**Тема:** **Разработка и исследование моделей прогнозирования основанных на технологи нейронных сетей в информационной системе электронного правительства.**

**Студент**: Гусейнов Рустам Ровшан оглы

**Группа: R6222**

**Научный руководитель**: доцент Салимов Вагиф Гасан оглы

В соответствии с планом диссертации, в первую очередь был проведен анализ обзоров существующей литературы, их ограничений и варианты решения этих ограничений. Актуальность этой темы обуславливается бурным развитием информационных технологий и возрастающей необходимостью их интеграции в институты управления, в том числе, государством.

Далее подчеркнул необходимость глубокого анализа политической природы процессов развития электронного правительства и учет сложной политической и институциональной среды. Затем были рассмотрены примеры интеграции технологий в государственное управление США, а также три ключевых события, ставшие основополагающими для этой интеграции:

1) Проект "Городские информационные системы"(URBIS), в рамках которого исследовалось влияние компьютеров на организации обслуживания. URBIS - первое систематическое исследование использования компьютеров в организациях комплексного обслуживания. Оно выявило продолжающиеся социальные и политические процессы, в которых технология ограничена окружающей средой, и подчеркивало взаимодействие между правительственными организациями и их внутренней и внешней средой. Исследователи утверждали, что вычислительная техника усиливает мощь и влияние субъектов и групп, которые уже обладают наибольшими ресурсами и властью в организации.

2) Рекомендация включения вычислительную технику в программы магистратуры государственного управления. Комитет NASPAA рекомендовал, чтобы вычислительная техника была основным навыком, преподаваемым в программах MPA. Их рекомендации включали в себя курсы по компьютеру для всех студентов, курсы "компьютерные приложения для управления", специализацию по управлению информацией и интеграцию компьютерных навыков в основные курсы государственного управления.

3) Создание правительственного портала firstgov, и последующее создание законов, поощряющих инвестиции в ИТ. Было установлено, что технологии влияют на правительство и нужно уделить им больше внимания. В 1990-х годах информационные технологии стали частью государственной реформы, благодаря созданию универсального правительственного портала, который в настоящее время называется "firstgov". Законодательные акты, такие как Закон Клингера-Коэна и Закон об электронном правительстве 2001 года, обеспечили инфраструктуру для распространения приложений электронного правительства.

Помимо этого, затронул тему пересмотра отношения к e-government и переопределение его основных целей, которые произошли после террористических актов 11 сентября 2001 года: теперь электронное правительство перестало рассматриваться только как инструмент улучшения доступа к государственным услугам, а стало рассматриваться как инструмент защиты от терроризма. Как следствие такого подхода, произошли реформы в обеспечении безопасности информационных систем.

Дальше описал меры, которые могут помочь в совершенствовании систем электронного правительства, такие как:

1) Более глубокое изучение моделей электронного правительства в контексте сложных политических условий

2) Решение проблемы недостаточной конкретизации в соответствующей литературе путем производства более обоснованных эмпирических исследований.

3) Прочное связывание предмета электронного правительства с основными исследованиями государственного управления.

Следующим этапом был обзор литературы по data mining. Первым делом были даны определения data mining:

1. Процесс извлечения значимых знаний из больших объемов цифровой информации.
2. Процесс интеллектуального анализа данных, включающий отбор, предварительную обработку, преобразование, интеллектуальный анализ и интерпретацию данных.

Потребность в этой технологии обуславливается тем, что с ростом объемов данных, человеческих возможностей для анализа стало недостаточно. Используя интеллектуальные алгоритмы и методы, этот процесс позволяет обнаруживать неизвестные закономерности и улучшать эффективность услуг, например, в медицине и бизнесе. Далее описал задачи data mining: они могут быть прогнозирующими или описательными, включая классификацию, регрессию, кластеризацию, моделирование зависимостей, обнаружение отклонений и обобщение. Также затронул проблемы этой технологии, такие как безопасность, методологические проблемы, проблемы пользовательского интерфейса и производительности. В конце параграфа описал сферы применения технологии data mining: она используется в здравоохранении, образовании, CRM, анализе рыночной корзины, спорте и т.д.

