

Zmodyfikowane zadanie 1.1. z listy 3.

Bazy danych 2023/24

Dominik Bojko
KPI WIT Politechnika Wrocławska

15 stycznia 2024

Zadanie 1. Dla danego zbioru zależności funkcyjnych podaj postać minimalną. Udowodnij, że $\{\mathcal{F}_{\min}\}^+ = \{\mathcal{F}\}^+$ oraz, że żadna z pozostałych zależności nie jest nadmiarowa.

$$\mathcal{F} = \{A \rightarrow DE, C \rightarrow D, AD \rightarrow F, BD \rightarrow C, AEG \rightarrow BCF, CEG \rightarrow A, BG \rightarrow D, DE \rightarrow EG, A \rightarrow E\}$$

Znajdź klucze główne.

Idea: W zbiorach zależności funkcyjnych, drukowane litery oznaczają kolumny, a strzałka — zależność, które kolumny (następnik relacji) można wyciągnąć, posiadając informacje z kolumn z poprzednika relacji. Innymi słowy: zależność relacyjna $XYZ \rightarrow TV$ oznacza, że z kolumn X, Y oraz Z, jesteśmy w stanie wywnioskować wartości w kolumnach T oraz V.

Można myśleć, że $\mathcal{F}$ obrazuje zależności w tabeli bazy danych. Chcielibyśmy możliwie zmniejszyć postać zbioru zależności relacyjnych, do tak zwanej postaci minimalnej $\mathcal{F}_{\min}$, to znaczy takiej, że żadnej z występujących relacji nie można uprościć, ale tak, żeby wszystkie możliwe zależności funkcyjne, które można uzyskać z $\mathcal{F}_{\min}$ (czyli elementy domknięcia zbioru zależności funkcyjnych $\{\mathcal{F}_{\min}\}^+$) były takie jak wszystkie możliwe zależności funkcyjne, które można uzyskać z wyjściowego zbioru $\mathcal{F}$ (tj. $\{\mathcal{F}\}^+$). Można powiedzieć, że celem znalezienia postaci minimalnej jest dojście do postaci, która spełnia drugą część zadania.

Po wykonaniu tej części zadania, łatwo znaleźć klucze główne. Potrzeba znaleźć takie zestawy kolumn, że ich domknięcie daje całą relację. Kluczami głównymi będą te zestawy, z których po wyjęciu dowolnej kolumny, domknięcie przestaje być całą relacją.

Rozwiązanie: Pierwszym krokiem zazwyczaj jest rozbicie następników zależności funkcyjnych (pomaga to w unikaniu błędów w dalszej części):

$$\mathcal{F}_0 = \{A \rightarrow D, A \rightarrow E, C \rightarrow D, AD \rightarrow F, BD \rightarrow C, AEG \rightarrow B, AEG \rightarrow C, AEG \rightarrow F, \\ CEG \rightarrow A, BG \rightarrow D, DE \rightarrow E, DE \rightarrow G, A \rightarrow E\}.$$

Następnie usuwamy zależności trywialne (takie, w których następnik jest zawarty w poprzedniku), czyli w tym przypadku $DE \rightarrow E$ oraz ewentualne powtórzenia (tu $A \rightarrow E$). Potem, dobrym trickiem jest sprawdzenie, które prawe strony są unikalne. Jeśli takie są, to na pewno pojawiają się w postaci minimalnej, bo w żaden inny sposób nie będziemy w stanie znaleźć takich zależności funkcyjnych. W naszym przykładzie będą to $A \rightarrow E$ (bo usunęliśmy już $DE \rightarrow E$), $AEG \rightarrow B$, $CEG \rightarrow A$, $DE \rightarrow G$.

$$\mathcal{F}_1 = \{A \rightarrow D, \underline{A \rightarrow E}, C \rightarrow D, AD \rightarrow F, BD \rightarrow C, \underline{AEG \rightarrow B}, AEG \rightarrow C, AEG \rightarrow F, \\ \underline{CEG \rightarrow A}, BG \rightarrow D, \underline{DE \rightarrow G}\}.$$

