

УДК 519.6

ОСОБЕННОСТИ ПРОГРАММНОЙ РЕАЛИЗАЦИИ ПРОГНОСТИЧЕСКОЙ СИСТЕМЫ КАФЕДРЫ КСМ

Сивяков А.С., Беловодский В.Н.

Донецкий национальный технический университет

E-mail: arteom.sivyakov@gmail.com

В настоящей работе описан макет программной системы для краткосрочного прогнозирования климатических данных, собранных с метеостанции установленной на кафедре КСМ. Изложена модификация прогностического комплекса и особенности его усовершенствования. Описан подход к использованию метода главных компонент в существующей программной системе. Изложена теория и практика использования метода при анализе и прогнозировании временных рядов.

Актуальность

Еще великий Михаил Васильевич Ломоносов заметил: «Человеку ничего не оставалось бы требовать от бога, если бы он научился правильно предсказывать погоду».

Национальные метеорологические и гидрологические службы выпускают ежегодно десятки миллионов прогнозов погоды. Прогнозы погоды являются одним из видов научного предвидения. Для улучшения качества этих прогнозов требуются комплексные и хорошо обслуживаемые метеорологические сети, стандартизация и координация обмена данными и продукцией при высоких скоростях, а также разработка и применение новых методов наблюдений и моделирования.

За последние несколько десятилетий благодаря улучшению средств наблюдений, возрастающему научному пониманию и более современным и сложным численным моделям, а также другим прогностическим механизмам, качество прогнозов погоды

значительно улучшилось, что привело к тому, что изменилось восприятие их населением. В самом деле, прогнозы на трое суток в настоящее время так же хорошо оправдываются, как прогнозы на сутки, которые выдавались 20 лет назад.

Однако несмотря на то, что теория и практика прогнозирования существуют и развиваются уже давно, несмотря на успехи, составление прогнозов погоды до сих пор остается очень трудной, инвариантно решаемой задачей. В гидрометеорологических прогнозах все еще остается неопределенность.

Обзор предметной области исследования

Прогнозистам жизненно необходимо идти в ногу с передовыми научными достижениями. За последние несколько десятилетий, благодаря значительному развитию науки, появились улучшенные и более эффективные методы и инструменты для проведения наблюдений, своевременного сбора, анализа и прогнозирования данных.

Одним из таких инструментов для своевременного сбора и анализа данных является многофункциональная метеорологическая станция Vantage Pro2. Vantage Pro II предназначен как для профессионального, так и для бытового применения, и имеет функцию прогноза погоды, которая учитывает координаты места расположения комплекса, время года, текущее значение атмосферного давления и его изменение, скорость и направление ветра, температуру и влажность воздуха, количество осадков. Эти параметры измеряются датчиками, которыми укомплектована метеостанция.

Анализ последних исследований

На сегодняшний день Vantage Pro II установлена на кафедре КСМ, факультета КНТ, ДонНТУ. Все данные сохраняются на сервере кафедры КСМ и «АКИАМ». Интервал замеров составляет 10 минут. Таким образом, в процессе его эксплуатации постоянно формируется и постепенно накапливается совокупность временных

рядов. Наличие этой информации делает вполне реальной задачу разработки прогностического комплекса. В связи с этим, для указанной выше метеостанции, в 2009 году, магистрантом кафедры КСМ была разработана программная система [1]. Работу данной системы можно представить схемой на рис. 1.

Работа была выполнена с использованием последних математических методов и современных инструментов моделирования. Тестирование и анализ существующего прогностического комплекса позволили сделать следующие выводы, связанные с доработкой данного приложения:

- работа по подготовке исходных данных занимает время, что снижает автоматизацию процесса прогнозирования. При имеющихся на кафедре КСМ возможностях, а именно непосредственного доступа к метеостанции, можно проводить обработку данных в постоянном режиме времени, делая выборку данных по временному параметру и проводя их автоматизированную подготовку;
- обзор литературных и интернет источников показал, что на сегодняшний день в области изучения и анализа климатических данных широкое распространение и использование получил метод главных компонент, что в свою очередь дает возможность в будущем улучшить



Рисунок 1 – Структурная схема работы программной системы прогнозирования метеоданных

качество прогноза. Поэтому возникает идея, на этапе определения размерности прогностической модели, перейти от метода ложных соседей к методу главных компонент;

- было предложено улучшить интерфейс путем добавления новых инструментов (графическое отображение загруженных данных, сохранение данных в файлы разных форматов, выбор тип доступа к базе данных и др.) и разбиения интерфейса на два вида – пользовательский и исследовательский. В свою очередь пользовательский интерфейс позволяет только формировать прогноз, а исследовательский – формировать прогноз, а так же проводить изменение параметров прогностической модели и оценивать при этом качество прогноза;
- для запуска прогностического комплекса необходимо производить предварительную установку и настройку инструментальных средств, а это в свою очередь делает открытым доступ к исходным кодам программной системы. Поэтому было решено преобразовать разработанный программный проект в исполняемый файл.

