Wrocław, 2020

Komputerowe wspomaganie diagnozowania stanów ostrego brzucha z wykorzystaniem algorytmu k-NN

PROJEKT – Zastosowanie informatyki w medycynie

AUTORZY:

Karolina Filipiak, 226008

Karol Maśluch, 235044

Spis treści

[1. Opis problemu medycznego 3](#_Toc38578528)

[2. Ranking cech 5](#_Toc38578529)

# Opis problemu medycznego

Naszym zadaniem było nabycie umiejętności zastosowania oraz implementacji algorytmu, który będzie wspomagał diagnostykę stanów ostrego brzucha. W tym celu, należało skorzystać z algorytmu minimalno-odległościowego k-NN, bazując na zbiorze danych dostarczonych przez prowadzącego.

Na początku przystąpiliśmy do zdefiniowania problemu medycznego. Na podstawie materiału empirycznego oraz implementacji odpowiedniego mechanizmu, należało dobrać taki zbiór cech, dzięki któremu z jak największym prawdopodobieństwem uda się wystawić prawidłową diagnozę dla pacjenta. Problemem było również określenie optymalnej liczby cech, dla których uda się wystawić trafną diagnozę pacjentowi.

Analizując dokument z danymi dotyczącymi stanów ostrego brzucha, wyodrębniliśmy:

* 476 pacjentów ze stwierdzonym stanem ostrego brzucha, gdzie każdy z pacjentów został opisany zestawem 31 cech, na podstawie których zdiagnozowano u nich jedną z 8 odmian powyższej przypadłości
* 31 cech – atrybuty/parametry uzyskane z wywiadu z pacjentem oraz badania wstępnego. Wszystkie poniżej opisane cechy mają charakter dyskretny (skokowy), ponieważ są opisane konkretnymi, dopuszczalnymi wartościami i jednocześnie nie mogą być wyrażone wartościami pośrednimi z danego przedziału (przykład: cecha „Apetyt” może przyjmować wartość 1, 2 lub 3 i nie może przyjmować żadnej innej wartości spoza tego przedziału - np. 1,5).

*Tabela 1. Opis cech.*

|  |  |  |  |
| --- | --- | --- | --- |
| Lp. | Nazwa cechy | Dopuszczalne wartości | Klasyfikacja cechy |
| 1 | Płeć | 1, 2 | Symptomy ogólne |
| 2 | Wiek | 1, 2, 3, 4, 5 |
| 3 | Lokalizacja bólu na początku zachorowania | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 | Ból |
| 4 | Lokalizacja bólu obecnie | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 |
| 5 | Intensywność bólu | 1, 2, 3 |
| 6 | Czynniki nasilające ból | 1, 2, 3, 4 |
| 7 | Czynniki przynoszące ulgę | 1, 2, 3 |
| 8 | Progresja bólu | 1, 2, 3 |
| 9 | Czas trwania bólu | 1, 2, 3, 4 |
| 10 | Charakter bólu na początku zachorowania | 1, 2, 3 |
| 11 | Charakter bólu obecnie | 1, 2, 3, 4 |
| 12 | Nudności i wymioty | 1, 2, 3 | Oddawanie moczu |
| 13 | Apetyt | 1, 2, 3 |
| 14 | Wypróżnienia | 1, 2, 3 |
| 15 | Oddawanie moczu | 1, 2 |
| 16 | Poprzednie niestrawności | 1, 2 | Historia |
| 17 | Żółtaczka w przeszłości | 1, 2 |
| 18 | Poprzednie operacje brzuszne | 1, 2 |
| 19 | Leki | 1, 2 |
| 20 | Stan psychiczny | 1, 2, 3 | Ogólne badanie |
| 21 | Skóra | 1, 2, 3 |
| 22 | Temperatura (pacha) | 1, 2, 3, 4, 5, 6 |
| 23 | Tętno | 1, 2, 3, 4, 5, 6, 7, 8, 9 |
| 24 | Ruchy oddechowe powłok brzusznych | 1, 2 | Oglądanie brzucha |
| 25 | Wzdęcia | 1, 2 |
| 26 | Umiejscowienie bolesności uciskowej | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 | Badania palpacyjne |
| 27 | Objaw Blumberga | 1, 2 |
| 28 | Obrona mięśniowa | 1, 2 |
| 29 | Wzmożone napięcie powłok brzusznych | 1, 2 |
| 30 | Opory patologiczne | 1, 2 |
| 31 | Objaw Murphy’ego | 1, 2 |

* 8 klas – dana klasa oznacza diagnozę uzyskaną na podstawie kombinacji cech

*Tabela 2. Opis klas.*

|  |  |  |
| --- | --- | --- |
| Lp. | Klasa (diagnoza) | Ilość wystąpień |
| 1 | Ostre zapalenie wyrostka robaczkowego | 141 |
| 2 | Zapalenie uchyłków jelit | 17 |
| 3 | Niedrożnośc mechaniczna jelit | 29 |
| 4 | Perforowany wrzód trawienny | 28 |
| 5 | Zapalenie woreczka żółciowego | 55 |
| 6 | Ostre zapalenie trzustki | 32 |
| 7 | Niecharakterystyczny ból brzucha | 157 |
| 8 | Inne przyczyny ostrego bólu brzucha | 17 |

