

Anna KOTULLA
Politechnika Śląska, Instytut Informatyki

NEGATYWNE REGUŁY ASOCJACYJNE – WYZNACZANIE, MIARY I OBSZARY ZASTOSOWANIA

Streszczenie. Artykuł opisuje pozytywne oraz negatywne reguły asocjacyjne. Przybliżone zostały najważniejsze miary dla reguł asocjacyjnych. Przykładowa analiza przeprowadzona została za pomocą środowiska R. Opisana została klasyfikacja bazująca na pozytywnych i negatywnych regułach asocjacyjnych.

Słowa kluczowe: analiza danych, eksploracja danych, analiza podobieństw, pozytywne reguły asocjacyjne, negatywne reguły asocjacyjne

NEGATIVE ASSOCIATION RULES – COMPUTING, MEASURES AND APPLICATION AREAS

Summary. This article presents positive and negative association rules. The most important measures for association rules are described. A sample analysis was done using the R environment. Classification based on positive and negative association rules was described.

Keywords: data analysis, data mining, affinity analysis, positive association rules, negative association rules

1. Wprowadzenie

Jedną z metod eksploracji danych jest analiza podobieństw [11, 13] (ang. *affinity analysis*), znana również pod nazwą analizy koszyka sklepowego (ang. *market basket analysis*) czy analizy koszykowej (ang. *basket analysis*). Analiza podobieństw jest metodą mającą na celu odkrycie zależności pomiędzy dwoma – lub więcej – atrybutami ukrytymi w dużych zbiorach danych. Wykryte zależności mogą być prezentowane w postaci tak zwanych reguł asocjacyjnych [2] (ang. *association rules*).

Pewne formy analizy koszyka sklepowego zagościły na dobre w internetowym świecie. Portale internetowe powszechnie starają się zachęcić do zakupu dalszych produktów czy do zatrzymania się dłużej na portalu, oferując odnośniki do powiązanych tematycznie zasobów – internauci stykają się z informacjami typu „Inni klienci oglądali również...”¹, „Ci, którzy kupili ten towar, kupili też...”².

W niniejszym opracowaniu opisane zostały pozytywne i negatywne reguły asocjacyjne. Reguły asocjacyjne wyznaczone zostały dla dwóch przykładowych zbiorów danych.

Przybliżona została klasyfikacja jako dziedzina, w której znajdują zastosowanie negatywne reguły asocjacyjne.

Opisane zostało również środowisko R , zwłaszcza dostępne w nim narzędzia umożliwiające stosowanie reguł asocjacyjnych w analizie podobieństw.

2. Podstawowe pojęcia

2.1. Reguły asocjacyjne

Niech $I = \{i_1, i_2, \dots, i_k\}$ będzie zbiorem k literalów, nazywanych elementami. Niech D będzie zbiorem transakcji T , nazywanym bazą danych, takim, że T jest podzbiorem I . Transakcja T zawiera X wtedy i tylko wtedy, gdy X jest podzbiorem T . Jako regułę asocjacyjną [6, 11, 13] rozumie się wyrażenie „jeżeli wystąpi X , to Y wystąpi z pewnym prawdopodobieństwem”, formalnie zapisywane jako:

$$X \rightarrow Y, \quad (1)$$

gdzie X oraz Y są podzbiórmi właściwymi I oraz zbiorami rozłącznymi ($X \cap Y = \emptyset$).

Ważne metryki dla reguł asocjacyjnych to:

- wsparcie s (ang. *support*) – jako wsparcie reguły asocjacyjnej rozumiana jest tzw. częstość względna, czyli stosunek liczby transakcji zawierających X oraz Y do całkowitej liczby transakcji N :

$$s(X \rightarrow Y) = \frac{n(X \cup Y)}{N}, \quad (2)$$

- ufność c (ang. *confidence*) – pod terminem ufność kryje się miara dokładności reguły asocjacyjnej, czyli stosunek liczby transakcji zawierających X oraz Y do liczby transakcji zawierających X :

$$c(X \rightarrow Y) = \frac{n(X \cup Y)}{n(X)} = \frac{s(X \cup Y)}{s(X)}. \quad (3)$$

¹ Za <http://merlin.pl>.

² Za <http://merlin.pl>.

Zazwyczaj pożądane są reguły o dużym wsparciu i dużej ufności.