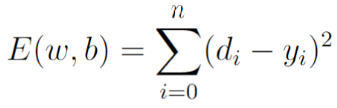
Далее перешел к параграфу 2 Главы 2: «Основы вычислительных нейронных сетей». Дал определения нейронных сетей, а также области их применения. Привел описание структуру и принцип работы биологического нейрона. На основании этого, привел описание искусственного нейрона и принцип устройства простейшей нейронной сети, дал определение слоям, озвучил их виды и назначение. Далее провел параллели между искусственными нейронными сетями и биологическими; дал определение весу связи нейронной сети, обозначил его смысл. Затем описал термин активационной функции, привел аргументы, почему эта функция не должна быть линейной. Привел примеры наиболее часто используемых функций активации:

* Сигмоид: f(x) = 1 / (1 + e^-x).
* ReLU: f(x) = max(0, x).
* Softmax: f(x) = e ^ x\_i / sum(e ^ x\_j).

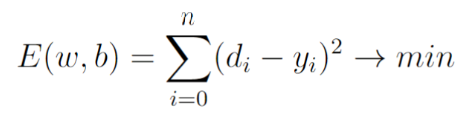
Затем перешел к описанию обучения нейронной сети. Дал определение понятию «обучение нейронной сети». Привел классификацию типов обучений:

* Обучение без учителя
* Обучения с учителем
* Трансфертное обучение

Дал определение функции ошибки



и описал, в чем, с математической точки зрения, заключается проблема обучения нейронной сети: задача сводится к поиску минимума функции E ошибки, то есть, задача принимает вид:



Привел пример метода решения такой задачи, а именно, метод градиентного спуска. Перед началом описания сути этого метода, дал определение градиенту, описал его геометрический смысл. Затем перешел непосредственно к самому методу градиентного спуска, описывая его на примере одномерного случая, то есть, когда функция Е зависит только от одного параметра. Описал геометрический смысл градиентного спуска. После описания алгоритма обратного распространения ошибки(backpropagation), перешел к описанию классификации типов нейронных сетей:

* Сети прямого распространения
* Сети обратного распространения
* Рекуррентные сети
* Сверточные сети

Дал определение каждой упомянутой нейронной сети, предоставил их схематический вид, описал различия и сферы применения:

**Сети прямого распределения:**

Сети прямого распространения, также известные как нейронные сети прямого действия (FNN), представляют собой фундаментальный тип нейронных сетей, в которых данные передаются в одном направлении, от входных уровней к выходным без циклов.

Эти сети состоят из слоев взаимосвязанных нейронов, где каждый нейрон в слое связан со всеми нейронами в последующем слое.

Обычно в качестве функции активации выступает ReLU

**Области применения**: Нейронные сети прямого действия используются для различных задач, включая классификацию изображений, регрессию, обработку естественного языка и распознавание речи.

**Сети обратного распространения:**

Обратное распространение — это алгоритм обучения, используемый в нейронных сетях, обычно в сочетании с сетями прямой связи.

Обучение включает в себя итеративную корректировку весов и смещений сети путем вычисления градиента функции потерь относительно этих параметров и использования этой информации для их обновления.

Обратное распространение является ключевым компонентом обучения нейронных сетей для повышения их производительности.

**Области применения**: Обратное распространение используется в нейронных сетях прямого действия для таких задач, как распознавание изображений, речи и обработка естественного языка.

**Рекуррентные сети:**

Рекуррентные нейронные сети (RNNS) — это тип нейронной сети, который позволяет информации передаваться циклически, что делает их пригодными для последовательных данных.

RNN имеют внутреннюю память, которая позволяет им фиксировать временные зависимости в данных.

Они часто используются для задач, где важен порядок входных данных, таких как прогнозирование временных рядов, языковое моделирование и распознавание речи.

**Области применения**: RNN широко используются в обработке естественного языка для таких задач, как машинный перевод, генерация текста и анализ настроений, а также в распознавании речи и прогнозировании временных рядов.

**Сверточные сети:**

Сверточные нейронные сети (CNN) — это специализированные нейронные сети, предназначенные для обработки данных в виде сетки, таких как изображения и видео.