*Wykres 1. Rozkład procentowy częstości występowania danej przypadłości w stanach ostrego brzucha.*

# Ranking cech

Każdy odrębny przypadek stanu ostrego brzucha jest opisany stałą ilością cech i na ich podstawie jest formułowana diagnoza. Daną diagnozę opisuje pewien zbiór cech wraz z konkretnymi wartościami tych cech, które są charakterystyczne dla danej klasy. Pewne wartości mogą być określone jako te, które z większym prawdopodobieństwem będą wskazywać na daną diagnozę, podczas gdy niektóre wartości są traktowane tylko jako pomocnicze, które mogą jedynie sugerować daną diagnozę. W związku z tym stworzyliśmy ranking cech, który posortował dane atrybuty w kolejności od najbardziej znaczących do najmniej znaczących przy kwalifikacji do diagnozy. W tym celu skorzystaliśmy z metody rekurencyjnej eliminacji cech. Za pomocą zewnętrznego estymatora wagi są przepisywane do poszczególnych cech, a następnie cechy o najgorszym współczynniku są eliminowane. Ostatnia wyeliminowana cecha zajmuję najwyższą pozycję w rankingu, wyznacza najlepszą cechę pod względem otrzymanego współczynnika. Jako estymator zewnętrzny wykorzystaliśmy współczynnik modelu liniowego.

*Tabela 3. Ranking cech.*

|  |  |  |
| --- | --- | --- |
| Pozycja w rankingu: | Numer cechy: | Nazwa cechy: |
| 1 | 27 | Objaw Blaumberga |
| 2 | 24 | Ruchy oddechowe powłok brzusznych |
| 3 | 12 | Nudności i wymioty |
| 4 | 19 | Leki |
| 5 | 21 | Skóra |
| 6 | 9 | Czas trwania bólu |
| 7 | 31 | Objaw Murphy’ego |
| 8 | 7 | Czynniki przynoszące ulgę |
| 9 | 20 | Stan psychiczny |
| 10 | 14 | Wypróżnienia |
| 11 | 6 | Czynniki nasilające ból |
| 12 | 5 | Intensywność bólu |
| 13 | 11 | Charakter bólu obecnie |
| 14 | 8 | Progresja bólu |
| 15 | 2 | Wiek |
| 16 | 30 | Opory patologiczne |
| 17 | 17 | Żółtaczka w przeszłości |
| 18 | 23 | Tętno |
| 19 | 29 | Wzmożone napięcie powłok brzusznych |
| 20 | 18 | Poprzednie operacje brzuszne |
| 21 | 15 | Oddawanie moczu |
| 22 | 26 | Umiejscowienie bolesności uciskowej |
| 23 | 13 | Apetyt |
| 24 | 22 | Temperatura (pacha) |
| 25 | 10 | Charakter bólu na początku zachorowania |
| 26 | 25 | Wzdęcia |
| 27 | 4 | Lokalizacja bólu obecnie |
| 28 | 1 | Płeć |
| 29 | 16 | Poprzednie niestrawności |
| 30 | 3 | Lokalizacja bólu na początku zachorowania |
| 31 | 28 | Obrona mięśniowa |

Dodatkowo w celu eliminacji mniej znaczących cechy otrzymane z rankingu cech skrzyżowaliśmy z dodatkową selekcją cech, którą była jednoczynnikową selekcją wyłaniającą   
k-najlepszych cech. Jako czynnik wybraliśmy metrykę X2 (chi kwadrat) . Jako k obraliśmy 25 cech. Tabela 4 zawiera ranking wybranych cech po skrzyżowaniu. Ostatecznie zostało wybrane 24 cechy z oryginalnych 31.

*Tabela 4. Ranking cech po skrzyżowaniu.*

|  |  |  |
| --- | --- | --- |
| Pozycja w rankingu: | Numer cechy: | Nazwa cechy: |
| 1 | 27 | Objaw Blaumberga |
| 2 | 24 | Ruchy oddechowe powłok brzusznych |
| 3 | 12 | Nudności i wymioty |
| 4 | 19 | Leki |
| 5 | 21 | Skóra |
| 6 | 9 | Czas trwania bólu |
| 7 | 31 | Objaw Murphy’ego |
| 8 | 7 | Czynniki przynoszące ulgę |
| 9 | 20 | Stan psychiczny |
| 10 | 14 | Wypróżnienia |
| 11 | 6 | Czynniki nasilające ból |
| 12 | 5 | Intensywność bólu |
| 13 | 11 | Charakter bólu obecnie |
| 14 | 8 | Progresja bólu |
| 15 | 2 | Wiek |
| 17 | 17 | Żółtaczka w przeszłości |
| 18 | 23 | Tętno |
| 19 | 29 | Wzmożone napięcie powłok brzusznych |
| 20 | 15 | Oddawanie moczu |
| 21 | 26 | Umiejscowienie bolesności uciskowej |
| 22 | 13 | Apetyt |
| 23 | 22 | Temperatura (pacha) |
| 24 | 10 | Charakter bólu na początku zachorowania |

# Informacje o środowisku implementacyjnym

Badania zostały przeprowadzone przy pomocy języka Python w wersji 3.7.3. Do stworzenia rankingu cech przy użyciu

# Opis badań eksperymentalnych

## Parametry klasyfikacji

## Badania