Do oceny reguł asocjacyjnych używane są także inne miary, przykładowo miara *lift* [4], dla której wyższe wartości oznaczają silniejsze asocjacje:

$$l(X \rightarrow Y) = \frac{s(X \cup Y)}{s(X)s(Y)}. \quad (4)$$

Popularnym narzędziem do odkrywania reguł asocjacyjnych jest algorytm Apriori [3], zastępowany obecnie przez bardziej efektywny algorytm FPGrowth [3].

Istotnym problemem w przypadku reguł asocjacyjnych jest fakt, iż wraz ze wzrostem liczby atrybutów, liczba możliwych reguł rośnie wykładniczo, w związku z czym dla większych zbiorów danych nie jest możliwe sprawdzenie wszystkich reguł. Liczbę możliwych reguł asocjacyjnych R , wygenerowanych dla zbioru danych zawierającego d atrybutów, przedstawić można jako [13]:

$$R = 3^d - 2^{d+1} + 1. \quad (5)$$

Reguły postaci $X \rightarrow Y$ to pozytywne reguły asocjacyjne.

2.2. Negatywne reguły asocjacyjne

Niech $I = \{i_1, i_2, \dots, i_k\}$ będzie zbiorem k literałów, nazywanych elementami. Niech D będzie zbiorem transakcji T , nazywanym bazą danych, takim, że T jest podzbiorem I . Transakcja T zawiera X wtedy i tylko wtedy, gdy X jest podzbiorem T . Niech X oraz Y będą podzbiórmi I , oraz zbiorami rozłącznymi ($X \cap Y = \emptyset$).

Jeżeli reguła postaci $X \rightarrow Y$ rozumiana jest jako pozytywna reguła asocjacyjna, to jako negatywne reguły asocjacyjne w niniejszym opracowaniu, bazując na [5,10,16], rozumiane są reguły postaci:

- „jeżeli wystąpi X , to Y nie wystąpi z pewnym prawdopodobieństwem”, formalnie:

$$X \rightarrow \neg Y, \quad (6)$$

- „jeżeli nie wystąpi X , to Y wystąpi z pewnym prawdopodobieństwem”, formalnie:

$$\neg X \rightarrow Y, \quad (7)$$

- „jeżeli nie wystąpi X , to Y nie wystąpi z pewnym prawdopodobieństwem”, formalnie:

$$\neg X \rightarrow \neg Y. \quad (8)$$

Wsparcie dla negatywnych reguł asocjacyjnych wyrazić można następująco:

$$s(\neg X) = 1 - s(X) \quad (9)$$

$$s(X \rightarrow \neg Y) = s(X) - s(X \cup Y) \quad (10)$$

$$s(\neg X \rightarrow Y) = s(Y) - s(X \cup Y) \quad (11)$$

$$s(\neg X \rightarrow \neg Y) = 1 - s(X) - s(Y) + s(X \cup Y). \quad (12)$$

Ufność dla negatywnych reguł asocjacyjnych wyrazić można następująco:

$$c(X \rightarrow \neg Y) = \frac{s(X) - s(X \cup Y)}{s(X)} \quad (13)$$

$$c(\neg X \rightarrow Y) = \frac{s(Y) - s(X \cup Y)}{1 - s(X)} \quad (14)$$

$$c(\neg X \rightarrow \neg Y) = \frac{1 - s(X) - s(Y) + s(X \cup Y)}{1 - s(X)}. \quad (15)$$

W przypadku analizy koszyka zakupów, negatywne reguły asocjacyjne pomóc mogą w wyłonieniu produktów, które zazwyczaj nie są kupowane razem. Załóżmy, że znaleziona reguła asocjacyjna postaci (6) mówi, że klienci, którzy kupują śmietanę, nie kupują jogurtu z obniżoną zawartością tłuszczu. Jeżeli wsparcie dla reguły jest odpowiednio wysokie, sklep może oferować produkty jednej i drugiej kategorii (śmietana oraz jogurt z obniżoną zawartością tłuszczu) równocześnie w promocji, licząc na przyciągnięcie do sklepu obydwu grup klientów.

3. Środowisko R

R [17] jest wolnym i otwartym językiem programowania oraz środowiskiem do obliczeń statystycznych. R dostępny jest na podstawie licencji *GNU GPL*³. R oferuje szeroką paletę funkcji przydatnych m.in. do eksploracji danych oraz wizualizacji wyników. Dostępne w R narzędzia mogą być poszerzone o dodatkowe skrypty czy pakiety.

Do testów zainstalowane zostało środowisko R w wersji 2.14.1, udostępnionej 22.12.2011.