Все выше сказанное делает доработку и оптимизацию программной системы весьма актуальной.

Обработка данных в постоянном режиме времени

Для обработки данных в постоянном режиме времени, средствами MatLab, была разработана и реализована схема (рис. 2), основанная на принципе клиент-серверного управления базами данных.

Работа прогностической системы по данной схеме позволяет осуществлять доступ к данным в непрерывном времени многочисленного числа пользователей. Таким образом, это делает более универсальной и мобильной разрабатываемую систему.

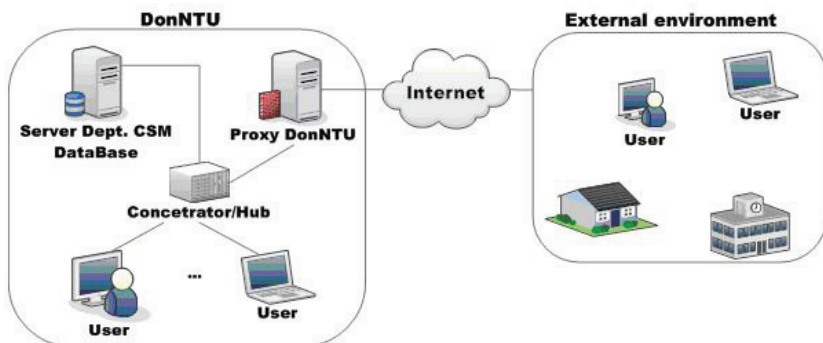


Рисунок 2 – Обработка данных в постоянном режиме времени

Использование метода главных компонент

В последние десятилетия в рамках нелинейной динамики получен ряд фундаментальных теоретических результатов и разработаны методики, обосновывающие принципиальную возможность прогнозирования физических процессов на базе их временных рядов. Теоретическим фундаментом этих разработок и методов является теорема Такенса [2]. Одной из его основополагающих идей является то, что при построении эмпирических моделей по временному ряду в качестве недостающих переменных можно использовать или последовательные значения доступной наблюдаемой величины, или ее последовательные производные. Было доказано, что при реконструкции по скалярной временной реализации динамической системы и метод временных задержек, и метод последовательных производных гарантируют, что в новых переменных будет получено эквивалентное описание исходной динамической системы при достаточно большой размерности восстановленных векторов D . А именно, должно выполняться условие $D > 2d$, где d – размерность множества M в фазовом пространстве исходной системы, на котором происходит моделируемое движение. Эти утверждения и составляют содержание знаменитых теорем Такенса.

На сегодняшний день существуют различные методы

оценки размерности модели: метод ложных соседей, метод главных компонент, метод Грассбергера – Прокаччия и др. Все они имеют свои особенности работы.

Анализ литературных и интернет источников показал, что из указанных выше методов оценки размерности и изучения климатических данных, широкую популярность получил метод главных компонент.

Метод главных компонент (Principal component analysis, PCA) - один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации.

Базовый вариант метода состоит в преобразовании одномерного ряда в многомерный с помощью однопараметрической сдвиговой процедуры, исследовании полученной многомерной траектории с помощью анализа главных компонент (сингулярного разложения) и восстановлении (аппроксимации) ряда по выбранным главным компонентам. Результатом применения метода является разложение временного ряда на главные компоненты и отбрасывания несущественных (компоненты с наименьшей выборочной дисперсией), а так же дальнейшее использование, отобранных главных компонент, в качестве новых переменных для аппроксимации временного ряда (прогнозирования) [3].

Алгоритм данного метода можно описать следующими этапами:

1. Развертка одномерного ряда в многомерный. Пусть $(f_i)_{i=1}^N$ - числовая последовательность. Тогда многомерный ряд можно представить следующим образом

$$V = \begin{pmatrix} f_1 & f_2 & \dots & f_n \\ f_2 & f_3 & \dots & f_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_\tau & f_{\tau+1} & \dots & f_N \end{pmatrix},$$

где N – размерность временного ряда;

n – параметр задаваемый исследователем либо разработчиком;

τ - рассчитываемый параметр. Определяется как $\tau = N - n + 1$.

2. Анализ главных компонент: сингулярное разложение выборочной ковариационной матрицы. Этот этап состоит в вычислении собственных чисел и собственных векторов матрицы V , т.е. разложение ее $V = PLP^T$, где L - диагональная матрица, на диагонали которой стоят упорядоченные по убыванию собственные числа и представляющие собой выборочные дисперсии собственных векторов матрицы P , а P - ортогональная матрица собственных векторов матрицы V . Вектора матрицы P и представляют собой главные компоненты.

