

06, июнь 2016

УДК 004.81

Сравнительный анализ методов обработки временных рядов

*Высочанский В.А., студент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Системы обработки информации и управления»*

*Научный руководитель: Тоноян С.А., к.т.н, доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Системы обработки информации и управления»
chernen@bmstu.ru*

Введение

В настоящее время задача распознавания образов лежит в основе большинства интеллектуальных систем, начиная от беспилотных транспортных средств и заканчивая медицинскими экспертными системами. Данная задача является междисциплинарной и задействует одновременно несколько областей знания: анализ данных, когнитивное познание, искусственные интеллектуальные системы и др.

Обработка временных рядов относится к одним из наиболее сложных задач распознавания образов. Классический пример подобной задачи – распознавание слитной речи, представленной в виде акустических и звуковых моделей. Работа с временными рядами усложняется за счёт необходимости фильтрации шумов, нестационарного характера ряда, частой необходимости быстрой обработки и т.д. Таким образом, среди существующего множества методов и технологий распознавания образов необходимо выделить те, которые наиболее приспособлены для анализа временных рядов.

1. Задачи, связанные с применением временных рядов

Временной ряд – это последовательность регистрируемого сигнала за несколько моментов или периодов времени. Такой ряд строится в том случае, когда невозможно описать динамический процесс некоторым уравнением функции [1].

Как видно из определения, временной ряд наиболее удобно представить в виде графика в двумерном пространстве, на одной оси которого отражено время, а на другой – значения некоторого показателя. На формирование этого показателя (уровня) оказывают

влияние следующие факторы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

Одна или несколько разновидностей перечисленных выше факторов характерна в той или иной степени различным временным рядам. Так, например, тенденция свойственна экономическим рядам – она возникает из-за того, что множество разнородных факторов в течение долгого времени влияют на один показатель и образуют функцию времени, называемую трендом. Циклические колебания также присущи рядам, связанным с финансово-экономической деятельностью, поскольку они часто выражаются в сезонных «скачках».

Временные ряды применяются в качестве аналитических моделей для решения множества задач, среди которых стоит отметить следующие:

1. распознавание речи – задача классификации образов акустических характеристик речевых сигналов, состоящая из двух подзадач [2]:

а. предварительная обработка речевых сигналов для выделения информативных признаков: фильтрация речевого сигнала, нарезка полученного сигнала перекрывающимися кадрами, вычисление кратковременной энергии и интенсивности всех перекрывающихся кадров;

б. классификация речевых сигналов при помощи сравнения с эталонными акустическими образами на основе, как правило, вейвлет-коэффициентов.

2. моделирование процессов болезней пациентов – задача анализа многомерного стохастического нестационарного временного ряда с нелинейными трендами, в которой тяжело собрать достаточно полную статистику [3];

3. прогнозирование финансовых рядов – задача, связанная с выявлением сложных закономерностей (не обнаруживаемых линейными методами) в нестационарных и быстро меняющихся рядах [4];

4. экологическое диагностирование и прогнозирование – задача анализа поведения экологических объектов, не имеющих полного аналитического описания и образующих во времени нестационарный многомерный случайный процесс [3];

5. адаптивное управление сложным динамическим слабо структурированным объектом в условиях неопределённости.

Таким образом, временные ряды являются единственной возможной математической моделью для большого числа процессов, требующих обработки. Несмотря на схожесть многих рядов, для каждой конкретной задачи существуют наиболее эффективные методы

анализа.

2. Обзор классических методов анализа временных рядов и выявление их недостатков

В качестве примера были выбраны финансовые временные ряды, для которых необходимо предсказывать цены, используя статистические данные за предыдущие периоды. Особенности таких рядов являются сильная зашумлённость, влияние широкого спектра постоянно меняющихся факторов на формирование цен и периодические резкие потрясения рынков. Пример финансового ряда, отражающего отношения курса доллара США к рублю за период в несколько месяцев 2015-2016 годов, показан на рисунке.



Таким образом, построение адекватной модели финансового ряда – это сложная многокритериальная задача, для которой в настоящий момент не найдено достаточно эффективного решения [4].

Методы спектрального анализа основываются на представлении наблюдаемого сигнала в виде суперпозиции некоторых функций-компонент [3]. Наиболее ярким представителем таких методов является преобразование Фурье. Как известно, суть разложения в ряд Фурье заключается в представлении произвольной сложной функции как суммы более простых функций. При решении задач таким методом осуществляется переход от временного ряда к совокупности случайных величин. Значения этих величин представляются в виде векторов в евклидовом пространстве. При проведении экспериментов с заранее известными результатами каждому состоянию исследуемой системы назначается набор векторов. Далее, при анализе системы в условиях неопределённости получившийся вектор относится к той или иной группе векторов (состоянию системы) по методу наименьших расстояний.

Недостатками данной совокупности методов являются:

– выборочные средние групп векторов, описывающих разные состояния системы, должны также различаться, что нехарактерно для финансовых рядов;

– невозможно представить непериодический случайный процесс, коим является практически любой финансовый ряд.

Таким образом, методы спектрального анализа неприменимы к обработке финансовых рядов, а также других рядов экономического происхождения.