Pakiet *arules* [8, 9] dostarcza narzędzia do reprezentowania, przetwarzania i analizowania danych dotyczących transakcji i wzorców. Pakiet ten dostarcza również interfejs do algorytmów *Apriori* oraz *Eclat*.

Do testów wykorzystana została wersja 1.0-7 pakietu, z dnia 04.11.2011.

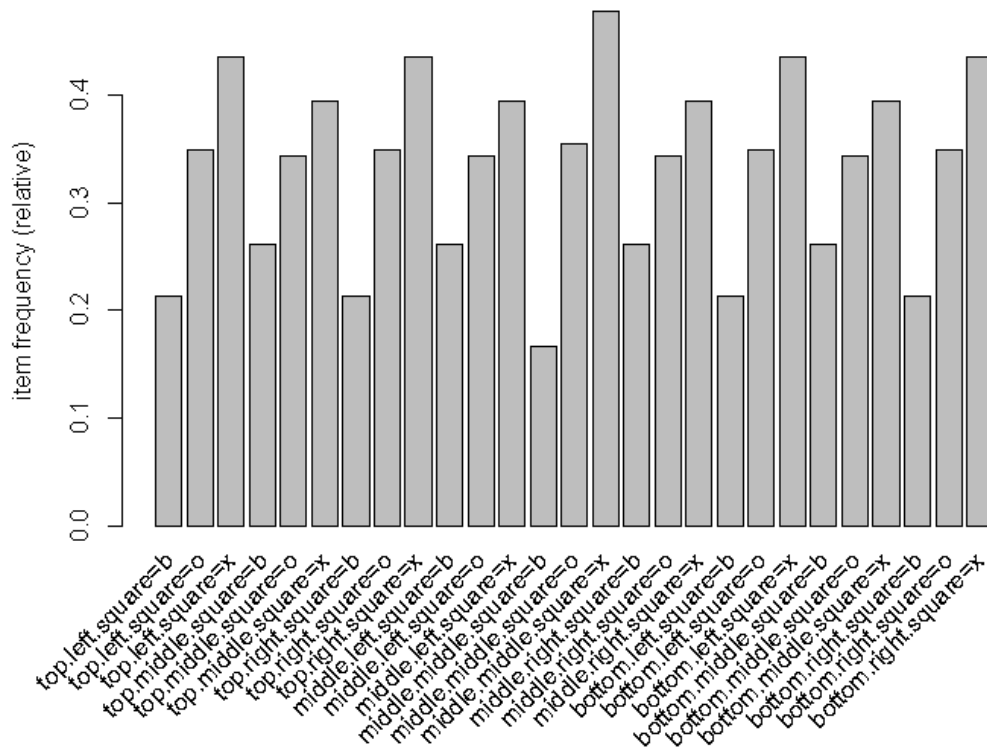
³ Licencje GNU General Public License dostępne są pod adresem <http://www.gnu.org/licenses/> (sprawdzone 12.01.2012).

4. Przeprowadzone testy

4.1. Test ze zbiorem „Tic-Tac-Toe Endgame Data Set”

Jako pierwszy zbiór danych testowych użyty został zbiór „Tic-Tac-Toe Endgame Data Set”, dostępny na stronach UCI Machine Learning Repository⁴. Zbiór zawiera informacje o popularnej grze kółko/krzyżyk, zawierając wszystkie 958 możliwych konfiguracji. Zbiór dobrze nadaje się do wstępnych testów, ponieważ liczba atrybutów jest mała (9 atrybutów⁵). Zbiór zawiera również informację, czy wynik był pozytywny czy negatywny – dane rozpatrywane są pod kątem wygranej dla x .

Histogram na rys. 1 pokazuje częstość poszczególnych cech.



Rys. 1. Częstość występowania cech, zbiór „Tic-Tac-Toe Endgame Data Set”

Fig. 1. Frequency of attributes, data set “Tic-Tac-Toe Endgame Data Set”

Dane zostały wczytane do środowiska R i przekształcone do binarnej macierzy incydencji:

```
> daneT<-as(dane,"transactions")
> daneT
transactions in sparse format with
```

⁴ Zbiór dostępny jest pod adresem <http://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/> (sprawdzono 12.01.2012).

⁵ Dostępne atrybuty: *top-left-square*, *top-middle-square*, *top-right-square*, *middle-left-square*, *middle-middle-square*, *middle-right-square*, *bottom-left-square*, *bottom-middle-square*, *bottom-right-square*, dla każdego z atrybutów podana jest wartość: x (gdy gracz x postawił w tym polu znak), o (gdy gracz o postawił w tym polu znak) lub b (gdy pole jest puste).

