

Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.657

Ляховец
Пётр Владимирович

Система интегрированного планирования
и прогнозирования финансовых показателей

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники
по специальности 1-40-81-03 Искусственный интеллект

Научный руководитель

Гулякина Наталья Анатольевна

Кандидат физико-математических наук, доцент

Минск 2015

Введение

Задача прогнозирования будущих значений временного ряда на основе его исторических значений является основой для финансового планирования в экономике и торговле, планирования, управления и оптимизации объемов производства, складского контроля. Она актуальна и является неотъемлемой частью ежедневной работы многих компаний.

Основная задача разрабатываемого модуля системы – получить на выходе прогноз финансовых показателей для каждого конкретного нового объекта приходящего в систему. Для этого каждый новый объект должен быть соотнесен с каким-либо кластером по обобщенному образу которого и будет построен график.

Общая характеристика работы

Актуальность темы диссертации. Оптимизация методов прогнозирования, включающая снижение человеческих трудозатрат и повышение точности и достоверности прогнозирования является одной из важнейших областей автоматизации не только в компаниях банковской и финансовой сфер.

Цель и задачи исследования. Целью работы является повышение качества прогнозирования финансовых показателей и снижение трудозатрат на прогнозирование приходно-расходных финансовых показателей.

Объект исследования. Объектом исследования является система финансового планирования, прогнозирования и управления контрактами.

Предмет исследования. Предметом исследования являются методы прогнозирования финансовых показателей в условиях неполных данных и методы кластеризации объектов.

Личный вклад магистранта. Постановка задач, выбор подходов, теоретические разработки, экспериментальная работа и программирование выполнены лично автором работы.

Структура и объем диссертации. Диссертация состоит из введения, общей характеристики работы, 3-х глав, заключения и списка использованной литературы. Объем работы составляет 61 стр., включая 1 рис. на 1 стр., 9 табл. на 10 стр. и список использованных источников из 50 наименований, в том числе 28 иностранных на 6 стр.

Доклад по теме работы был представлен на 50-й научной конференции БГУИР (2014 г.) в секции «Интеллектуальные информационные технологии».

Краткое содержание работы

В **Главе 1** рассматриваются современные методы и подходы к прогнозированию и кластеризации.

В настоящее время насчитывается свыше 100 классов моделей прогнозирования. Число общих классов моделей, которые в тех или иных вариациях повторяются в других, гораздо меньше. Часть моделей и соответствующих методов относится к отдельным процедурам прогнозирования. Часть методов представляет набор отдельных приемов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения.

Для решения задач, требующих изучения отношения между двумя и более переменными, используется регрессионный анализ. В настоящее время регрессия получила широкое применение, включая задачи прогнозирования и управления. Целью регрессионного анализа является определение зависимости между исходной переменной и множеством внешних факторов (регрессоров). При этом коэффициенты регрессии могут определяться по методу наименьших квадратов или методу максимального правдоподобия.

В основу авторегрессионных моделей заложено предположение о том, что значение процесса $Z(t)$ линейно зависит от некоторого количества предыдущих значений того же процесса $Z(t-1)$, ..., $Z(t-p)$.

Модель экспоненциального сглаживания применяется для моделирования финансовых и экономических процессов. В основу экспоненциального сглаживания заложена идея постоянного пересмотра прогнозных значений по мере поступления фактических. Эта модель присваивает экспоненциально убывающие веса наблюдениям по мере их старения. Таким образом, последние доступные наблюдения имеют большее влияние на прогнозное значение, чем старшие наблюдения.

Классификация — один из разделов машинного обучения, посвященный решению задачи определения классовой принадлежности произвольного объекта из исходного множества, при наличии обучающей выборки.

Кластерный анализ — задача разбиения заданной выборки объектов на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

Определение множества переменных — выбор признаков для кластеризации, которые являются наиболее информативными для решения конкретной задачи. На одном и том же наборе данных для разных задач информативными признаками будут являться различные множества признаков. Выбор адекватного множества признаков (фильтрация) — отдельная задача призванная решить набор вопросов по адекватности кластеризации, также как и требуемой производительности ЭВМ и необходимым объемам памяти.

Установлено, что к настоящему времени разработано множество моделей для решения задачи прогнозирования временного ряда, среди которых наибольшую применимость имеют авторегрессионные, нейросетевые модели, и модели экспоненциального сглаживания.

