

О. А. Вдовиченко, А. В. Аверченков

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ПОСТРОЕНИЯ АССОЦИАТИВНЫХ ПРАВИЛ ДЛЯ ВЫЯВЛЕНИЯ ГРУПП РИСКА В ДАННЫХ ОБСЛЕДОВАНИЯ ПАЦИЕНТОВ

В статье рассматривается применение алгоритма построения ассоциативных правил Apriori для целей анализа результатов ультразвукового обследования щитовидной железы. Алгоритм применяется к решению конкретной задачи организационного обеспечения обследования щитовидной железы. Разработан программный инструментарий, позволяющий медикам применять указанный алгоритм для проведения необходимых исследований в процессе решения задач диагностики.

Ключевые слова: здравоохранение, информационно-аналитические системы, data mining, apriori.

O.A. Vdovichenko, A.V. Averchenkov

USING ASSOCIATIVE RULE CONSTRUCTION METHODS TO IDENTIFY RISK GROUPS IN PATIENTS' DIAGNOSTIC FINDINGS

The article considers the application of the Apriori association rule construction algorithm to analyze the results of the thyroid gland ultrasound examination. The algorithm is applied to solve a specific problem of organizational support of the thyroid gland examination. A software toolkit has been developed that allows physicians to apply the specified algorithm to carry out the necessary research in the process of solving diagnostic problems.

Keywords: healthcare, information and analytical systems, data mining, apriori.

Введение

В современных условиях перед многими учреждениями здравоохранения РФ стоит задача повышения эффективности деятельности по выявлению заболеваний и сохранению здоровья населения. Зачастую деятельность учреждений здравоохранения основывается на проведении обследований только для граждан, изъявивших желание пройти обследование, что создает проблему недостаточно высокой эффективности обследований.

Автоматизированный анализ данных о пациентах является одним из ключевых направлений исследований в сфере здравоохранения. Область применения информационно-аналитических систем в этой отрасли охватывает многие направления – включая непосредственный анализ технических данных для выявления симптомов определенных заболеваний, помощь в принятии решений лечащим врачом, анализ косвенно связанных с заболеванием персональных и иных данных.

Особую роль на территории ряда регионов, в том числе Брянской области, занимает проведение эндокринологических исследований. Важность данного вопроса обусловлена негативным влиянием радиоактивных веществ, попавших на территорию региона в результате аварии на Чернобыльской АЭС.

Процесс ультразвукового обследования щитовидной железы у населения региона

позволяет накапливать большие массивы информации. Проблема повышения эффективности обследований и общей деятельности по профилактике заболеваний щитовидной железы населения может быть решена при использовании полученной информации в качестве входных данных для применения алгоритмов анализа. Современные методы анализа позволяют выявить группы признаков, целевые группы, для которых проведение регулярных обследований, направленных на выявление заболеваний, является необходимой составляющей обеспечения длительного сохранения здоровья.

В данной статье рассматривается применение алгоритма поиска ассоциативных правил *Argioi*, относящегося именно к группе методов *Data Mining*. Этот алгоритм позволяет выявлять на основе характеристик пациентов и их диагнозов закономерности формата «если условие, то результат» и интерпретабельные базы правил. Впервые данный алгоритм был описан в работе Rakesh Agrawal, посвященной поиску взаимосвязанных данных в крупноразмерных БД [1]. Рассмотрим применение данного алгоритма к задачам управленческого обеспечения профилактики заболеваний щитовидной железы.

Описание исходных данных для алгоритма

Существует большое количество различных реализаций алгоритма *Argioi* в его исходном виде и с модификациями, предназначенными для решения конкретных прикладных задач. Описание некоторых из них можно найти в работах [2 – 5]. Большинство из описанных применений ориентированы на анализ потребительских корзин. Применение модификаций данного алгоритма непосредственно для задач сферы здравоохранения описано в работе [6].