958 transactions (rows) and
27 items (columns)

Pozytywne reguły asocjacyjne wyznaczone zostały za pomocą algorytmu Apriori. Liczbę wygenerowanych reguł przy minimalnym wsparciu 1%, dla różnych poziomów ufności, przedstawiono w tabeli 1.

Wygenerowane reguły asocjacyjne zostały zapisane do pliku tekstowego za pomocą funkcji WRITE. Na podstawie znanych pozytywnych reguł asocjacyjnych utworzone zostały negatywne reguły asocjacyjne, dla których obliczono wsparcie (9), (10), (11), (12) i ufność (13), (14), (15).

Tabela 1

Liczba wygenerowanych reguł asocjacyjnych

Poziom ufności	10%	20%	30%	40%	50%
Liczba reguł	43051	41598	33404	22815	14744
Poziom ufności	60%	70%	80%	90%	95%
Liczba reguł	6134	2412	704	292	108

4.1.1. Przykładowa negatywna reguła asocjacyjna postaci (6)

Za pomocą środowiska R wygenerowana została reguła mówiąca, że jeżeli lewy górny narożnik jest pusty oraz w prawy dolnym narożniku znajduje się znak postawiony przez gracza x , to wygrywa gracz x :

```
{top.left.square=b,bottom.right.square=x} => {Class=positive}
```

Dla powyższej reguły wsparcie wynosi 0,0709812108559499, poziom ufności to 0,739130434782609.

Reguła negatywna postaci (6) mówi, że jeżeli lewy górny narożnik jest pusty oraz w prawy dolnym narożniku znajduje się znak postawiony przez gracza x , to gracz x przegrywa:

```
{top.left.square=b,bottom.right.square=x} => {Class=negative};
```

Dla powyższej reguły wsparcie wynosi 0,0250521920668058, poziom ufności to 0,260869565217391.

Na podstawie wartości wsparcia oraz ufności dla powyższych dwóch reguł bardziej prawdopodobna jest wygrana gracza x przy opisanej wartości pól

```
{top.left.square=b,bottom.right.square=x}
```

4.1.2. Przykładowe negatywne reguły asocjacyjne postaci (7)

Wygenerowana została reguła mówiąca, że jeżeli gracz x postawił znak w środkowym polu, to gracz x wygrywa:

```
{middle.middle.square=x} => {Class=positive}
```

Dla powyższej reguły wsparcie wynosi 0,382045929018789, poziom ufności to 0,799126637554585.

W przypadku negatywnej reguły asocjacyjnej postaci (7) jest istotne, aby w środkowym polu nie znajdował się znak postawiony przez gracza x , a przy tym gracz x wygrał. Zatem wyznaczone mogą być dwie negatywne reguły asocjacyjne, ponieważ możliwe jest:

- $\{middle.middle.square=o\}$, to znaczy że w środkowym polu znak postawił gracz o . Wówczas negatywna reguła asocjacyjna o wsparciu 0,154488517745303 oraz poziomie ufności 0,435294117647059 ma postać:

```
{middle.middle.square=o} => {Class=positive}
```

- $\{middle.middle.square=b\}$, to znaczy że środkowe pole jest puste. Wówczas negatywna reguła asocjacyjna o wsparciu 0,116910229645094 oraz poziomie ufności 0,7 ma postać:

```
{middle.middle.square=b} => {Class=positive}
```

4.1.3. Przykładowe negatywne reguły asocjacyjne postaci (8)

Wygenerowana została reguła mówiąca, że jeżeli gracz x wygrywa, to w dolnym lewym polu znajduje się znak postawiony przez gracza x :

```
{Class=positive} => {bottom.left.square=x}
```

Dla powyższej reguły wsparcie wynosi 0,307933194154488, poziom ufności to 0,471246006389776.