Другим классическим методом обработки временных рядов является метод сингулярного спектрального анализа (метод главных компонент), разработанный для исследования автокорреляционной зависимости. Он заключается в использовании скользящего окна размера L и некоторой выборке, оформленной в виде матрицы, зависящей от размера окна. После нахождения ранга матрицы происходит сингулярное разложение – определяется базис в L -мерном пространстве (главные компоненты), в результате чего получается стационарный временной ряд, определить характеристики которого не составляет труда [3]. Иными словами, суть метода заключается в преобразовании одномерного временного ряда, который может быть нестационарным (как финансовые ряды), в несколько многомерных стационарных рядов.

В сравнении со спектральными методами, к которым относится, в частности, преобразование Фурье, метод сингулярного анализа более приспособлен к обработке финансовых рядов. Именно поэтому он часто используется для, например, определения вероятного поведения цен в ближайшем будущем. В этом случае каждая цена в фиксированной точке времени является состоянием системы. Серия таких состояний во времени образует процесс, характеристики которого неизвестны. Метод главных компонент раскладывает этот процесс на сумму отдельных составляющих (паттернов), каждый из которых даёт информацию о тренде, шумовых и осциллирующих компонентах.

К недостаткам этого метода относятся:

- сложность реализации;
- большая длительность расчётов;
- отсутствие аналитического модельного представления ряда.

Экспертные методы являются следующей разновидностью инструментов обработки временных рядов. Они основываются на субъективной оценке текущего момента и перспектив развития, причём каждая оценка может быть получена в одной из следующих форм:

- точечный прогноз – наиболее простая и наименее эффективная форма, выраженная в прогнозе конкретного показателя;

– интервальный прогноз – предполагает наличие границ, внутри которых лежит прогнозируемое значение показателя с заданным уровнем значимости;

– прогноз распределения вероятностей – связан с определением вероятности попадания фактического значения показателя в один из нескольких интервалов.

Интервалы, с которыми имеют дело последние две формы экспертной оценки, обычно делят на пессимистический (низкий уровень исследуемого показателя), наиболее вероятный (средний уровень показателя) и оптимистический (высокий уровень показателя). Очень важным моментом в данной группе методов является обобщение оценок каждого из экспертов. Для этого, как правило, используются следующие методы взвешивания по критериям:

– равные веса в случае, если эксперты имеют одинаковые компетентности;

– веса, пропорциональные степени важности мнений экспертов;

– веса, пропорциональные уровням самооценки экспертов;

– веса, пропорциональные точности последних прогнозов экспертов.

Таким образом, экспертные методы в чистом виде не могут дать по-настоящему объективного прогноза временного ряда. Для устранения проблемы взвешивания индивидуальных прогнозов удобно использовать метод Дельфи, сближающий точки зрения экспертов путём последовательного объединения идей. Собранные оценки подвергаются статистической обработке, после чего вновь предоставляются экспертам для уточнения оценок. Иными словами, на каждой итерации эксперты могут изменить свои оценки, приближая их таким путём к объективным мнениям.

Несмотря на то, что экспертные методы достаточно универсальны (их можно применять для обработки любых временных рядов), они сильно зависят от компетентности участвующих в оценке экспертов. Кроме того, для некоторых разновидностей финансовых временных рядов, критичных ко времени анализа, экспертные оценки могут занимать слишком длительное время.

Реляционный подход также применяется для обработки временных и, в частности, финансовых временных рядов [4]. Суть его заключается в том, что параллельно с обнаружением закономерностей, предсказывающих значения ряда, обнаруживаются также высоковероятные закономерности, описывающие динамику ряда на обучающем интервале, чтобы в дальнейшем можно было выявлять нарушения этих зависимостей. Высоковероятные закономерности описывают нормальную динамику ряда, тогда как нарушения – аномальное поведение ряда. Важно отметить, что именно такие спонтанные изменения тренда характерны большинству финансовых рядов – например, к ним может относиться обвал рынка по причине какого-либо чрезвычайного происшествия, и т.д.

Получается, что в случае обнаружения аномалии данный метод указывает на невозможность дальнейшего применения составленного прогноза, в то время как другие методы продолжали бы работать по одному и тому же алгоритму.

Реляционный подход состоит из двух основных методов, работающих параллельно:

1. метод обнаружения вероятностных закономерностей, основанный на методологии семантического вероятностного вывода [4], заключается в последовательном наращивании условной части правил с проверкой выполнимости условий;

2. метод обнаружения нарушений динамики, в котором происходит сравнение гипотез нормального протекания процесса с показаниями системы на анализируемом временном интервале.

Как показывает практика [4], реляционный подход хорошо показывает себя в задачах прогнозирования финансовых временных рядов. Во многом этом достигается за счёт наличия метода обнаружения нарушений динамики, поскольку подобные ряды склонны к влиянию случайных факторов. Тем не менее, к недостаткам данного метода можно отнести использование ресурсоёмкого и негибкого алгоритма поиска закономерностей.