Исходными данными при обследовании щитовидной железы пациента являются:

- персональные характеристики пациентов (возраст, пол, место проживания);
- данные о выявленных заболеваниях (история обследований);
- данные о характеристиках щитовидной железы пациентов (размеры долей, экзогенность, структура, наличие гипо- или гипертериоза);
- данные об образованиях на щитовидной железе пациентов (тип образования, размер образования, экзогенность, характеристики контура, характеристики включений).

На основе этих данных лицом, принимающим решения (врачом) устанавливается диагноз, а также формируются рекомендации для пациента по повторным обследованиям, дальнейшим мерам предотвращения развития заболевания при его выявлении.

Таким образом, в контексте работы использование данного метода позволяет произвести анализ персональных данных пациентов (возраста, пола, места проживания, места рождения) и их медицинских данных для выявления признаков принадлежности к группам риска по определенным заболеваниям; например, определить степень влияния возраста пациента на факт наличия признаков аутоиммунного тиреоидита. Аналитические функции в данной области применения позволяют снизить трудовые затраты на осуществление деятельности кабинета, снизить время, необходимое для обследования отдельного пациента, выявлять развитие хронического заболевания или болезни, которая еще не была должным образом диагностирована, определить принадлежность пациента к группе риска по определенному заболеванию, что позволит врачу обратить особое внимание на выявление симптомов, характерных для такого заболевания. Отсутствие аналитического функционала негативно влияет на процесс осуществления основной деятельности организации.

Для оценки применимости метода поиска ассоциативных правил *Argioi* для задач выявления групп риска среди пациентов была сформирована выборка, состоящая из возраста обследуемого пациента, населенного пункта его проживания, пола, диагноза, величины щитовидной железы и её экзогенности. Пример данных, включенных в итоговую отфильтрованную выборку, представлен на рис. 1.

	Age	Sex	Diag	Size	Echo
	0	9	male	Healthy	Normal Middle
	1	25	female	DiffusedStruma	Hyperplasia Middle
	2	29	male	DiffusedStruma	Hyperplasia Low
	3	57	male	DiffusedStruma	Hyperplasia Low
	4	32	male	Healthy	Normal Middle

20988	17	female	Healthy	Normal	Middle
20989	41	female	KnotStruma	Normal	Middle
20990	18	female	Healthy	Normal	Middle
20991	18	male	Healthy	Normal	Middle
20992	18	male	Healthy	Normal	Middle

Рис. 1. Состав полученных данных

Количество данных в выборке составило 20992 записей об обследованиях. Выборка состояла из пациентов г. Новозыбкова, проживающих в Чернобыльской зоне, данные были получены на основе обследований.

Результаты применения метода

Для применения метода экспериментальным путем были установлены параметры поддержки, доверия и лифта. Для поддержки было установлено исходное значение 0.005, что соответствует наличию правила в 0,5 % записей от общего количества значение было выбрано с учетом недостаточного количества записей для нахождения достоверных правил. Значение доверия – 0.3, что соответствует уровню в 30 % подтверждения правила в записях выборки. Значение лифта – 3, что позволяет предполагать повышение вероятности встречи правила при наличии n-1 элементов из правила в данных. При обработке алгоритмом представленных исходных данных был получен набор из 619 ассоциативных правил. Наиболее достоверные правила представлены в таблице 1.

По результатам поиска ассоциативных правил удалось выявить взаимосвязь возраста с определенными диагнозами.

Наибольшим значением поддержки и доверия характеризуются правила, связанные с диагнозом «Аутоиммунный тиреоидит».

Наибольшим значением поддержки и доверия характеризуются правила для диагноза «Узловой зоб».

При пониженной экзогенности щитовидной железы наиболее часто проявляется диагноз «аутоиммунный тиреоидит» – следует отметить, что верно и обратное, кроме того, для женщин этот диагноз гораздо более характерен, чем для мужчин – поддержка подобного правила для мужского пола составила лишь 0,0018 % от общей выборки.

для возраста от 7 до 18 лет наиболее характерным заболеванием щитовидной железы стал диффузный зоб.

