

О ПРИМЕНЕНИИ МЕТОДА АНАЛИЗА СЕМАНТИЧЕСКИХ СЕТЕЙ В ОПРЕДЕЛЕНИИ СТЕПЕНИ ТЯЖЕСТИ ЗАБОЛЕВАНИЯ

С.В. Аксенов, к.т.н., доц. ОИТ ИШИТР,
М.Н. Кушеева, студент гр. 8ИИМ9М.
Томский политехнический университет
E-mail: marykush98@mail.ru

Введение

Семантические сети, как способ визуализации данных, имеют огромное количество сфер применения. Данная же работа посвящена применению семантических сетей в структуризации и анализе текстовых медицинских данных, конкретно для определения степени тяжести заболевания. Данное исследование является достаточно актуальным, потому что применение семантических сетей в таких случаях в перспективе должно упростить задачу анализа медицинских документов, ускоряя развитие сектора здравоохранения.

Итак, объектом исследования являются семантические сети, а предметом исследования – метод их использования в определении степени тяжести заболевания.

Основной целью исследования можно считать изучение способа применения семантических сетей в обработке медицинских данных для повышения производительности в сфере здравоохранения.

Описание алгоритма

Проблемам создания систем обработки естественного языка, представленных в текстовом формате, посвящены работы огромного количества исследователей (Баранов А.А., Намазова-Баранова Л.С., Смирнов И.В., Девяткин Д.А., Шелманов А.О.) [1,2,3]. Анализ их работ показал, что семантические сети достаточно эффективно могут быть применены в области обработки текстовых медицинских данных и имеют огромный потенциал в использовании.

Суть предлагаемого способа использования семантических сетей заключается в сравнении двух семантических сетей, одна из которых построена на основе данных осмотра пациента врачом и отражает текущее состояние пациента, а вторая описывает одно из степеней тяжести конкретного заболевания (легкое, среднее, тяжелое, крайне тяжелое). В зависимости от полученного коэффициента сравнения сетей можно будет сделать вывод о состоянии тяжести пациента. На рисунке 1 изображен пример семантической сети пациента. Стоит отметить, что семантическая сеть болезни будет иметь схожую структуру для получения более конкретных результатов.



Рис.1. Прототип семантической сети, построенной на основе данных пациента

В настоящее время существует метод оценки качества семантических сетей [4], суть которого заключается в выведении коэффициента совпадения (1) одной семантической сети, называемой эталонной, с другой сетью, называемой кандидатом.

$$Score = F_{mean}(1-p), \quad (1)$$

где F_{mean} – это среднее гармоническое значение, объединяющее в себе отношение двух переменных: количеств общих терминов и отношений в каждой из сравниваемых семантических сетей; а p – это штрафной коэффициент за несовпадающие термины и отношения в обеих сетях.

Значение $Score$ по своей сути является оценкой силы корреляционной связи двух семантических сетей. Используя этот метод, но принимая в качестве сетей-кандидатов сети, построенные по уже

известным характеристикам определенного заболевания, а в качестве эталонной сети сеть, построенную по данным осмотра пациента с уже известным диагнозом, можно вычислить некоторый коэффициент совпадения. На рисунке 2 представлена визуализация анализируемых семантических сетей, на которой видно, что сеть-пациент будет поочередно сравниваться с сетями степеней тяжести болезни в поисках наивысшего совпадения.

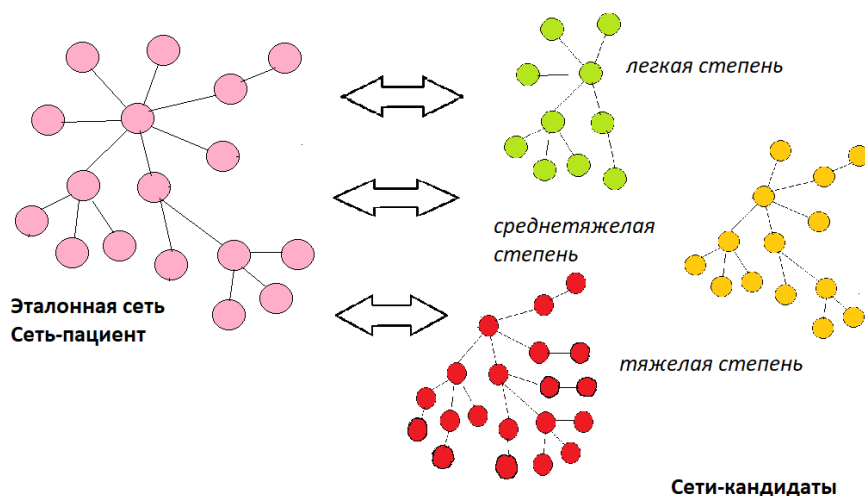


Рис. 2. Визуальное представление эталонной семантической сети и сетей-кандидатов.

Сравнивая эталонную сеть с каждой из представленных сетей-кандидатов, получаем определенное значение $Score \in [0,1]$. Чем ближе полученный коэффициент к 1, тем больше сети совпадают друг с другом. Таким образом, в процентном отношении рассматриваемый пациент может иметь ту или иную степень тяжести своего заболевания.

Заключение

Исследование и разработка метода определенно внесут свой вклад в автоматизацию и компьютеризацию работы в сфере здравоохранения. Определение степени тяжести заболевания необходимо в первую очередь лечащему врачу для отслеживания течения болезни, тактики лечения и прогноза состояния пациента, также степень тяжести болезни определяет сумму выплачиваемой страховщиками компенсации и позволяет подобным организациям эффективно вести свою деятельность. Однако, в настоящий момент данное исследование нуждается в качественно проведенных экспериментах, для проведения которых необходимо создать программный прототип, сделать выборку историй болезни, и построить соответствующие им семантические сети. Вне зависимости от качества тестирования данный метод все равно будет иметь некоторую погрешность, которую также необходимо вычислить в ходе экспериментов.

Список использованных источников

1. Баранов А.А., Намазова-Баранова Л.С., Смирнов И.В. и др. Технологии комплексного интеллектуального анализа клинических данных // Вестник Российской академии медицинских наук. – 2016. – с. 160-171
2. Лапаев М.В. Система обработки текстовых медицинских данных // Известия СПбГЭТУ «ЛЭТИ». – 2016. – № 9. – с. 11-15
3. Шелманов А.О. Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа / А.О. Шелманов – М.: Изд-во ИСА РАН, 2015. – 210 с
4. Аюшеева Н.Н., Гомбожапова Т.Н., Кушеева М.Н. Метод оценки качества семантических сетей // Современные наукоемкие технологии. – 2019. – № 11 (часть 1) – с. 77-81.
5. Понятие и назначение семантических сетей [Электронный ресурс]. – URL: <https://libtime.ru/expertsystems/ponyatie-i-naznachenie-semanticheskikh-setey.html> (дата обращения: 25.02.2021)
6. Аюшеева Н.Н. Разработка методов построения семантической сети текста. Монография. / Н.Н. Аюшеева, Т.Н. Гомбожапова. – Улан-Удэ: Изд-во ВСГУТУ, 2016. – 124 с.