3. Отбор главных компонент. Этот этап в большей степени является эмпирическим, а следовательно требует дальнейшего исследования и опытов. Идея данного этапа основана на анализе матрицы собственных чисел. Так как собственные числа представляют собой выборочные дисперсии собственных векторов, то отбор собственного числа, согласно заданному правилу, сопровождается отбором соответствующего собственного вектора.

4. Переход к многомерному ряду меньшей размерности (переход к новым переменным). Реализуется путем повторного перемножения исходной матрицы на отобранные собственные вектора.

На момент написания статьи метод главных компонент был внедрен в прогностический комплекс, однако переход к новым переменным привели к задаче корректного их использования в нейронных сетях, в частности формирование обучающих и тестирующих выборок.

Функциональность прогностического комплекса

В общем виде работу с прогностическим комплексом можно представить схемой изображенной на рисунке 3. Как видно из рисунка, в работе прогностического комплекса участвуют два актёра – «Администратор/разработчик» (далее по тексту администратор) и «Пользователь/Исследователь». Введение актёра – «Администратор» выполнена с целью полноценного функционирования программной

системы, актуализации данных в БД и поддержки работы сервера.

Необходимо так же отметить, что неоднозначное определение актера «Пользователь/Исследователь» обусловлено необходимостью разделения интерфейса на два – пользовательский и исследовательский. Режим пользовательского интерфейса обеспечивает задание момента и дальности прогноза, а так же просмотр результата. Создание исследовательского интерфейса выполнено с целью выполнения варьирования входных параметров и исследования их влияния на итоговый результат, а так же дальнейшего анализа метеопараметров.

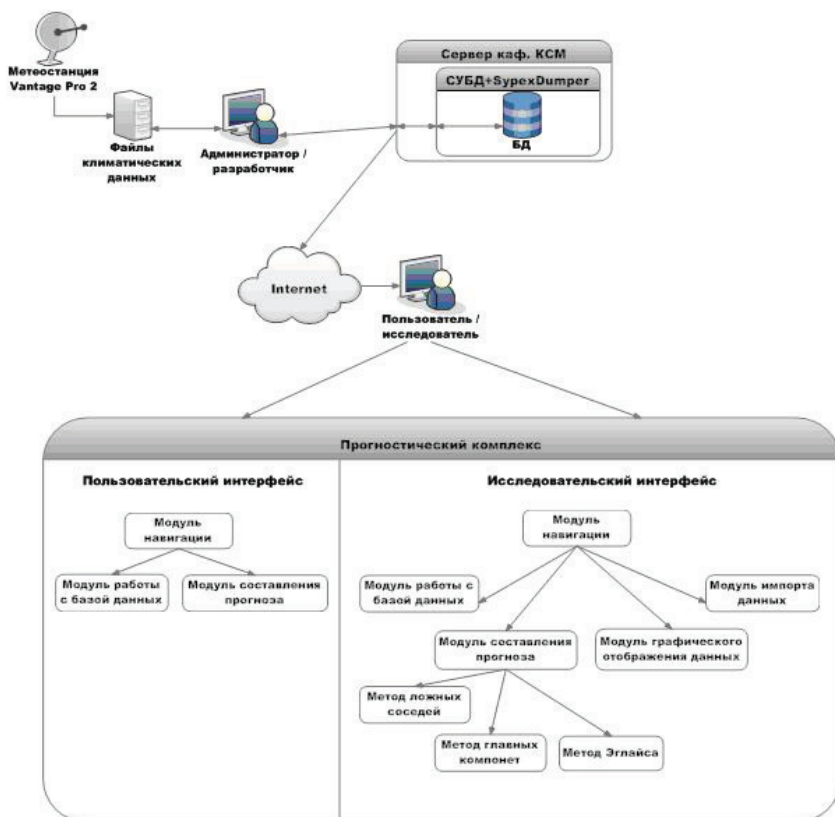


Рисунок 3 – Обобщенная схема работы прогностического комплекса

Переход к исполняемому файлу

На момент написания статьи данная задача еще не была реализована. Однако необходимо отметить, что переход к исполняемому файлу будет осуществляться средствами компилятора MatLab и языка программирования C++. Компилятор MatLab позволяет из m-функций создавать автономные приложения, C и C++ библиотеки совместного использования.

Решение данной задачи позволит сделать прогностический комплекс более мобильным и решит проблему открытого кода.

Литература

- [1] Гриценко А.В. Реконструкция уравнений и прогнозирование метеопараметров по их временным рядам. – Донецк, ДонНТУ, 2010. – 149 с.
- [2] Takens F. Detecting strange attractors in turbulence // Lec. Notes in Math., 1981. V. 898. P. 366-381.
- [3] Ефимов В.М., Галактионов Ю.К., Шушпанова И.Ф. Анализ и прогноз временных рядов методом главных компонент. – Новосибирск: Наука. Сиб. отделение, 1988. – 71 с.