Выявлены достоинства и недостатки рассмотренных моделей. Установлено, что существенным недостатком авторегрессионных моделей является большое число свободных параметров, требующих идентификации; недостатками нейросетевых моделей является ее непрозрачность моделирования и сложность обучения сети. И в текущих реалиях системы наиболее

оправданным будет применение моделей экспоненциального сглаживания с учетом сезонности и трендов.

Определено, что наиболее перспективным направлением развития моделей прогнозирования с целью повышения точности является создание модели, выполняющей на первом этапе кластеризацию, а затем прогнозирование временного ряда для объекта внутри установленного кластера с учетом усредненных по кластеру функций для сезонности и тренда функции.

В **Главе 2** описывается модель решения задачи. Приводится общая схема и краткий анализ системы. Производится анализ этапов решения задачи, и проектируется гибкая модель решения.

Решение задачи можно разделить на этапы:

- Кластеризация
 - Подбор набора признаков
 - Подбор меры подобия
 - Подбор параметров кластеризации
 - Разбиение множества объектов на кластеры
- Синтез образа
 - Оценка сезонности
 - Оценка тренда

- Проверка результатов

Далее итеративно прогоняются все этапы решения по разным наборам входных данных (наборам признаков объектов, мерам подобия и параметрам алгоритма кластеризации), и анализируются результаты с целью выявления наилучшей комбинации.

В **Главе 3** описывается программная реализация алгоритмов и приводится оценка полученных результатов. Приводится описание аппаратно-программного обеспечения и критерии их выбора. Описываются этапы проведения испытаний и приводятся их результаты.

В результате проведенных исследований выявлено, что наиболее точные прогнозы получены при таком наборе параметров контракта: TCV, продолжительность контракта (DUR), страна контрагента (SC) и принадлежность контракта к конкретному финансовому отделу (PRC).

Слабые кластеры следует исключать из дальнейшей обработки т.к. создаваемая в них усредненная модель объекта будет нерепрезентативна.

Определено, что степенные меры расстояния в большинстве случаев дают одинаковые (или крайне схожие) результаты, а качество кластеризаций, использовавших манхэттенское расстояние, значительно ниже.

Также намечаются дальнейшие пути усовершенствования системы, ключевые потенциальные проблемы и примерные пути решения этих проблем при работе в промышленной системе на больших объемах данных.

Заключение

В результате проведенной работы разработан модуль автоматического прогнозирования финансовых показателей объекта, выявлен набор свойств контракта наиболее сильно отображающих его схожесть/отличие относительно других контрактов.

Разработанный модуль позволяет системе освободить работников от рутинной работы – как следствие сокращение человеко-часов что ведет к снижению затрат предприятия.

Также модуль имеет высокую точность прогнозирования. Как показано ранее, модуль предоставляет статистически более точные прогнозы, чем существовавшие ранее авто-прогнозы, что существенно влияет на качество управления финансовыми потоками и рисками, что обычно ведет к более правильным и взвешенным решениям по управлению предприятием, и, как следствие, более высоким прибылям из-за точной стратегии.

Для запуска использования модуля планируется добавить для изначального анализа еще несколько методов кластеризации и провести ряд испытаний с различными наборами параметров кластеризации и на основе данных с других региональных серверов. Намеченные пути повышения надежности прогноза:

- Увеличение параметра «мощность исключений» позволит отбросить “слабые” кластеры, которые ввиду небольшого количества объектов все-равно не предоставят достаточно данных для формирования достоверного образа.
- Увеличение количества кластеров (как минимального, так и максимального) позволит, при использовании на намного бóльших массивах, создать кластеры хоть и меньшие по мощности (что может негативно сказаться на репрезентативности данных), но с более высокой плотностью, т.е. будут объединять более схожие объекты, что должно привести к увеличению точности прогнозирования.

Предложенная методика и ее реализация не привязана к объекту или платформе, таким образом может быть легко адаптирована под различные объекты исследования и перенесена на различные системы (в контексте реализации данный пункт особенно важен, т.к. DB2 является проприетарной СУБД).

Приведенная методика прогнозирования представляет из себя перспективное направление разработки, тем более ввиду общей тенденции к машинной аналитике больших объемов данных.