

# Vytěžování dat - Hledání množin častých položek a asociačních pravidel

Jan Hrdlička

November 6, 2011

- Cílem této úlohy je zjistit, které produkty zákazníci kupují společně a z kterých produktů lze vytvořit asociační pravidla

# Analýza nákupního koše

- Cílem této úlohy je zjistit, které produkty zákazníci kupují společně a z kterých produktů lze vytvořit asociační pravidla
- Jedná se o aplikaci úlohy "Vyhledávání častých množin" kterou znáte z přednášky.

- V matlabu načtěte soubor `marketBasket.mat` obsahující transakční databázi supermarketu vhodnou k analýze nákupního košíku.

- V matlabu načtěte soubor marketBasket.mat obsahující transakční databázi supermarketu vhodnou k analýze nákupního košíku.
- V daném souboru se nachází proměnné "tranDb" a "info".

- V matlabu načtěte soubor marketBasket.mat obsahující transakční databázi supermarketu vhodnou k analýze nákupního košíku.
- V daném souboru se nachází proměnné "tranDb" a "info".
- Proměnná "tranDb" je transakční databáze v maticové booleanovské formě. Každý řádek je transakce - jeden nákupní košík jednoho zákazníka.

- V matlabu načtete soubor marketBasket.mat obsahující transakční databázi supermarketu vhodnou k analýze nákupního košíku.
- V daném souboru se nachází proměnné "tranDb" a "info".
- Proměnná "tranDb" je transakční databáze v maticové booleanovské formě. Každý řádek je transakce - jeden nákupní košík jednoho zákazníka.
- Každý sloupec je jedna možná položka v košíku (item). Slovní popis těchto položek je v proměnné "info".

- Stáhněte si `aprioriFPM.m`, tato funkce vygeneruje množinu častých položek pro danou transakční databázi algoritmem `apriori`. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help aprioriFPM"`.



- Stáhněte si `aprioriFPM.m`, tato funkce vygeneruje množinu častých položek pro danou transakční databázi algoritmem `apriori`. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help aprioriFPM"`.
- Najděte množinu častých položek pro vámi zvolenou minimální relativní podporu. Tu odůvodněte.

- Stáhněte si `aprioriFPM.m`, tato funkce vygeneruje množinu častých položek pro danou transakční databázi algoritmem `apriori`. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help aprioriFPM"`.
- Najděte množinu častých položek pro vámi zvolenou minimální relativní podporu. Tu odůvodněte.
- Nalezené množiny vypište do souboru pomocí funkce `printFreqSets.m`

- Stáhněte si `aprioriFPM.m`, tato funkce vygeneruje množinu častých položek pro danou transakční databázi algoritmem `apriori`. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help aprioriFPM"`.
- Najděte množinu častých položek pro vámi zvolenou minimální relativní podporu. Tu odůvodněte.
- Nalezené množiny vypište do souboru pomocí funkce `printFreqSets.m`
- Mezivýsledek: Pro minimální relativní podporu rovnou 0.02 vám má vyjít 1302 častých množin.

- Spusťte aprioriFPM pro množinu položek lexikálně srovnanou sestupně podle častosti výskytu a srovnanou vzestupně. Porovnejte časy obou běhů.

- Spusťte aprioriFPM pro množinu položek lexikálně srovnanou sestupně podle častosti výskytu a srovnanou vzestupně. Porovnejte časy obou běhů.
- Výsledek popište a zdůvodněte.

- Použijte funkci `associationRules.m` pro vygenerování asociačních pravidel. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help associationRules"`.

# Asociační pravidla

- Použijte funkci `associationRules.m` pro vygenerování asociačních pravidel. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help associationRules"`.
- Vygenerujte všechna asociační pravidla pro vámi zvolenou minimální relativní spolehlivost. Tu odůvodněte.

# Asociační pravidla

- Použijte funkci `associationRules.m` pro vygenerování asociačních pravidel. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem `"help associationRules"`.
- Vygenerujte všechna asociační pravidla pro vámi zvolenou minimální relativní spolehlivost. Tu odůvodněte.
- Mezivýsledek: Pro minimální relativní spolehlivost rovnou 0.7 (a dříve danou minimální relativní podporu 0.02) vám vyjde 480 asociačních pravidel.



# Asociační pravidla

- Použijte funkci `associationRules.m` pro vygenerování asociačních pravidel. Způsob použití funkce `aprioriFPM.m` zjistíte příkazem "`help associationRules`".
- Vygenerujte všechna asociační pravidla pro vámi zvolenou minimální relativní spolehlivost. Tu odůvodněte.
- Mezivýsledek: Pro minimální relativní spolehlivost rovnou 0.7 (a dříve danou minimální relativní podporu 0.02) vám vyjde 480 asociačních pravidel.
- Nalezená pravidla vypište do souboru pomocí funkce `printRules.m`

# Faktické náležitosti protokolu

Váš protokol by měl obsahovat:

- Časté množiny položek pro vámi vybranou minimální relativní podporu, minimální relativní podporu, kterou jste použili, její zdůvodnění

# Faktické náležitosti protokolu

Váš protokol by měl obsahovat:

- Časté množiny položek pro vámi vybranou minimalní relativní podporu, minimální relativní podporu, kterou jste použili, její zdůvodnění
- Časy běhů funkce aprioriFPM pro obě lexikální řazení, váš komentář k časům

# Faktické náležitosti protokolu

Váš protokol by měl obsahovat:

- Časté množiny položek pro vámi vybranou minimální relativní podporu, minimální relativní podporu, kterou jste použili, její zdůvodnění
- Časy běhů funkce aprioriFPM pro obě lexikální řazení, váš komentář k časům
- Asociační pravidla pro vámi vybranou minimální relativní spolehlivost, minimální relativní spolehlivost, kterou jste použili, její zdůvodnění