Бази от данни

Функционални зависимости (втора част)

доц. д-р Димитър Димитров

Преговор

- $A \to B \Leftrightarrow$ за всеки два кортежа от R е изпълнено, че ако те съвпадат по атрибутите A, то те съвпадат и по B
- Аксиоми на Армстронг
  - A1. Ако $Y \subseteq X$, то $X \to Y$
  - A2. Ако $X \to Y$, то $XW \to YW$
  - A3. Ако $X \to Y$ и $Y \to Z$, то $X \to Z$
- Следствия

Обвивка
на множество от ФЗ

- F – множество от ФЗ
- F^+ – обвивка на F
- F^+ е множеството от ФЗ, които логически следват от F
- По-късно ще видим и по-добра дефиниция

Задача

- $F = \{ AB \to C, CD \to E \}$
- $ABD \to E \in F^{+}$?

Решение

- $AB \rightarrow C$ (дадено)
- $ABD \rightarrow CD$ (A2)
- $CD \rightarrow E$ (дадено)
- $ABD \rightarrow E$ (A3)

Надеждност (soundness) на аксиомите на Армстронг

- Лема: аксиомите на Армстронг са надеждни
- т.е. ако дадена ФЗ е изведена от *F* чрез аксиомите, то тя е вярна за всяка релация, в която важат *F*

Пример: приложение на аксиомите на Армстронг

- Дадено: $A \to B$, $B \to C$, $A \to D$, $CE \to HG$
- Какви ФЗ можем да изведем?

Пример: приложение на аксиомите на Армстронг

- Дадено: $A \rightarrow B$, $B \rightarrow C$, $A \rightarrow D$, $CE \rightarrow HG$
- $A \rightarrow B$, $B \rightarrow C \Rightarrow A \rightarrow C$ (транзитивност)
- $A \rightarrow B$, $A \rightarrow D \Rightarrow A \rightarrow BD$ (обединение)
- $CE \rightarrow HG$, $A \rightarrow C \Rightarrow AE \rightarrow HG$ (псевдотранзитивност)
- $AE \rightarrow HG \Rightarrow AE \rightarrow H$, $AE \rightarrow G$ (декомпозиция)

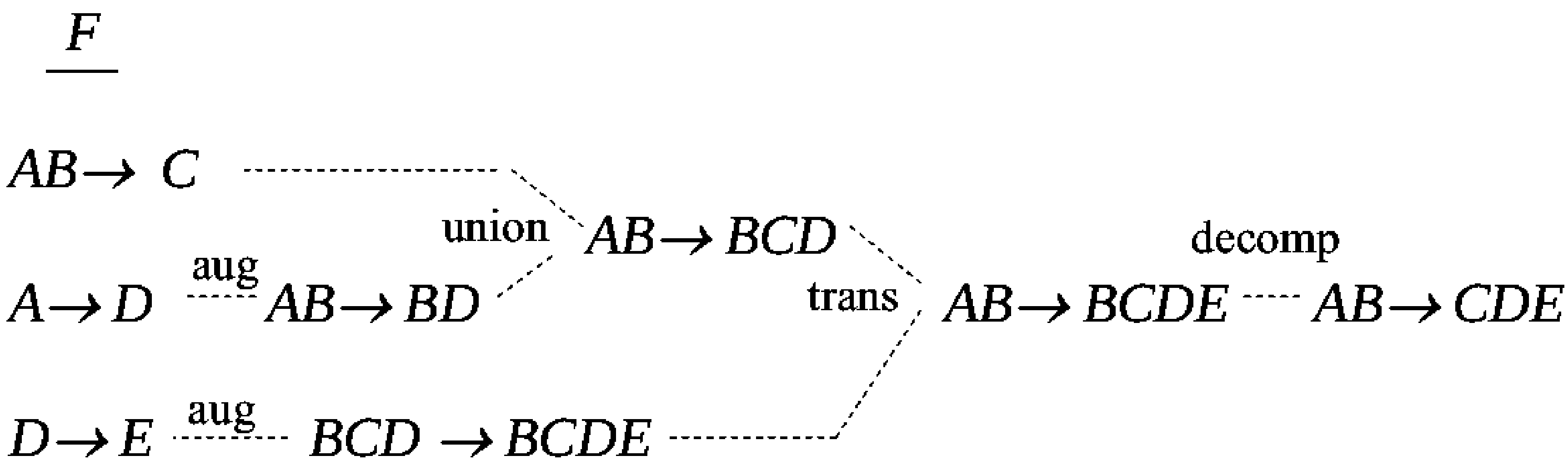
Пълнота (completeness) на аксиомите на Армстронг