Podobnie, można sprawdzić, czy jest tylko jedna zależność funkcyjna, której poprzednikiem jest konkretna pojedyncza kolumna. Takie zależności muszą zostać w postaci minimalnej, gdyż usunięcie jej nie pozwalałoby wywnioskować czegokolwiek z tej pojedynczej kolumny. U nas takie jest $C \rightarrow D$ (nie ma innych zależności postaci $C \rightarrow \cdot$):

$$\mathcal{F}_1 = \{A \rightarrow D, \underline{A \rightarrow E}, \underline{C \rightarrow D}, AD \rightarrow F, BD \rightarrow C, \underline{AEG \rightarrow B}, AEG \rightarrow C, AEG \rightarrow F, \\ \underline{CEG \rightarrow A}, BG \rightarrow D, \underline{DE \rightarrow G}\}.$$

Następnie, najlepiej rozważać czy domknięcie zbioru zależności funkcyjnych po usunięciu pojedynczej zależności jest takie jak domknięcie obecnego zbioru zależności funkcyjnych. Najbezpieczniej zacząć od (niepodkreślonych) zależności, których

poprzednikiem jest jedna kolumna. Gdy takich nie ma, to wypada stopniowo zwiększać liczbę kolumn w poprzednikach relacji. Rozważmy $\mathcal{F}'_1 = \mathcal{F}_1 \setminus \{A \rightarrow D\}$. Sprawdźmy, czy $\mathcal{F}'_1{}^+ = \mathcal{F}_1^+$. Należy szukać domknięcia $\{A\}^+$ w $\mathcal{F}'_1$: mamy $A \rightarrow E$, wobec czego $\{AE\} \subset \{A\}^+$. A z lewej strony występuje tylko raz jako singleton, a AE nie występuje jako poprzednik żadnej zależności funkcyjnej, czyli $\{AE\} = \{A\}^+$. Oznacza to, że nie możemy się pozbyć $A \rightarrow D$. Rozważamy kolejną zależność (nie mamy już pojedynczych kolumn po lewych stronach): $AD \rightarrow F$. Szukamy dopełnienia $\{AD\}^+$ w $\mathcal{F}_1 \setminus \{AD \rightarrow F\}$. $\{AD\}^+ = \{ADE\}^+$, bo $A \rightarrow E$, a następnie otrzymujemy $\{ADEG\}^+$, ponieważ $DE \rightarrow G$. Dalej z $AEG \rightarrow BCF$, otrzymujemy $\{AD\}^+ = \{ABCDEFG\}$. Dostaliśmy w szczególności $\{AD \rightarrow F\}$, czyli ta zależność jest zbędna:

$$\mathcal{F}_2 = \{\underline{A \rightarrow D}, \underline{A \rightarrow E}, \underline{C \rightarrow D}, \underline{BD \rightarrow C}, \underline{AEG \rightarrow B}, \underline{AEG \rightarrow C}, \underline{AEG \rightarrow F}, \underline{CEG \rightarrow A}, \underline{BG \rightarrow D}, \underline{DE \rightarrow G}\}.$$

Jako, że w dopełnieniu $\{AD\}^+$ otrzymaliśmy całą relację, to oznacza, $\{AD\}$ jest podkluczem. D nie występuje w lewych stronach zależności jako singleton (czyli $\{D\}^+ = \{D\}$), więc $\{D\}$ nie jest kluczem głównym (minimalnym podkluczem w sensie inkluzji). Natomiast jest nim $\{A\}$, bo $\{AD\} \subset \{A\}^+$, a $\{AD\}$ jest podkluczem, czyli $\{A\}^+ = \{AD\}^+ = \{ABCDEFG\}$. Podsumowując: pojedynczy atrybut A jest kluczem głównym, bo jego domknięcie stanowi całą relację.

Wracając do pierwszej części naszego zadania, kolejno, powinniśmy rozważać usunięcie zależności $BD \rightarrow C$ oraz $BG \rightarrow D$. Można zauważyć, że będą one w postaci minimalnej z prostej przyczyny: w obu przypadkach, kolumny, który występują po lewych stronach, nie występują w lewych stronach innych zależności zarówno jako singletony, jak i w parach BD oraz BG . Oznacza to, że po usunięciu którejkolwiek z nich, $\{BD\}^+ = \{BD\}$ (czy w drugim przypadku $\{BG\}^+ = \{BG\}$):

$$\mathcal{F}_3 = \{\underline{A \rightarrow D}, \underline{A \rightarrow E}, \underline{C \rightarrow D}, \underline{BD \rightarrow C}, \underline{AEG \rightarrow B}, \underline{AEG \rightarrow C}, \underline{AEG \rightarrow F}, \underline{CEG \rightarrow A}, \underline{BG \rightarrow D}, \underline{DE \rightarrow G}\}.$$

Gdyby jednak istniała przykładowo zależność $B \rightarrow A$, to należałoby oba te przypadki rozważyć.