Они состоят из слоев, выполняющих свертки, которые по сути являются скользящими окнами, сканирующими входные данные для извлечения локальных шаблонов.

CNN особенно эффективны при автоматическом изучении иерархических признаков на изображениях, что делает их подходящими для задач классификации изображений и обнаружения объектов.

**Области применения**: CNN широко используются в компьютерном зрении для таких задач, как классификация изображений, обнаружение объектов, распознавание лиц и генерация изображений.

Итого: сети прямой связи универсальны и могут использоваться для различных применений. Обратное распространение — это метод обучения, используемый в сетях прямой связи. Рекуррентные сети превосходно справляются с обработкой последовательных данных, в то время как сверточные сети адаптированы для задач, связанных с изображениями. Выбор сетевой архитектуры зависит от решаемой задачи и характера входных данных.

После описания видов ИНС перешел к Главе 3: «Обзор программного обеспечения data mining». Интеллектуальный анализ данных (data mining) выявляет закономерности и взаимосвязи. Для удовлетворения растущего спроса появились различные программные инструменты, включая RapidMiner, SAS, IBM SPSS Modeler, Oracle Data Mining, Knime, Weka, Alteryx, Tableau и Orange. Они предлагают уникальные возможности - от анализа данных до визуализации. Этот обзор помогает выбрать правильный инструмент для конкретных задач.

**RapidMiner** - это программное обеспечение для анализа данных с открытым исходным кодом, известное своим удобным интерфейсом и обширными возможностями предварительной обработки данных и моделирования. Он хорошо интегрируется с другими инструментами обработки данных, такими как Python, R и SQL, и предлагает множество вариантов развертывания.

**Преимущества**:

* Удобный для пользователя: Простой в использовании графический интерфейс для непрограммистов.
* Открытый исходный код: настраивается в соответствии с конкретными потребностями.
* Интеграция: легко работает с Python, R и SQL.
* Богатая функциональность: предлагает широкий спектр инструментов для обработки данных и моделирования.
* Параметры развертывания: поддерживает различные методы развертывания.

**Недостатки**:

* Ограниченная бесплатная версия: Бесплатная версия имеет ограничения.
* Скудная документация: Документация может иметь темную глубину.
* Ограниченная масштабируемость: проблемы с большими наборами данных или сложными моделями.
* Базовая визуализация: предлагает основные параметры визуализации.
* Ограниченный контроль: может не подходить пользователям, нуждающимся в детальном контроле рабочего процесса.

**SAS** (система статистического анализа) широко используется в финансах, здравоохранении и государственном управлении для интеллектуального анализа данных. Он обрабатывает структурированные и неструктурированные данные из различных источников, предлагая инструменты очистки, преобразования данных и статистического/ машинного обучения. SAS включает в себя гибкий язык сценариев (скриптов) SAS Language для пользовательской автоматизации и управления рабочим процессом.

**Преимущества**:

* Комплексный: предлагает обширную обработку данных, анализ и моделирование для структурированных и неструктурированных данных.
* Гибкость: используется универсальный язык сценариев для настройки и автоматизации.
* Интеграция: интегрируется с Python и R, обеспечивая использование разнообразных инструментов и языков.
* Активное сообщество: извлекает выгоду из большого, поддерживающего сообщества пользователей.
* Высокая производительность: Предназначен для больших, сложных наборов данных и высокопроизводительных вычислений.

**Недостатки**:

* Кривая обучения: сложная задача для непрограммистов или тех, кто новичок в статистике.
* Ограниченная настройка: Сложная настройка, выходящая за рамки встроенных функций, может быть затруднена.
* Базовая визуализация: Базовые параметры визуализации, лишенные расширенных функций.
* Ограничения масштабируемости: могут возникнуть проблемы с очень большими наборами данных или сложными моделями.
* Проприетарный: Ограниченная модификация или расширение из-за его проприетарного характера.

