления энергосбережением и реализован прототип [1, 4].

1. Прогнозирование потребления электроэнергии в автоматическом режиме

Предлагаемый подход автоматизации процесса прогнозирования потребления электроэнергии основан на CRISP DM и на оригинальных подходах применения коннективистских систем для прогнозирования потребления электроэнергии. В качестве объектов прогнозирования рассматривается широкий класс потребителей электроэнергии: жилые и нежилые помещения и здания различных типов. Данные о потреблении электроэнергии собираются и передаются на сервер сбора данных (система EcoSCADA, разработанная Бельгийско-Польской компанией Порта Капена) [5]. Фиксация информации о потреблении осуществляется в режиме реального времени каждые 15 минут. Дадим общее описание подхода.

Шаг 1. Добавление объекта потребителя. Данная процедура включает следующие основные шаги: 1) указание название объекта; 2) позиционирование объекта на карте и 3) загрузка данных о потреблении электроэнергии (исторические значения). Позиционирование на карте позволяет использовать информацию о месторасположении и получать информацию о погоде, которая используется для построения более точных и полных моделей прогнозирования. Для этого используются API погодных сервисов <http://www.worldweatheronline.com>, <http://www.noaa.gov/> и других.

Шаг 2. Предварительная обработка информации. Можно выделить две основные процедуры, связанные с оценкой качества данных. Первая процедура выявление и заполнение пропусков в данных, вторая выявление выбросов и аномалий.

Шаг 3. Автоматическое построение семейства прогнозирующих моделей, параметрическая оптимизация и выбор наиболее точной модели.

Шаг 4. Оценка качества прогнозирующих моделей и выбор наиболее адекватной модели.

Шаг 5. Реализация и визуализация результатов прогноза.

Рассмотрим метод более подробно, с выделением процедур обработки данных. Пусть в системе фиксируется информация о m зданиях, для каждого из которых необходимо сформировать прогноз на h значений вперед с таким же дискретным интервалом (15 минут). Формально информация о здании имеет представление

$$B = \{\{x\}, \{y\}, \{z\}\}, \quad (1)$$

где $\{x\}$ – информация о расположении здания, $\{y\}$ – информация о потреблении электроэнергии, $\{z\}$ – информация о погоде (температура, влажность, давление).

На первом шаге реализуются следующие процедуры. Процедура позиционирования здания на карте ВР. Для упрощения будем считать, что пользователь формирует отметку на карте (MicrosoftBing) и в $\{x\}$ содержатся значения широты и долготы. Для получения значений $\{y\}$ для произвольного здания на основании $\{x\}$ используется процедура веб-майнинга

$$WM(\{x\}, \{y\}) \quad (2)$$

Процедура работает с погодными сервисами.

На втором шаге реализована процедура выявления и заполнения пропусков в данных

$$GF(\{y\}, \{z\}) \quad (3)$$

Регрессионные модели и интерполяция кубическим сплайном используются в качестве базовых, реализованных в процедуре.

Процедура выявления выбросов и аномалий

$$OAD(\{y\}, \{z\}) \quad (4)$$

использует подход AD и ADAM [6].

Процедура

$$FMC(\{y^*\}, \{z^*\}, \{m\}) \quad (5)$$

синтеза и настройки моделей $\{m\}$ использует обработанные на предыдущих этапах данные

$\{y^*\}, \{z^*\}$. Исходя из предыдущих исследований [3, 4] входными переменными являются: 1)

потребление электроэнергии в предыдущие моменты времени, 2) информация о погоде (для обучения) и о прогнозе погоды (для расчета прогнозирования) и 3) календарные данные. В качестве вычислительного базиса используются коннективистские системы: конструктивные многослойные нейронные сети и эволюционные коннективистские системы Касабова на базе многослойных нейронных сетей. [7].

Сформированы 2 модели прогнозирования, отличающиеся набором входных переменных. Модель 1 включает 4 входные авторегрессионные переменные:

$$y_i(t - i*672), i = 1, \dots, 4. \quad (6)$$

Модель 2, помимо авторегрессионных, включает еще 3 входные переменные: режим работы здания (бинарная переменная, 0 – не работает, 1 – работает), день недели $\{1, \dots, 7\}$, и номер квартала $\{1, 2, 3, 4\}$.

Для оценки качества и выбора реализована процедура

$$QE(\{m\}, \{c\}) \quad (7)$$

оценки качества моделей на основе множества оценок $\{c\}$ [8]. Для расчета прогноза реализована процедура

$$CF(\{m\}, \{c\}) \quad (8)$$

2. Концепция интеллектуальной системы

Общее описание. Для реализации предложенного подхода и рассмотренных механизмов вычислительного интеллекта разработан прототип интеллектуальной системы.