W przypadku negatywnych reguł asocjacyjnych postaci (8) interesujące są te reguły, które mówią, że gracz x nie wygrywa oraz w dolnym lewym polu nie znajduje się znak postawiony przez gracza x . Możliwe są więc następujące reguły:

- jeżeli gracz x przegrywa, to dolne lewe pole jest puste. Wówczas negatywna reguła o wsparciu 0,0657620041753653 oraz poziomie ufności 0,189759036144578 ma postać:

```
{Class=negative} => {bottom.left.square=b}
```

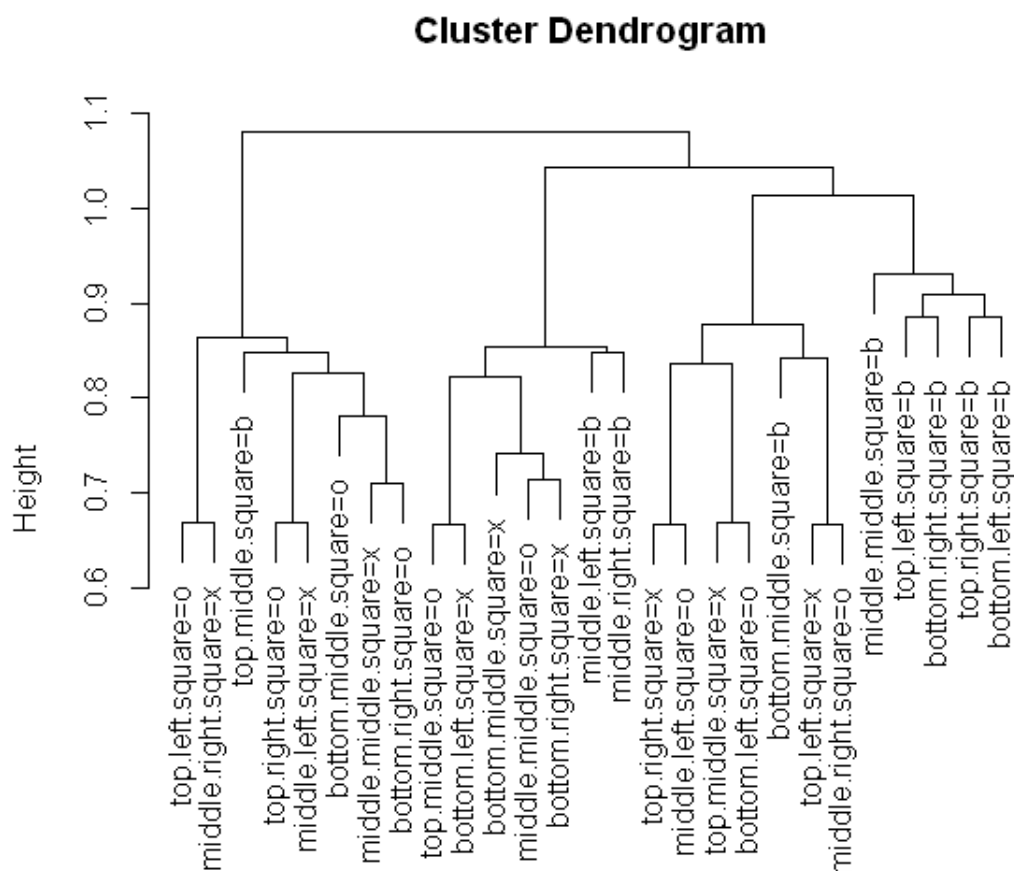
- jeżeli gracz x przegrywa, to w dolnym lewym polu znajduje się znak postawiony przez gracza o – pole jest puste. Wówczas negatywna reguła asocjacyjna o wsparciu 0,152400835073069 oraz poziomie ufności 0,439759036144578 ma postać:

```
{Class=negative} => {bottom.left.square=o}
```

4.1.4. Przykład – klasteryzacja hierarchiczna

W celu zademonstrowania możliwości pakietu R, dla danych testowych przeprowadzona została klasteryzacja hierarchiczna metodą minimalnej wariancji według Warda. Wygenero-

wany został dendrogram – rys. 2. Do klasteryzacji wykorzystane zostały tylko te grupy, które spotykane są w conajmniej 1% transakcji.



Rys. 2. Dendrogram dla “Tic-Tac-Toe Endgame Data Set”

Fig. 2. Dendrogram for “Tic-Tac-Toe Endgame Data Set”

4.2. Test ze zbiorem „Mushroom Database”

Jako zbiór danych testowych użyty został zbiór „Mushroom Database”, dostępny na stronach UCI Machine Learning Repository⁶. W zbiorze zawarte są informacje o 8124 grzybach, zawierające dla każdego grzyba wartości 22⁷ cech opisujących wygląd, zapach, rozmieszczenie, podłoże oraz informację⁸, czy jest jadalny, czy trujący.