**IBM SPSS Modeler**, используемый в финансах, здравоохранении и маркетинге, имеет явное преимущество перед RapidMiner: удобный интерфейс. Пользователи могут легко создавать и изменять рабочие процессы обработки данных без кодирования. Он предлагает рабочее пространство с возможностью перетаскивания. IBM SPSS Modeler включает в себя инструменты обработки данных, поддерживает различные статистические методы и методы машинного обучения и интегрируется с R и Python. Он развертывает модели в виде веб-служб или пакетных процессов. Однако это дорого и имеет ограниченную масштабируемость.

**Преимущества:**

* Удобный интерфейс: легко создавайте и изменяйте рабочие процессы без кодирования.
* Всесторонняя функциональность: поддерживает структурированный и неструктурированный анализ данных.
* Интеграция: легко интегрируется с R и Python.
* Варианты развертывания: обеспечивает интеграцию модели в бизнес-процессы.
* Активное сообщество пользователей: Преимущества большого и активного сообщества пользователей.

**Недостатки**:

* Стоимость: Высокие цены ограничивают доступность для небольших организаций.
* Ограниченная масштабируемость: проблемы с очень большими наборами данных или сложными моделями.
* Ограниченная настройка: трудно настроить рабочие процессы, выходящие за рамки предоставленной функциональности.
* Базовая визуализация: предоставляет основные параметры визуализации.
* Проприетарный: Проприетарный инструмент, аналогичный SAS.

**Oracle Data Mining**, часть Oracle Advanced Analytics for Oracle Database, легко интегрируется с базой данных для анализа больших наборов данных. Он предлагает функции подготовки данных, поддерживает как контролируемое, так и неконтролируемое обучение и позволяет развертывать модели в виде функций SQL. Однако его не очень удобный графический интерфейс требует опыта. Гибкость инструмента ограничена для решения конкретных задач.

**Преимущества**:

* Интеграция с базой данных Oracle: Разработан для бесперебойной работы с базой данных Oracle.
* Эффективная обработка: Оптимизирован для крупномасштабного интеллектуального анализа данных на многоядерных процессорах и кластерах.
* Развертывание модели: позволяет интегрировать модель в существующие приложения баз данных.
* Поддержка и ресурсы: Всесторонняя поддержка, документация, учебные пособия и форумы.

**Недостатки**:

* Сложность: Сложная задача для пользователей, не знакомых с Oracle Database или SQL.
* Стоимость: требуется приобрести опцию Oracle Advanced Analytics, которая может быть дорогостоящей.
* Ограниченная интеграция: Интеграция с инструментами или источниками данных, отличными от Oracle, может быть затруднена.
* Проприетарное программное обеспечение: Oracle Data Mining является проприетарным.
* Ограниченная визуализация: предоставляет ограниченные возможности визуализации.

Научный руководитель: доц. В.Г.Салимов

Источники:

* Brown, M. M., & Brudney, J. L. (2003). Learning organizations in the public sector? A study of police agencies employing information and technology to advance knowledge. Public Administration Review, 63(1).
* Doty, P., & Erdelez, S. (2002). Information micro-practices in Texas rural courts: Methods and issues for Egovernment. Government Information Quarterly, 19.
* Aldrich, J., Bertot, J. C., & McClure, C. R. (2002). E-government: initiatives, developments, and issues. Government Information Quarterly, 19.
* Bozeman, B., & Bretschneider, S. (1986). Public management information systems: Theory and prescription. Public Administration Review, 46.
* Halchin, L. E. (2004). Electronic government: Government capability and terrorist resource. Government Information Quarterly, 21.
* Yildiz, M. (2003). A general evaluation of the theory and practice of e-government (In Turkish). In M. Acar & H. Ozgur (Eds.), Cagdas Kamu Yonetimi-1. Istanbul: Nobel Publications.
* Jaeger, P. T. (2003). The endless wire: E-Government as a global phenomenon.
* Piatetsky-Shapiro, Gregory. "Knowledge discovery in databases: 10 years after." ACM SIGKDD Explorations Newsletter 1.2 (2000).
* Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3 (1996).
* Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "The KDD process for extracting useful knowledge from volumes of data." Communications of the ACM 39.11 (1996).
* Rygielski, Chris, Jyun-Cheng Wang, and David C. Yen. "Data mining techniques for customer relationship management." Technology in society 24.4 (2002).