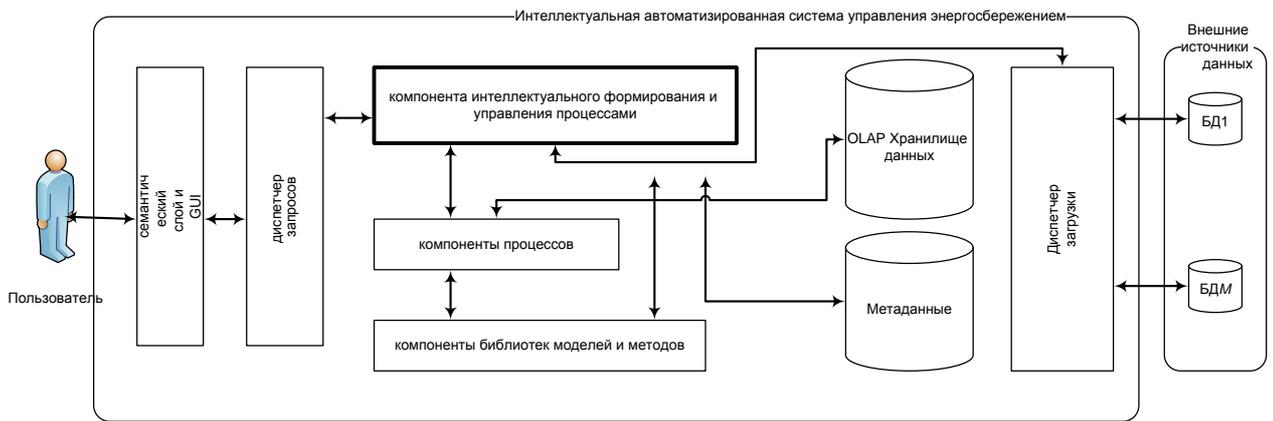


Рис 1. Архитектура интеллектуальной системы

Система имеет следующие компоненты (рис 1): 1) семантический слой для выбора (описания) проблемы; 2) диспетчер загрузки и определения качества данных (процедуры BP , V , O); 3) диспетчер запросов; 4) хранилище данных со структурой OLAP с диспетчером загрузки информации со внешних источников данных и с прослойкой, обеспечивающей доступ к хранилищу (например, объектная схема данных, $linq$); 5) компоненты библиотек моделей и методов; 6) компоненты библиотек процессов (процесс – последовательность вызовов методов для решения проблемы, процедуры F , G); метаданные или база знаний; компонента интеллектуального формирования и управления процессами и потоками потока метаданных.

Реализация в виде веб приложения и облачного приложения.



Рис 2. Диаграмма развертывания приложения

Облачная реализация представляет собой SaaS-систему, состоящую из хранилища данных со статистикой измерений и рассчитанными прогнозными значениями иоблачной интернет-службы WorkerCloudService, развернутых в облаке WindowsAzure, а также, непосредственно, системы автоматического прогнозирования, развернутой на обыкновенном хостинге и обращающейся к данным развернутого в облаке хранилища.

Приложения-клиенты, собирающие статистику измерений с датчиков в зданиях и помещениях, в автоматическом режиме с определенной периодичностью отправляют данные облачному сервису в виде XML-пакетов.

Структура XML-пакета представлена ниже:

```
<data>
  <userid>1875</userid>
  <num>148</num>
  <measures>
    <measure>
      <datetime>10:06:41 01.06.2008</datetime>
    </measure>
  </measures>
</data>
```

```

        <value>0,2</value>
    </measure>
    <measure>
        <date>10:26:41 01.06.2008</date>
        <value>0,8</value>
    </measure>
    ...
</measures>
</data>

```

Пакет данных содержит в себе идентификатор пользователя `tegeuserid`, число измерений в `tegenum` и, непосредственно, массив измеренных значений, заключенный в `tegemeasures`. Каждое измерение в массиве содержит в себе дату и время произведения измерения, а также полученное значение.

После отправки XML-пакет обрабатывается сервисом, который добавляет вновь полученные данные в хранилище, а также передает системе автоматического прогнозирования команду на обновление значений прогноза с учетом вновь загруженных данных.

Полученные прогнозные значения загружаются системой в хранилище данных, откуда, в свою очередь, их выгружает сервис и возвращает клиенту XML-пакет с результатами прогноза (структура XML с данными прогноза идентична структуре пакета с данными измерений).

Выводы

Для апробации предложенного подхода были проведены эксперименты с использованием информации за 2009 – 2010 годы о 17 коммерческих зданиях. Выборка данных была разделена на обучающую и тестовую. Все процедуры были реализованы в автоматическом режиме, средняя ошибка составила 11% (средняя относительная ошибка, MAPE).

Одним из направлений развития данной работы является формирование записей в таблицу метаданных для выявления закономерностей между данными и структурами наиболее точных моделей.

Литература:

1. Камаев, В.А. Интеллектуальные системы автоматизации управления энергосбережением / В.А. Камаев, М.В. Щербаков, А. Бребельс // Открытое образование. - 2011. - № 2. - С. 227-231.
2. Larose, D. T. Discovering knowledge in data: an introduction to data mining / Daniel T. Larose // John Wiley & Sons, Inc. ISBN 0-471-66657-2 (cloth). – 2005.
3. Камаев В. А., Щербаков М. В., Панченко Д. П., Щербакова Н. Л., Бребельс А. Применение коннективистских систем для прогнозирования потребления электроэнергии в торговых центрах / Управление большими системами. Выпуск 31. М.: ИПУ РАН, 2010. С.92-109.
4. Щербаков, М.В. Методика выявления потенциала энергосбережения на основе интеллектуального анализа данных / М.В. Щербаков, Т.А. Яновский, А. Бребельс, Н.Л. Щербакова // Прикаспийский журнал: управление и высокие технологии. - 2011. - № 2. - С. 51-55.
5. Hill, D.J., Minsker, B.S., Anomaly detection in streaming environmental sensor data: A data-driven modeling approach, Environ. Model. Softw. (2009), doi:10.1016/j.envsoft.2009.08.010

6. Коннективистские модели идентификации динамики систем на коротких интервалах наблюдения с заданным множеством классов поведения / М.В. Щербаков, Н.Л. Щербакова, И.П. Козлов, В.А. Камаев // Известия ЮФУ. Технические науки. - 2012. - № 1. - С. 83-91.
7. Brebels A., Shcherbakov M. (2011) Calculation method for the selection of the best scale- and domain- independent forecast model In The 31st Annual International Symposium on Forecasting (ISF 2011), 26-29 June, 2011, Prague, ISSN 1997-4, pp. 157. http://www.forecasters.org/isf/pdfs/ISF11_Proceedings.pdf