Dane zostały wczytane do środowiska R i zostały przekształcone do binarnej macierzy in-cydencji:

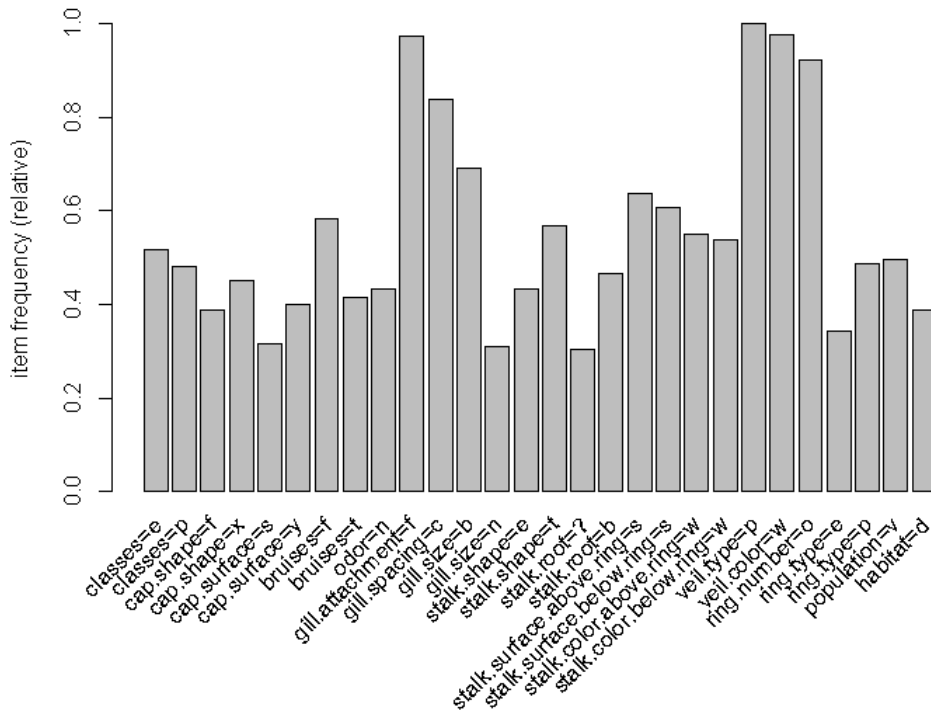
⁶ Zbiór dostępny jest pod adresem <http://archive.ics.uci.edu/ml/datasets/Mushroom> (sprawdzono 12.01.2012).

⁷ Uwzględnione zostały następujące cechy: *cap-shape*, *cap-surface*, *cap-color*, *bruises*, *odor*, *gill-attachment*, *gill-spacing*, *gill-size*, *gill-color*, *stalk-shape*, *stalk-root*, *stalk-surface-above-ring*, *stalk-surface-below-ring*, *stalk-color-above-ring*, *stalk-color-below-ring*, *veil-type*, *veil-color*, *ring-number*, *ring-type*, *spore-print-color*, *population*, *habitat*.

⁸ Występuje pod terminem *classes*.


```
> grzybyT<-as(grzyby,"transactions")
> grzybyT
transactions in sparse format with
8124 transactions (rows) and
119 items (columns)
```

Histogram na rys. 3 pokazuje częstość poszczególnych cech. Aby diagram był czytelny, ograniczono się do takich wartości, dla których wsparcie wynosi przynajmniej 30%.



Rys. 3. Częstość występowania cech dla zbioru „Mushroom Database”, wsparcie min. 30%

Fig. 3. Frequency of attributes for data set “Mushroom Database”, support min. 30%

Pozytywne reguły asocjacyjne wyznaczone zostały za pomocą algorytmu Apriori, przy wsparciu równym 1% oraz ufności równej 60%:

```
> reguly1<-apriori(grzybyT,parameter = list(support = 0.1, confidence = 0.6));
```

Podstawowe informacje na temat wygenerowanych reguł:

```
> summary(reguly1)
set of 3430601 rules

rule length distribution (lhs + rhs):sizes
 1  2  3  4  5  6  7  8  9 10
 8 569 7805 45380 153486 353970 605373 797473 817364 649173

  Min. 1st Qu. Median Mean 3rd Qu. Max.
1.000  7.000  8.000  8.034  9.000 10.000

summary of quality measures:
 support confidence lift
Min. :0.1002 Min. :0.6000 Min. :0.6827
1st Qu.:0.1064 1st Qu.:1.0000 1st Qu.:1.0849
Median :0.1064 Median :1.0000 Median :1.7110
Mean :0.1242 Mean :0.9806 Mean :1.8619
3rd Qu.:0.1142 3rd Qu.:1.0000 3rd Qu.:2.1515
Max. :1.0000 Max. :1.0000 Max. :6.8718
```

```

mining info:
  data ntransactions support confidence
  grzybyT 8124 0.1 0.6

