Секція 3. Системи штучного інтелекту

УДК 004.932.2:616-006.6

МЕТОДИ ПОШУКУ АСОЦІАТИВНИХ ПРАВИЛ В БАЗІ ДАНИХ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

Вербовий С.О.[1)], Зубко В.С.[2)]
Тернопільський національний економічний університет
[1)] аспірант; [2)] магістрант

I. Постановка проблеми

Аналіз біомедичних зображень має важливе значення в сучасній медицині. Сьогодні опрацювання зображень є важливим напрямком застосування сучасної медичної техніки. Задачами опрацювання зображень є опис, аналіз та оброблення зображень. Проблеми аналізу зображень включаючи класичну задачу розпізнавання фігур заданої форми, важлива також експертна оцінка, яка зараз є дорогою. Виникають проблеми, які зумовлені новими завдання опису зображення та пошуком закономірностей або наборів закономірностей, що одночасно зустрічаються в багатьох наборах. Оскільки наборів може бути велика кількість необхідно здійснювати цей пошук автоматично. Тому актуальною задачею є розробка алгоритму для пошуку асоціативних правил бази даних цитологічних та гістологічних зображень, що містять кількісні і якісні ознаки мікрооб'єктів [1].

II. Мета роботи

Метою дослідження є пошук асоціативних правил в базі даних цитологічних та гістологічних зображень диспластичних і ракових процесів молочної залози, використовуючи різні алгоритми інтелектуального аналізу даних.

III. Алгоритми пошуку асоціативних правил

Виділяють 7 основних алгоритмів пошуку асоціативних правил, такі як Apriori (використовує генерування і тестування підходу – генерує набори кандидата і тестує, якщо вони є частими), FilteredAssociator (виконує довільну асоціацію на вхідних даних, передану через довільний фільтр), PredectiveApriori (виконує пошук зі збільшенням порогу підтримки для кращих 'N' правил, що стосуються скоригованого значення достовірності на основі підтримки.), Tertius (генерує і знаходить «цікаві» правила відповідно до міри їх підтвердження), FP-Growth (дозволяє виявити часті набори елементів без генерування наборів кандидатів), GeneralizedSequentialPatterns (виявлення послідовних шаблонів в послідовному наборі даних), HotSpot (набір правил, відображених у вигляді дерева, які максимізують/мінімізують цільову функцію і відповідність значення інтересу сегментів даних) [2].

В більшості випадків використовується алгоритм Apriori. Якщо в структурній одиниці даних зустрівся деякий набір елементів X, то на підставі цього можна зробити висновок про те, що інший набір елементів Y також має з'явитися в цій одиниці. Ці правила мають такий вигляд (1):

$$X \Rightarrow Y. \quad (1)$$

Припустимо, що правило X⇒Y має підтримку (support) s, якщо s% транзакцій з D, містять множину X∪Y(2),

$$size\ D - 100\%. \quad (2)$$

Достовірність (confidence) правила показує, яка ймовірність того, що з X випливає Y. Правило X⇒Y справедливе з достовірністю c, якщо c% транзакцій з D, що містять X, також містять Y(3) [3],

$$supp\ (X) \cdot 100\%. \quad (3)$$

Програмним засобом для пошуку і виділення асоціативних правил є WEKA. Файл, над яким пізніше будуть здійснюватись всі дії повинен бути формату *.arff.

V. Узагальнений алгоритм пошуку асоціативних правил

Розроблений алгоритм включає в себе попередню обробку даних (класифікація),пошук усіх асоціативних правил, виділення з них корисних та візуалізація результатів.

На етапі попередньої обробки даних варто провести класифікацію вхідних екземплярів. Це було зроблено одним з найкращих алгоритмів класифікації в середовищі WEKA, названим J48. Варто

зазначити, що правильно прокласифіковані екземпляри складають близько 76%, тобто 216 з 286 загальної кількості. В таблиці 1 представлені порівняння результатів по заданим критеріям і показникам при знаходженні пошуку асоціативних правил в базі даних біомедичних зображень.

