

Министерство образования Республики Беларусь
Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ
И РАДИОЭЛЕКТРОНИКИ

УДК _____

Сироткин
Александр Владимирович

Натуральный текстовый анализатор

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 40.80.02 «Системный анализ, управление и обработка
информации»

Научный руководитель
Герман О.В.
кандидат технических наук, доцент

Минск 2015

ХАРАКТЕРИСТИКА

Целью данной работы является разработка системы автоматической постановки диагноза на основании анамнеза в свободной форме. Разрабатываемое приложение должно предоставлять возможность пользователю вести свои симптомы на естественном языке, после чего система должна вывести наиболее подходящие заболевания. Также система должна предоставлять экспертам возможность её настройки. Таким образом, предстоит решить ряд задач:

- изучение существующих систем автоматической постановки диагноза;
- изучение и анализ существующих алгоритмов по классификации текстовой информации;
- разработка метода для автоматической постановки диагноза;
- сбор данных для работы системы;
- разработка программного обеспечения автоматической постановки диагноза.

Разработанное программное обеспечение предоставляет пользователю возможность вводить свой анамнез или загружать текстовый файл, предоставляет экспертам методы по настройке системы для более эффективной постановки диагноза. Эксперты имеют возможность:

- просматривать доступные болезни;
- редактировать доступные болезни (добавлять, изменять, удалять);
- изменять текстовую информацию, соответствующую данному заболеванию (описание течения болезни, симптоматика, различные осложнения);
- просматривать симптомы для каждого из заболеваний;
- добавлять симптомы, изменять их важность, удалять симптомы.

СОДЕРЖАНИЕ

1	Постановка задачи	Ошибка! Закладка не определена.
2	Обзор существующих аналогов	Ошибка! Закладка не определена.
2.1	Поисковые системы.....	Ошибка! Закладка не определена.
2.2	Системы поиска симптомов.....	Ошибка! Закладка не определена.
2.3	Системы на базе анкетирования пользователя	Ошибка! Закладка не определена.
2.4	Семантические сети	Ошибка! Закладка не определена.
3	Проектирование модели	Ошибка! Закладка не определена.
3.1	Предполагаемая модель	Ошибка! Закладка не определена.
3.2	Классификация в рамках группы	Ошибка! Закладка не определена.
3.3	Классификация в рамках класса	Ошибка! Закладка не определена.
3.4	Итоговая модель	Ошибка! Закладка не определена.
4	Программная реализация.....	Ошибка! Закладка не определена.
4.1	Используемое программное обеспечение	Ошибка! Закладка не определена.
4.2	Описание структуры разрабатываемой системы	Ошибка! Закладка не определена.
4.3	Руководство пользователя	Ошибка! Закладка не определена.
	Заключение	8
	Список использованных источников	Ошибка! Закладка не определена.
	Список собственных публикаций	9

ВВЕДЕНИЕ

В последнее время неуклонно возрастает значение информационного обеспечения различных медицинских технологий. Использование современных информационных технологий становится критическим фактором развития большинства отраслей знания и областей практической деятельности, поэтому разработка и внедрение информационных систем является одной из самых актуальных задач[1].

В медицинских учреждениях большинство персональных компьютеров применяется лишь для обработки текстовой документации, хранения и обработки баз данных, ведения статистики и выполнения финансовых расчетов. Отдельная, специализированная часть машин используется совместно с различными диагностическими и лечебными приборами.

Во многих лечебно-диагностических технологиях возможности современных компьютеров практически не используются. Прежде всего это диагностика, назначение лечебных мероприятий, прогнозирование течения заболеваний и их исходов. Основными причинами недостаточно полного использования современных компьютерных технологий в медицине являются слабо развитая техническая база, недостаточный уровень подготовки участников этих технологий в области современного аппаратного и программного обеспечения, плохая оснащенность специализированными пакетами прикладных программ и др. Большое значение имеет психологический аспект применения компьютерных приложений. Это серьезная причина, связанная с особенностями работы врача. Врач является исследователем, его работа носит творческий характер, однако он несет прямую ответственность за результат своей деятельности. Принимая решение о диагнозе или лечении, он опирается на знания и опыт – свои собственные и коллег, являющихся для него авторитетом. Очень важно при этом обоснование решения, особенно если оно подсказывается со стороны.

Современные технические возможности позволяют выйти на качественно новый уровень диагностику заболеваний, а именно на основе соответствующих математических моделей диагностировать то или иное заболевание. Возникает дилемма: либо приходится жертвовать полнотой анализа информации, либо необходимо шире использовать различные методы компьютерной поддержки принятия решений. Медицинские системы позволяют врачу не только проверить собственные диагностические

предположения, но и обратиться к компьютеру за консультацией в трудных диагностических случаях.

Область исследований, посвященная формализации способов представления знаний и построению экспертных систем (ЭС), называют «инженерией знаний». Этот термин введен Е. Фейгенбаумом и в его трактовке означает «привнесение принципов и средств из области искусственного интеллекта в решение трудных прикладных проблем, требующих знаний экспертов». Иными словами, экспертные системы применяются для решения неформализованных проблем, к которым относятся задачи, обладающие одной или несколькими характеристиками из следующего списка:

- задачи не могут быть представлены в числовой форме;
- исходные данные и знания о предметной области неоднозначны, неточны, противоречивы;
- цели нельзя выразить с помощью четко определенной целевой функции;
- не существует однозначного алгоритмического решения задачи.