Zauważmy, że w momencie, gdy usunęliśmy zależność $AD \rightarrow F$, pozostała nam jedna zależność z prawą stroną F , tj. $AEG \rightarrow F$. Zatem ze względu na wcześniej użyty trick, zostanie ona w postaci minimalnej:

$$\mathcal{F}_4 = \{\underline{A \rightarrow D}, \underline{A \rightarrow E}, \underline{C \rightarrow D}, \underline{BD \rightarrow C}, \underline{AEG \rightarrow B}, \underline{AEG \rightarrow C}, \underline{AEG \rightarrow F}, \underline{CEG \rightarrow A}, \underline{BG \rightarrow D}, \underline{DE \rightarrow G}\}.$$

Pozostała zależność $AEG \rightarrow C$ do rozważenia. Zamiast rozważać odpowiednie dopełnienie $\{AEG\}^+$, możemy przeprowadzić rozumowanie abdukcyjne. Istotnie, C możemy otrzymać tylko z $BD \rightarrow C$. B możemy otrzymać tylko (z naszego wyjściowego AEG) poprzez $AEG \rightarrow B$. Natomiast D można otrzymać na trzy sposoby, jednak nie chcemy używać do tego C , czyli można wykorzystać $A \rightarrow D$ lub $BG \rightarrow D$. Oznacza to, że $\{C\} \subset \{AEG\}^+$ w $\mathcal{F}_3 \setminus \{AEG \rightarrow C\}$, zatem można się pozbyć roważanej zależności funkcyjnej:

$$\mathcal{F}_5 = \{\underline{A \rightarrow D}, \underline{A \rightarrow E}, \underline{C \rightarrow D}, \underline{BD \rightarrow C}, \underline{AEG \rightarrow B}, \underline{AEG \rightarrow F}, \underline{CEG \rightarrow A}, \underline{BG \rightarrow D}, \underline{DE \rightarrow G}\}.$$

Ostatnim krokiem jest pozbycie się zbędnych części zależności, występujących po lewych stronach. Jako że $A \rightarrow E$, to możemy uprościć AEG do AG . Co więcej, z $A \rightarrow D$ i $A \rightarrow E$, dostajemy DE , a z $DE \rightarrow G$ mamy $A \rightarrow G$, czyli AG możemy uprościć do samego A . Podobnie, z $C \rightarrow D$ oraz $DE \rightarrow G$, mamy $CE \rightarrow G$, czyli CEG można uprościć do CE :

$$\mathcal{F}_{\min} = \{\underline{A \rightarrow B}, \underline{A \rightarrow D}, \underline{A \rightarrow E}, \underline{A \rightarrow F}, \underline{C \rightarrow D}, \underline{BD \rightarrow C}, \underline{CE \rightarrow A}, \underline{BG \rightarrow D}, \underline{DE \rightarrow G}\}.$$

Oczywiście ten krok można rozbić na części i wykonywać w międzyczasie.

Wobec uwagi na wstępie, pokazaliśmy tym samym, że $\mathcal{F}^+ = \mathcal{F}_{\min}^+$.

Wcześniej pokazaliśmy, że $\{A\}$ jest kluczem głównym. Poszukajmy innych. Jeśli klucz główny składałby się z jednego atrybutu, to byłby singletonem po lewej stronie którejś z zależności. Policzmy $\{C\}^+ = \{CD\}$, czyli $\{C\}$ nie jest kluczem. Teraz należy szukać kluczy podwójnych i potrójnych, które nie zawierają $\{A\}$: Jednak, jest tylko jedna zależność funkcyjna, która daje A : $CE \rightarrow A$. Wobec tego $\{CE\}$ jest kluczem głównym. Jako że $BD \rightarrow C$ i $BG \rightarrow D$, to również $\{BDE\}$ oraz $\{BGE\}$ są kluczami głównymi.