В возрасте от 51 до 59 лет наиболее характерно наличие у пациента узлового зоба.

Для данных в возрастах от 20 от 47 правила отсутствуют – это вызвано большой дисперсией диагнозов. Для установления более точной связи необходимо пополнение

выборки.

Таблица 1. Пример выявленных в результате анализа правил

№	Базовый набор	Добавленное значение	Поддержка	Доверие	Лифт
1	эхогенность: понижена	диагноз: АИТ	0,161816	0,89749	4,239651
2	эхогенность: понижена, пол: жен	диагноз: АИТ	0,149335	0,906069	4,280179
3	эхогенность: понижена, величина: норма	диагноз: АИТ	0,113943	0,897561	4,239986
4	эхогенность: понижена, пол: жен, величина: норма	диагноз: АИТ	0,105654	0,909016	4,2941
5	диагноз: Диффузный зоб	величина: гиперплазия	0,06183	0,912799	6,260172
6	пол: жен, диагноз: Диффузный зоб	величина: гиперплазия	0,032344	0,908969	6,233907
7	пол: жен, диагноз: ДУЗ	величина: гиперплазия	0,008717	0,88835	6,092493

Выявленные правила коррелируют с наблюдениями врачей-эндокринологов [7], что подтверждает их пригодность для дальнейшего использования при обследовании пациентов

Рассмотрим иные результаты, полученные методом поиска ассоциативных правил:

При аутоиммунном тиреоидите наиболее часто наблюдается гиперплазия щитовидной железы – т.е. увеличение размеров относительно возрастного норматива. Подобная картина наблюдается и при диффузном зобе: наиболее часто данный диагноз проявляется именно при гиперплазии щитовидной железы [7].

Диагноз «диффузный зоб» встречается с примерно одинаковой частотой как у женщин, так и у мужчин – так как данный диагноз характерен для возрастов 7-18 лет и увеличенной щитовидной железы, выборка состояла из пациентов г. Новозыбкова, проживающих в Чернобыльской зоне, при йододефиците наиболее часто изменяется размер щитовидной железы, возможно предположить, что существует взаимосвязь между йододефицитом, характерным для зон, пораженных радиацией, и большой частотой данного диагноза (поддержка правила «диффузный зоб»-«гиперплазия» составляет около 6 % от выборки с доверием около 91 %).

Наименьшая поддержка правил для диагноза «Узловой зоб» связана с неопределенностью такого диагноза в медицинской практике: фактически он определяется для всех пациентов, щитовидная железа которых содержит образования, не подходящие под критерии других диагнозов.

Заключение

В рамках задачи анализа данных обследований щитовидной железы пациентов использование алгоритма Argio1 позволило выявить ряд взаимосвязанных характеристик щитовидной железы, которые могут использоваться для отнесения определенных пациентов к группам риска. Текущие результаты анализа показали низкий уровень поддержки, однако при дальнейшем расширении выборки в ходе работы кабинета возможно выявление большего количества ассоциативных правил и установление более достоверных взаимосвязей между характеристиками щитовидной железы и её заболеваниями.

Алгоритм с успехом был применен для решения конкретной задачи получения данных

для профилактики заболеваний учреждениями здравоохранения.

Предложенный алгоритм и программный инструментарий могут с успехом применяться при решении различных задач диагностики в здравоохранении, а также в других прикладных областях.

Список литературы:

1. Agrawal, R. Fast algorithms for mining association rules in large databases / Agrawal, R., Srikant, R. // Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 12-15 сентября 1994 г. / Morgan Kaufmann Publishers; под общ. ред. J. Bocca – Сантьяго, Чили, 1994. – С. 487-499. — ISBN 1-55860-153-8
2. Федоров, С. В. Клиническое использование системы обработки и анализа информации на основе искусственной нейронной сети типа "многослойный перцептрон" / С. В. Федоров, М. Ш. Кашаев, Т. Р. Кашаев // Пермский медицинский журнал. – 2013. – Т. 30. – № 4. – С. 97-102.
3. Карасева, Т. С. Решение задач медицинской диагностики методами интеллектуального анализа данных / Т. С. Карасева // Решетневские чтения. – 2015. – Т. 2. – С. 46-47.
4. Ефимов, А. С. Решение задачи кластеризации методом конкурентного обучения при неполных статистических данных * / А. С. Ефимов // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2010. – № 1. – С. 220-225.
5. Леденева, Т. М. Обобщенные ассоциативные правила / Т. М. Леденева, Е. А. Кретов // Вестник Воронежского государственного технического университета. – 2014. – Т. 10. – № 5. – С. 46-49.
6. Биллиг, В. А. Построение ассоциативных правил в задаче медицинской диагностики / В. А. Биллиг, О. В. Иванова, Н. А. Царегородцев // Программные продукты и системы. – 2016. – № 2. – С. 146-157.
7. Делягин, В. М. Диагностика и дифференциальная диагностика заболеваний щитовидной железы / В. М. Делягин // Практическая медицина. – 2008. – № 3(27). – С. 38-42.

References:

1. Agrawal, R. Fast algorithms for mining association rules in large databases / Agrawal, R., Srikant, R. // Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), September 12-15, 1994 / Morgan Kaufmann Publishers; under the general ed. J. Bocca – Santiago, Chile, 1994. – S. 487-499. — ISBN 1-55860-153-8
2. Fedorov, S. V. Klinicheskoye ispol'zovaniye sistemy obrabotki i analiza informatsii na osnove iskusstvennoy neyronnoy seti tipa "mnogosloynnyy perseptron" / S. V. Fedorov, M. SH. Kashayev, T. R. Kashayev // Permskiy meditsinskiy zhurnal. – 2013. – T. 30. – № 4. – S. 97-102.
3. Karaseva, T. S. Resheniye zadach meditsinskoй diagnostiki metodami intellektual'nogo analiza dannykh / T. S. Karaseva // Reshetnevskiye chteniya. – 2015. – T. 2. – S. 46-47.
4. Yefimov, A. S. Resheniye zadachi klasterizatsii metodom konkurentnogo obucheniya pri nepolnykh statisticheskikh dannykh * / A. S. Yefimov // Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. – 2010. – № 1. – S. 220-225.
5. Ledeneva, T. M. Obobshchennyye assotsiativnyye pravila / T. M. Ledeneva, Ye. A. Kretov // Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta. – 2014. – T. 10. – № 5. – S. 46-49.
6. Billig, V. A. Postroyeniye assotsiativnykh pravil v zadache meditsinskoй diagnostiki / V. A. Billig, O. V. Ivanova, N. A. Tsaregorodtsev // Programmnyye produkty i sistemy. – 2016. – № 2. – S. 146-157.
7. Delyagin, V. M. Diagnostika i differentsial'naya diagnostika zabolevaniy shchitovidnoy zhelezy / V. M. Delyagin // Prakticheskaya meditsina. – 2008. – № 3(27). – S. 38-42.

Статья поступила в редколлегию 27.04.2021 г.

Рецензент: к.т.н., доцент Брянского государственного технического университета Леонов Ю.А.

Статья принята к публикации 12.05.2021 г.

Сведения об авторах:

Вдовиченко Олег Антонович

аспирант, ассистент кафедры «Компьютерные технологии и системы» Брянского государственного технического университета
E-mail: olegwdoa@gmail.com

Аверченков Андрей Владимирович

д.т.н., доцент заведующий кафедрой «Компьютерные технологии и системы»
Тел.: +7 (4832) 58-83-62
E-mail: mahar@mail.ru

Information about authors:

Vdovichenko O.A.

postgraduate student, assistant of the Department “Computer Technologies and Systems” of Bryansk State Technical University
E-mail: olegwdoa@gmail.com

Averchenkov A.V.

Doctor of Technical Sciences, Associate Professor
Head of the Department “Computer Technologies and Systems”
Тел.: +7 (4832) 58-83-62
E-mail: mahar@mail.ru