Все вышеперечисленные свойства являются типичными для медицинских задач, так как в большинстве случаев они представлены большим объемом многомерных, запутанных, а порой и противоречивых клинических данных. ЭС позволяют решать задачи диагностики, дифференциальной диагностики, прогнозирования, выбора стратегии и тактики лечения и др.

Среди экспертных медицинских систем особое место занимают так называемые самообучающиеся интеллектуальные системы (СИС). Они основаны на методах автоматической классификации ситуаций из реальной практики или на методах обучения на примерах.

Искусственные нейронные сети (ИНС; artificial neural networks) представляют собой нелинейную систему, позволяющую классифицировать данные гораздо лучше, чем обычно используемые линейные методы. В приложении к медицинской диагностике ИНС дают возможность значительно повысить специфичность метода, не снижая его чувствительность.

Наиболее важным отличием ИНС от остальных методов прогнозирования является возможность конструирования экспертных систем

самим врачом-специалистом, который может передать нейронной сети свой индивидуальный опыт и опыт своих коллег или обучать сеть на реальных данных, полученных путем наблюдений. Нейронные сети способны принимать решения, основываясь на выявляемых ими скрытых закономерностях в многомерных данных. Положительное отличительное свойство ИНС состоит в том, что они не программируются, т.е. не используют никаких правил вывода для постановки диагноза, а обучаются делать это на примерах. Нейронные сети нашли применение во многих областях техники, где они используются для решения многочисленных прикладных задач: в космонавтике, автомобилестроении, банковском и военном деле, страховании, робототехнике, при передаче данных и др. Другое, не менее важное, свойство нейронной сети состоит в способности к обучению и обобщению полученных знаний. Натренированная на ограниченном множестве обучающих выборок, она обобщает накопленную информацию и вырабатывает ожидаемую реакцию применительно к данным, не обрабатывавшимся в процессе обучения.

Примером другой перспективной технологии обработки и обобщения больших объемов информации для решения задач классификации и прогнозирования является так называемая технология анализа и добычи данных Data Mining. Методы и инструментальные средства анализа и добычи данных представляют собой дальнейшее развитие таких известных статистических инструментов разведочного анализа, как метод главных и метод независимых компонент, факторный анализ, множественная регрессия, редуцирование пространства признаков с использованием метода многомерного шкалирования, кластерного анализа и распознавания образов и других. Программно реализованные и снабженные удобным пользовательским интерфейсом, а также поддерживаемые гибкими алгоритмами визуализации многомерных данных, средства Data Mining позволяют проводить соответствующие исследования даже начинающему пользователю. В арсенал методов кластерного анализа и распознавания образов систем Data Mining обычно входят метод опорных векторов (Support Vector Machine, или SVM), метод деревьев решений (decision trees), метод «ближайшего соседа» в пространстве признаков, байесовская классификация и др. [2].

Примеры использования экспертных систем в медицине нельзя назвать единичными, они применяются во многих областях здравоохранения. Так что же дает использование таких систем в медицине? Для пациента дает возможность узнать список возможных заболеваний, и тем самым сузить круг специалистов, к которым следует обратиться. Для специалистов – возможность сразу просмотреть анамнез, и по возможности без дополнительных вопросов дать заключение.

Библиотека БГУИР

ЗАКЛЮЧЕНИЕ

В результате работы над магистерской диссертацией были рассмотрены системы по автоматической постановке диагноза. Но на данный момент не одна из систем не предоставляет достаточно качественного решения этой задачи.

Для решения поставленной задачи предлагается двухступенчатый подход: классификация в пределах группы заболеваний (эндокринные, сердечные и т.д.), классификация в пределах релевантных представителей подмножества группы (сахарный диабет, гипертиреоз). Описаны методы, использованные на каждом из этапов классификации. Разработанный двухступенчатый классификатор объединяет методику частотного анализа с методом нечеткого логического вывода, причем в качестве параметров первого метода используется не обычная частота, а взвешенная частота с учетом семантической оценки. Таким образом, новизна данной модели заключается в следующем: двухступенчатая классификация, модифицированный метод частотного анализа, синтез методов частотного анализа и нечеткого логического вывода.

Программно данная модель реализована как пилот-проект, но на тестовых данных хорошо себя показавшая. Дальнейшее её использование, разработка и проверка предполагается в период обучения в аспирантуре.

Данное приложение будет полезно как для пациентов, так и для специалистов. Для пациента дает возможность узнать список возможных заболеваний, и тем самым сузить круг специалистов, к которым следует обратиться. Для специалистов – возможность сразу просмотреть анамнез и иметь дополнительное мнение по данному анамнезу.

Данная методика также может служить основанием для применения в различных областях, где требуется text mining – online консультации, обучение без учителя, системы интерактивной диагностики в интернет и т.п. Подход использует комплексное применение моделей распознавания, его цель состоит как в повышении производительности, так и в обеспечении более релевантных ответов.

СПИСОК СОБСТВЕННЫХ ПУБЛИКАЦИЙ

1. Герман О.В., Сироткин А.В.. Модель распознавания на основании анамнеза и её приложения. //Информационные технологии и системы: мат. Международной конференции БГУИР, Минск, 28 октября 2015 г. – Минск, БГУИР, 2015. – С. 304-305.

Библиотека БГУИР