- Теорема:
- Аксиомите на Армстронг са надеждни и пълни

Следствие: еквивалентна дефиниция на F^+

- F – множество от ФЗ
- F^+ – обвивка на F
- F^+ е множеството от ФЗ, които могат да бъдат извлечени от F чрез аксиомите на Армстронг

В търсене на F^+ - пример



- AB → BD, AB → BCD, AB → BCDE, AB → CDE са елементи на F^+

Обвивка на атрибут

- Attribute closure
- X е атрибут или множество от атрибути
- F е множество от ФЗ
- Обвивка на X е множеството от атрибути
  $X^+ = \cup \{ Y \mid X \rightarrow Y \in F^+ \}$

Свойство

- $X \subseteq X^+$

Алгоритъм за намиране на обвивка на атрибут

Вход: X, F

Closure := X

do

  PreviousClosure := Closure

  for each $A \to B$ in F do

    if $A \subseteq$ Closure and $B \not\subseteq$ Closure

    then Closure := Closure $\cup$ B

while PreviousClosure $\neq$ Closure

Резултат: Closure

Задача

- Дадено:

  Релация R(A,B,C,D,E,F)

  ФЗ:

  $AB \rightarrow C$

  $BC \rightarrow AD$

  $D \rightarrow E$

  $CF \rightarrow B$

- Търси се:

  $\{A,B\}^+ = ?$

Решение

- Започваме с {A,B}
- Имаме AB→C, разширяваме до {A,B,C}
- Имаме BC→AD, разширяваме до {A,B,C,D}
- D→E, разширяваме до {A,B,C,D,E} и повече не можем да продължаваме

Едно приложение на обвивките на атрибути

- Можем да проверим дали ФЗ следва от множеството от ФЗ F
- ФЗ $X \rightarrow Y$ следва от множеството от ФЗ F т.с.т.к. $Y \subseteq X^+$

Примери

- Дадени са R(A,B,C,D,E,F) и множество от ФЗ F:

  $AB \rightarrow C$

  $BC \rightarrow AD$

  $D \rightarrow E$

  $CF \rightarrow B$

- $AB \rightarrow D$ следва ли от F?
  - Да, защото $D \in \{A,B\}^+ = \{A,B,C,D,E\}$
- $D \rightarrow A$ следва ли от F?

Преговор: ключ и суперключ

- Множеството от атрибути X на релацията $R(A_1, A_2, \ldots, A_n)$ е ключ на R, ако:

  1) $X \rightarrow A_1 A_2 \ldots A_n$

  2) за никое $Y \subset X$ не е изпълнено $Y \rightarrow A_1 A_2 \ldots A_n$

- Суперключ

Обвивки и суперключове

- Множеството от атрибути X на релацията R(A_1, A_2, …, A_n) е суперключ на R, т.с.т.к. $X^+=\{A_1, A_2, \ldots, A_n\}$

Доказателство

- X е суперключ, следователно $X \to A_1, A_2, \ldots, A_n$, следователно $A_1, A_2, \ldots, A_n \in X^+$
- Обратното – $X^+ = \{A_1, A_2, \ldots, A_n\}$, следователно $X \to A_1, A_2, \ldots, A_n$, следователно X е суперключ

Коректност и пълнота на алгоритъма за обвивка

- Алгоритъмът за намиране на обвивка на множество от атрибути намира:
  - Само верни ФЗ
  - Всички верни ФЗ

Коректност

- Ако $B \in A^+$, то $A \to B$ е в сила
- Може да се докаже индуктивно

Пълнота

- Алгоритъмът за A^+ генерира всички ФЗ, които могат да се изведат от F
  - Няма други, които да са изводими от F и валидни в R
  - Т.е. не съществува $A \rightarrow B$, валидна за R, за която $B \notin A^+$
- Може да се докаже с допускане на противното, конструирайки релация с два кортежа, съвпадащи по A^+ и несъвпадащи по останалите атрибути

Еквивалентни множества от ФЗ

- Множествата от ФЗ *F* и *G* се наричат еквивалентни, ако $F^+ = G^+$
- Пример:
  $\{AB \rightarrow C, A \rightarrow B\}$
  $\{A \rightarrow C, A \rightarrow B\}$
- Множествата от екземпляри релации, които удовлетворяват съответно F и G, съвпадат

Покритие
на множество от ФЗ

- Покритие на F:
  всяко множество от ФЗ, което е еквивалентно на F^+

Минимално покритие (1)

- F^+ съдържа много ФЗ
- Множеството от ФЗ F е минимално, ако:

1. За всяка ФЗ $X \to Y$ от F
   Y е единичен атрибут
2. Ако от F отстраним която и да е ФЗ, няма да получим еквивалентно множество ФЗ
3. Ако от F сменим която и да е ФЗ $X \to A$ с $Y \to A$, където $Y \subseteq X$, няма да получим еквивалентно множество ФЗ

Минимално покритие (2)

- Всяко множество от ФЗ F е еквивалентно на някакво минимално множество F′

Пример

- R(A,B,C)
- Всеки атрибут функционално определя другите два атрибута

  $F = \{A \to B, A \to C, B \to A, B \to C, C \to A, C \to B, AB \to C, AC \to B, BC \to A, A \to BC, B \to AC, C \to AB\}$

- Минимални покрития:
- $\{A \to B, B \to C, C \to A\}$,
  $\{A \to B, B \to A, B \to C, C \to B\}$, ...