Таблиця 1 - Порівняння результатів експерименту

Критерії/ показники	Підтримка/ Достовірність	Підтримка/ Достовірність	ACC (перше/останнє)	Аналізовані гіпотези/ Досліджені гіпотези
	Apriori	FilteredAssociator	PredectiveApriori	Tertius
На 10 правилах	0,5/0,9 57 екземплярів	0,35/0,9 43 екземпляри	0,99481/0,99368	299985/155268
На 100 правилах	0,2/0,9 143 екземпляри	0,15/0,9 100 екземплярів	0,99481/0,99011	352939/199078
Час знаходження 10/100 правилах	0,15 сек/0,45 сек	0,18 сек/0,33 сек	4,03 сек/8,25 сек	17, 85 сек/25, 07 сек

На рисунку 1 представлена залежність двох атрибутів: розмір пухлини та ступінь злоякісної пухлини.

Відповідно до результатів експерименту можна сказати що на великій кількості правил алгоритми PredectiveApriori і Tertius працюють довше, але точність краще обчислюється. Коли ж результати Apriori і FilteredAssociator напряму залежать від кількості оброблених екземплярів даних.

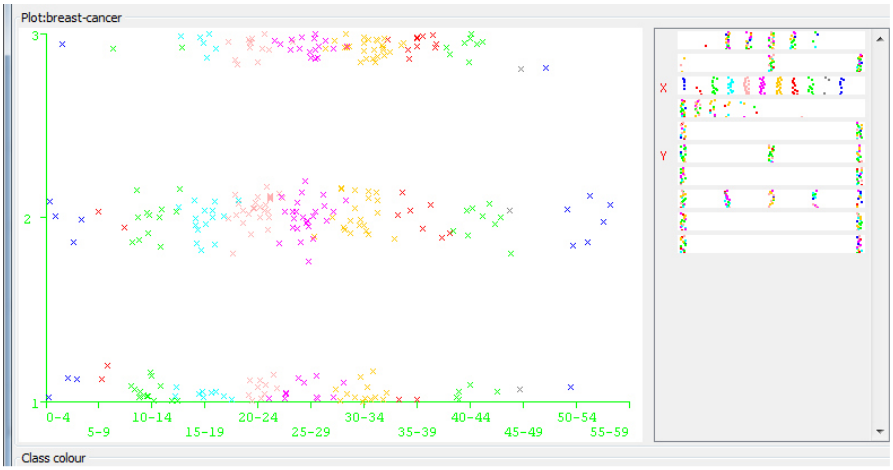

Рисунок 1 – Залежність розміру пухлини(X) від ступеню злоякісної(Y)

На основі отриманих результатів можна отримати наступні приклади повних асоціативних правил передракових станів молочної залози, використовуючи алгоритм Tertuis:

/* 0,911736 0,001808 */ Діагноз_Цитологія = Папілярний рак ==> Клітина_Форма = Кубічна or Угрупування клітин або клітинні комплекси_Розташування = Багаточисельними округлими стурктурами or Угрупування клітин або клітинні комплекси_Форма = Папілярні

/* 0,911736 0,001808 */ Діагноз_Цитологія = Папілярний рак ==> Клітина_Форма = Призматична or Угрупування клітин або клітинні комплекси_Розташування = Багаточисельними округлими стурктурами or Угрупування клітин або клітинні комплекси_Форма = Папілярні

/*0,911736 0,001808 */ Діагноз_Цитологія = Папілярний рак ==> Угрупування клітин або клітинні комплекси_Розташування = Багаточисельними округлими стурктурами or Угрупування клітин або клітинні комплекси_Форма = Папілярні or Ядерце_Кількість = Одиничні дрібні ядерця

Приклад неповного правила:

/* 0,930113 0,000000 */ Клітина_Колір = Насичений and Клітина_Цитоплазма = Насичена цитоплазмою ==> Угрупування клітин або клітинні комплекси_Розташування = Багаточисельними округлими стурктурами or Хроматин_Тип = Сітчастий

Висновок

У роботі проведено аналіз алгоритмів інтелектуального аналізу даних та основних алгоритмів для побудови асоціативних правил. Розроблено узагальнений алгоритм для пошуку асоціативних правил та проведено порівняльний аналіз по критеріях і показниках даного експерименту. Дослідження проведено за допомогою програмного засобу WEKA, за рахунок можливостей пошуку асоціації та візуалізації результатів дослідження. В результаті отримано асоціативні правила для

діагностики передракових та ракових станів раку молочної залози та відповідні їм лінгвістичні змінні.

Список використаних джерел

1. Березький О.М. Інтелектуальна система для діагностування різних форм раку молочної залози на основі аналізу гістологічних та цитологічних зображень / О.М. Березький, Г.М. Мельник, Ю. М. Батько, Т. В. Дацко // Науковий вісник НЛТУ України - 2013. - № 23.13. - С. 357-367
2. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод: Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP: БХВ-Петербург, 2007
3. Дюк В.А., Самойленко А.П. Data Mining: учебный курс. -СПб.: Питер, 2001. - 368 с.

УДК 004.932.2:616-006.6

АЛГОРИТМИ ПОБУДОВИ НЕЧІТКИХ ПРОДУКЦІЙНИХ ПРАВИЛ НА ОСНОВІ АНАЛІЗУ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

Вербовий С.О.[1)], Мартинчук Т.О.[2)]

Тернопільський національний економічний університет

[1)] аспірант; [2)] магістрант

I. Постановка проблеми

Основним методом цитологічного та гістологічного досліджень клітин, тканин органів є світлова мікроскопія. Для кількісного опису мікрооб'єктів використовують такі ознаки: площа та периметр клітин, геометричні ознаки форми, а для кількісного опису патологічних змін у структурах використовують кількість шарів клітин у тканині, коефіцієнт структурної атипії та інші. При тестуванні програмного забезпечення аналізу зображень використовуються тестові бази цитологічних та гістологічних зображень із поставленим діагнозом, недоліком яких є відсутність детального опису мікрооб'єктів у якісних категоріях [1]. Тому актуальною задачею є розробка алгоритмів для побудови нечітких продукційних правил для бази даних цитологічних і гістологічних зображень, що містить якісні та кількісні ознаки мікрооб'єктів.

II. Мета роботи

Метою роботи є аналіз існуючих алгоритмів побудови нечітких продукційних правил на основі аналізу біомедичних зображень.

III. Нечітка система побудови продукційних правил діагностування диспластичних процесів молочної залози

Нечітка система на основі продукційних правил є найбільш розповсюдженою при моделюванні складних систем, тому що вона використовує лінгвістичні змінні. Лінгвістичні змінні можуть бути природним чином представлені в нечітких множинах та у ролі логічних зв'язків цих множин. Нечіткий рівень розуміння і опису складної системи виражається у вигляді набору обмежень на виході за рахунок певних умов вводу. Обмеження, як правило, моделюються нечіткими множинами та зв'язками типу «AND», «OR», «THEN» [2].

В якості експериментальних досліджень використано тестову навчальну вибірку цитологічних та гістологічних зображень [3]. Вхідними змінними є геометричні ознаки даних зображень, а саме розмір клітини та її форма. Змінна "розмір клітини" включає такі елементи терм-множини: малі (small), середні (medium) та великі (large). Змінна "форма клітини" включає елементи терм-множини: циліндрична (cylindrical), кубічна (cubic), овальна (oval).

Функція належності базується на отриманих експертом знаннях з мікрооб'єктів та їх числових ознаках. Після проведених експериментальних досліджень виміру розмірів та форми нормальних клітин та одного із диспластичних процесів, а саме проліферативна мастопатія, отримано числові значення, які наведені в таблиці 1.