3. Анализ применение генетических алгоритмов и нейронных сетей для выявления скрытых закономерностей

Одной из наиболее перспективных в настоящее время групп методов обработки временных рядов является техническая диагностика. Эти методы способны решать задачи распознавания состояний системы, обнаружения нарушений работоспособности, а также установления вида и места дефекта. Так или иначе, многие технологии из этой группы основаны на алгоритмах минимизации ошибки – следовательно, необходимо наличие обучающей выборки с заранее известными состояниями. Самый распространённый метод технической диагностики связан с применением нейронных сетей и алгоритма обратного распространения ошибки для поиска оптимальных весовых коэффициентов связей сети [5]. Однако у данного подхода имеется ряд недостатков, не позволяющих в полной мере использовать его для предсказания состояний финансовых рядов:

- очень долгий процесс обучения;
- возможный «паралич сети», заключающийся в невозможности продолжения процесса обучения;
- незащищённость алгоритма от попадания в локальный минимум, следствием чего является ошибочный прогноз.

Эти недостатки преодолеваются путём использования генетических алгоритмов для обучения (настройки весов) или поиска архитектуры (числа нейронов на каждом слое) нейронной сети. Сравнительно недавно эта технология получила широкое распространение [6] благодаря простой реализации, практически неограниченным возможностям масштабирования и распараллеливания вычислений, а также отсутствия перечисленных выше недостатков.

Важной особенностью генетических алгоритмов является их нечувствительность к размерности множества оптимизации. Это означает, что вычислительная сложность слабо возрастает при увеличении объёма обрабатываемых данных. В финансовых рядах, связанных с отношениями курсов различных валют, такое качество алгоритмов обработки является крайне полезным, т.к. приходится строить прогнозные модели сразу для огромного количества слабосвязанных временных рядов (по одному на каждое соотношение валют).

Кроме того, генетические алгоритмы поддаются гибкой настройке благодаря множеству изменяемых параметров, среди которых:

- длина хромосомы (влияет на количество решений за одну итерацию);
- наполнение хромосомы;
- параметры оператора кроссовера;
- параметры оператора мутации;
- параметры оператора инверсии;
- тип алгоритм выбора лучших особей и его параметры;
- параметры генерации начальной популяции.

Следовательно, можно предположить, что для конкретной задачи выявления закономерностей временного ряда возможно подобрать набор некоторых оптимальных параметров генетического алгоритма, обучающего нейронную сеть. Например, если целью является быстрый подбор весов сети с достаточной (но не максимальной) точностью, то следует использовать турнирный алгоритм выбора лучших особей, малый процент мутации и среднюю длину хромосомы.

Как было отмечено ранее, саму задачу прогнозирования вероятностных закономерностей (нормального состояния) временного ряда способны решать нейронные сети. При этом предпочтительным является выбор в пользу трёхслойных нейронных сетей, поскольку они позволяют реализовать разделение любых множеств посредством нескольких гиперплоскостей в пространстве состояний [7]. В рамках поставленной задачи обработки финансовых рядов это означает, что нейронная сеть будет способна обучаться

на множествах произвольного характера и, следовательно, адекватно составлять прогноз.

Выводы

Задача обработки временных рядов, будь то ряд финансово-экономических показателей, характеристики работы промышленного объекта или кардиограмма пациента, остаётся актуальной в настоящее время в виду трудностей практической реализации многих методов, хорошо описанных в теории: большинство из описанных технологий либо ориентированы на решение узкоспециализированных задач, либо слишком требовательны к вычислительным ресурсам.

Из множества известных методик выделены некоторые методы технической обработки рядов, связанные с использованием нейронных сетей и генетических алгоритмов. В случае объединения положительных качеств реляционного подхода (поиск нарушений динамики) с гибкостью и простотой реализации генетических алгоритмов, а также способностями прогнозирования многослойных нейронных сетей, возможно создание нового эффективного метода обработки временных рядов.

Список литературы

- [1]. Лоскутов А.Ю. Анализ временных рядов: курс лекций. Режим доступа: http://chaos.phys.msu.ru/loskutov/PDF/Lectures_time_series_analysis.pdf (дата обращения 20.02.2016).
- [2]. Панченко Д.П., Нгуен Л.В. Распознавание речи с использованием вейвлет-преобразования // Альманах современной науки и образования. Тамбов: Грамота, 2011. № 5. С. 60-64.
- [3]. Букреев В.Г., Колесникова С.И., Янковская А.Е. Выявление закономерностей во временных рядах в задачах распознавания состояний динамических объектов. Томск: Издательство Томского политехнического университета, 2011. 254 с.
- [4]. Демин А.В., Витяев Е.Е. Финансовые временные ряды: прогнозирование и распознавание нарушений динамики. Новосибирск: Институт систем информатики СО РАН, 2009. 8 с.
- [5]. Шумков Е.А., Чистик И.К. Использование генетических алгоритмов для обучения нейронных сетей. Краснодар: Научный журнал КубГАУ, № 91(07), 2013. 10 с.
- [6]. Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. Новосибирск: НГУ, 2006. 293 с.
- [7]. Местецкий Л.М. Математические методы распознавания образов: курс лекций. М.: МГУ, ВМиК, 2004. 85 с.