```

Liczba wygenerowanych reguł asocjacyjnych jest duża – 3430601 reguł. Do wyodrębnienia podzbiorów, np. grzybów jadalnych oraz trujących, posłużyć może funkcja *subset()*. Przykładowe reguły asocjacyjne przedstawiono w tabeli 2.

Tabela 2

Przykładowe wygenerowane reguły asocjacyjne

Reguła	Wsparcie	Ufność	Miara <i>lift</i>
{cap.shape=k} => {veil.type=p}	1	1	1.0
{habitat=l} => {ring.number=o}	0.1024126	1	1.084936
{habitat=l} => {veil.type=p}	0.1024126	1	1.000000
{stalk.shape=e, ring.type=l} => {gill.size=b}	0.1595273	1	1.447612

Wyznaczone reguły asocjacyjne zostały zapisane do pliku tekstowego za pomocą funkcji WRITE. Na podstawie znanych pozytywnych reguł asocjacyjnych utworzone zostały negatywne reguły asocjacyjne, dla których obliczono wsparcie (9), (10), (11), (12) i ufność (13), (14), (15).

4.2.1. Przykładowa negatywna reguła asocjacyjna postaci (6)

Znaleziona została reguła asocjacyjna postaci

```
{veil.type=p,veil.color=w,ring.number=o} => {classes=p},
```

dla której wsparcie wynosi 0,467749876907927, a poziom ufności to 0,521405049396268. Powyższa reguła mówi, że jeżeli kapelusz grzyba jest częściowo połączony z trzonem (*veil.type=p[artial]*), ma kolor biały (*veil.color=w[hite]*) oraz grzyb ten posiada jeden pierścień (*ring.number=o[ne]*), to jest trujący.

Negatywna reguła asocjacyjna postaci (6) mówi, że jeżeli grzyb ma wymienione powyżej cechy, to jest jadalny:

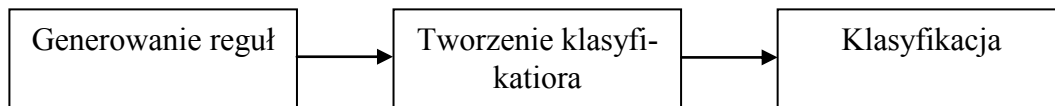
```
{veil.type=p,veil.color=w,ring.number=o} => {classes=e}
```

Dla wymienionej negatywnej reguły asocjacyjnej wsparcie wynosi 0,429345150172329, a poziom ufności to 0,478594950603732.

5. Klasyfikacja z zastosowaniem negatywnych reguł asocjacyjnych

Algorytmy klasyfikacyjne bazujące na asocjacjach [1, 12, 14, 15] zazwyczaj działają według schematu przedstawionego na rys. 4.

W czasie generowania reguł asocjacyjnych selekcjonowane są często występujące wzorce, zawierające reguły klasyfikacyjne. W fazie tworzenia klasyfikatora usuwane są powtarzające się reguły, a przydatne reguły są sortowane. W ostatnim kroku dokonywana jest klasyfikacja danych.



Rys. 4. Algorytm klasyfikacyjny bazujący na asocjacjach
Fig. 4. Associative classification algorithm

6. Podsumowanie

W artykule omówiono pozytywne oraz negatywne reguły asocjacyjne. Przedstawiono najważniejsze miary określające reguły asocjacyjne. Wygenerowane zostały reguły asocjacyjne dla przykładowych zbiorów danych. Omówiony został sposób utworzenia negatywnych reguł asocjacyjnych, przedstawione zostały przykładowe negatywne reguły asocjacyjne.

Nie jest wskazane rozszerzenie przestrzeni atrybutów przez dodatnie negatywnych atrybutów, ponieważ liczba możliwych reguł asocjacyjnych rośnie wykładniczo. Wspomnieć trzeba, że już dla zbioru „Mushroom Database” wystąpił problem – zbyt mało dostępnej pamięci (testy wykonywano na komputerze z 2 gigabajtami RAM). Dla pokazanych w artykule przykładów negatywne reguły asocjacyjne generowane były na podstawie utworzonych pozytywnych reguł asocjacyjnych, poza środowiskiem R. Planowane jest zaimplementowanie w środowisku R algorytmów umożliwiających bezpośrednio wyznaczanie także negatywnych reguł asocjacyjnych. W dalszej perspektywie planowane jest również zaimplementowanie w tymże środowisku klasyfikatora bazującego na asocjacjach.