ФЗ при проекция (1)

- Проекция
- Нека R1(A,B,C,D) има следните ФЗ:
- $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$
- Нека премахнем атрибута B
  - Защо бихме го направили? Ще разберем в следваща лекция
- Кои ФЗ важат в новата релация R2(A,C,D)?
- $C \rightarrow D$?

ФЗ при проекция (2)

- Да, но не само $C \to D$
- $A \to C$?
- Трябва да видим кои ФЗ следват от дадените и да изберем тези, които включват само атрибути на R2

ФЗ при проекция (3)

- Универсален подход: намираме обвивките на всички подмножества на {A,C,D}
  - Без $\emptyset$ и {A,C,D}, понеже не могат да предизвикат нетривиални ФЗ
  - Ако $X^+=\{A,C,D\}$, няма да открием нови ФЗ от подмножества, съдържащи X
- Започваме от единични атрибути, после множества от два атрибута и т.н.
- За всяко подмножество X добавяме ФЗ от вида $X \rightarrow Y$, където $Y \in X^+$, но $Y \notin X$ (за да избегнем тривиални ФЗ)

ФЗ при проекция (4)

- $\{A\}^+=\{A,B,C,D\}$, получаваме $A \to C$, $A \to D$
  - $A \to B$ включва атрибути, липсващи в R2, затова я пропускаме
- $\{C\}^+=\{C,D\}$, получаваме $C \to D$
- $\{D\}^+=\{D\}$ – не получаваме нищо
- От оптимизацията следва, че няма нужда да обхождаме множествата, съдържащи A
  - (няма да получим нови ФЗ, които да не са били вече открити)
- Единствено има смисъл да обходим {C,D}
  - $\{C,D\}^+=\{C,D\}$, но не получаваме нищо
- Краен резултат: $A \to C$, $A \to D$, $C \to D$

Намиране на всички (кандидат-)ключове

- Дадена е релация R(A_1,…,A_n) и множество от ФЗ F
- Ще ни бъде необходимо за следващи лекции
- Намиране на всички суперключове с пълно изчерпване
  - 2^n-1 множества от атрибути, за всеки намираме обвивката
- По-ефективни алгоритми
  - Идея за оптимизация:
    - Всеки кандидат-ключ трябва да съдържа атрибутите, които не се срещат в дясната страна на никоя ФЗ (група 1)
    - Ако атрибут се среща в дясна страна на ФЗ, но не и в лявата страна на нито една ФЗ, този атрибут не е част от никой ключ (група 2)
    - Така разделяме атрибутите на три групи
    - Намираме обвивката на група 1, ако съвпада с A_1,…,A_n, то атрибутите в група 1 са ключ
    - Ако не, постепенно пробваме да добавяме атрибути от група 3 към група 1 ...
  - Допълнителен материал: https://en.wikipedia.org/wiki/Candidate_key

Въпроси?

Следват задачи

Задача 1

- Дадено:

  $AB \rightarrow C$

  $A \rightarrow D$

  $D \rightarrow E$

  $AC \rightarrow B$

- Търси се:

  $\{A,B\}^+ = ?$

Решение на задача 1

- Започваме с {A,B}
- Имаме AB → C, разширяваме до {A,B,C}
- A → D, …
- Стигаме до $\{A,B\}^+ = \{A,B,C,D,E\}$

Задачи за намиране на всички ключове

Дадена е R(A,B,C,D) и множество от нетривиални ФЗ F, равно на…

a) $\emptyset$
(дайте пример за такава релация и неин екземпляр)

b) $\{A \to B, B \to C, C \to A\}$

Нека вече R(A,B,…,G):

c) $\{AB \to C, C \to A, B \to AD, D \to CE, E \to FG\}$

d) $\{AB \to C, B \to D, CD \to E, D \to A, C \to B, E \to F\}$

Отговори

a) ABCD
релация, получена при преобразуване на връзка “много към много” без собствени атрибути

b) Не трябва да пропускаме D…

c) Първа група – B, втора група – FG, B+={A,B,…,G} – единствен ключ е B

d) Първа група – G, втора – F, G не е ключ, пробваме последователно AG, BG, …, FG, намираме BG и CG