Ciekawym obszarem zastosowań negatywnych reguł asocjacyjnych jest wykrywanie anomalii – planowane są badania w tej dziedzinie.

BIBLIOGRAFIA

1. Antonie M.-L., Zaïane O. R.: An Associative Classifier based on Positive and Negative Rules. 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, Paryż 2004.
2. Agrawal R., Imielinski T., Swami A.: Mining Association Rules between Sets of Items in Large Databases. SIGMOD Conference Washington, D.C. 1993, s. 207÷216.

3. Agrawal R., Srikant R.: Fast Algorithms for Mining Association Rules. VLDB Conference, Santiago de Chile 1994, s. 487÷499.
4. Brin S., Motwani R., Ullman J. D., Tsur S.: Dynamic Itemset Counting and Implication Rules for Market Basket Data. ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA 1997, s. 255÷264.
5. Ghaderi R., Minaei-Bidgoli B.: Detecting Data Errors with Employing Negative Association Rules. International Journal of Digital Content Technology and its Applications, Vol. 3, No. 3, 2009, s. 91÷95.
6. Han J., Kamber M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco 2006.
7. Han J., Pei J., Yin Y., Mao R.: Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. Data Mining and Knowledge Discovery, Kluwer Academic Publishers Hingham, MA 2004.
8. Hahsler M., Buchta Ch., Gruen B., Hornik K.: arules: Mining Association Rules and Frequent Itemsets. Manual, <http://CRAN.R-project.org/> (sprawdzono 12.01.2012).
9. Hahsler M., Gruen B., Hornik K.: arules – A Computational Environment for Mining, Association Rules and Frequent Item Sets. Journal of Statistical Software, Vol. 14, No.15, 2005, s. 1÷25.
10. Kavitha Rani B., Srinivas K., Ramasubbareddy B., Govardhan A.: Mining Negative Association Rules. International Journal of Engineering and Technology, Vol. 3(2), 2011, s. 100÷105.
11. Markov Z., Larose D. T.: Eksploracja zasobów internetowych. Wydawnictwo Naukowe PWN, Warszawa 2009.
12. Ramasubbareddy B., Govardhan A., Ramamohanreddy A.: Classification Based on Positive and Negative Association Rules. International Journal of Data Engineering, Vol. 2, 2011.
13. Tan P.-N., Steinbach M., Kumar V.: Introduction to Data Mining. Addison-Wesley, 2005.
14. Thabtah F., Cowling P., Peng Y.: MMAC: A New Multi-class, Multi-label Associative Classification Approach. 4th IEEE International Conference on Data Mining, Brighton UK 2004, s. 217÷224.
15. Thabtah F., Cowling P., Peng Y.: MCAR: Multi-class Classification Based on Association Rule Approach. 3rd IEEE International Conference on Computer Systems and Applications, Kair 2005, s. 1÷7.

16. Wu X., Zhang Ch., Zhang Sh.: Efficient Mining of Both Positive and Negative Association Rules. ACM Transactions on Information Systems, Vol. 22, No. 3, 2004, s. 381÷405.
17. The R Foundation for Statistical Computing: The R Project for Statistical Computing, <http://www.r-project.org> (sprawdzono 12.01.2012).

Wpłynęło do Redakcji 15 stycznia 2012 r.

Abstract

Affinity analysis [11, 13] is a data mining method; know also as market basked analysis or basket analysis. Affinity analysis explores dependencies between two or more attributes for large data sets. Those dependencies could be presented as association rules [2].

This paper describes positive and negative association rules. Association rules were explored for two sample data sets, „Tic-Tac-Toe Endgame Data Set” and „Mushroom Database”.

Classification is one of the areas, where the association rules can be used, to build an associative classifier.

The open programming language and environment R was described and used for both tests. Especially methods implemented in the package arules [8, 9] were used.

It is planned to implement in R algorithms to compute negative rules directly and also to implement an associative classifier in R.

Another interesting area is the application of negative association rules for the detection of anomalies.

Adres

Anna KOTULLA: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, anna.kotulla@polsl.pl.