Tutorial: Boyce-Codd Normal Form

Your company, Apasaja Private Limited, is commissioned by Toko Kopi Luwak to design relational schema for the management of their coffee beans, drinks, and cafes.

A coffee bean is fully identified by its unique brand name or a combination of its cultivar and region (since the same cultivar can be grown in different region). For instance, we may have a coffee bean named "The Waterfall" which comes from a Tabi cultivar grown in Colombia.

A drink can be made utilizing a particular coffee bean. The name of the drink is only unique for a particular coffee bean. This means, we can have an "Espresso" made with "The Waterfall" or "La Bella" (which is a Pacamara cultivar grown in Guatemala). The price of the drink is also recorded.

A branch identified by its branch name may then sell the drink. A drink may be sold by zero or more branches. A branch may sell zero or more drinks. Additionally, the address of the branch is also recorded. Lastly, for each drink sold by a branch, we record the quantity sold to see which branch is the most profitable.

We are only given an abstract schema for this application as follows.

$$R = \{A, B, C, D, E, F, G, H\}$$

$$\begin{aligned}\Sigma = \{\ & \{A\} \rightarrow \{C, E\},\quad \{A, B\} \rightarrow \{D\},\quad \{F\} \rightarrow \{H\},\quad \{C, E\} \rightarrow \{A\},\quad \{B, C, E\} \rightarrow \{D\},\\ & \{A, B, F\} \rightarrow \{D, G\},\quad \{B, C, E, F\} \rightarrow \{G\}\ \}\end{aligned}$$

This tutorial continues from the computation of candidate keys and minimal cover in "Tutorial: Functional Dependencies". You are advised to compute them before continuing.

Questions

Not all questions will be discussed during tutorial. You are expected to attempt them before coming to the tutorial. You may be randomly called to present your answer during tutorial. You are encouraged to discuss them on Canvas Discussion.

1. **Boyce-Codd Normal Form**.

   (a) Is R with Σ in BCNF?

> **Comments:**
> **No.**
> Let us look at the non-trivial functional dependencies of the form $X \to \{A\}$ where X is a set of attributes such that $X \to \{A\}$ can be *derived* from Σ.
> Consider $\{A\} \to \{C\}$. This is a BCNF violation because:
>
> - It is non-trivial (since $\{C\} \not\subseteq \{A\}$), and
> - $\{A\}$ is not a superkey (since $\{A\}^+ = \{A, C, E\} \subset R$).
>
> Note that another way to check if $\{A\}$ is a superkey is by checking if it is a *superset of a key*. Here, $\{A\}$ is not a superset of either $\{A, B, F\}$ or $\{B, C, E, F\}$. $\Rightarrow$ This is a simpler way to check superkey if we have computed the keys.

2. **Normalization**.

   (a) Decompose R with Σ into a lossless-join BCNF decomposition using the algorithm from the lecture.

> **Comments:**
> We have found that $\{A\} \to \{C\}$ violates BCNF condition; let's use it to decompose R into two fragments. Then we recursively check for violation and further decompose if needed.
>
> - $\{A\} \to \{C\}$ violates BCNF condition of R with Σ.
>   Compute $\{A\}^+ = \{A, C, E\}$ with Σ. Decompose R into the following two fragments.
>   - $R_1 = \{A, C, E\}$ with
> $\Sigma_1 = \{\ \{A\} \to \{C, E\},\ \ \{C, E\} \to \{A\}\ \}$
> R_1 with Σ_1 is in BCNF.
>   - $R_2 = \{A, B, D, F, G, H\}$ with
> $\Sigma_2 = \{\ \{A, B\} \to \{D\},\ \ \{A, B, F\} \to \{G\},\ \ \{F\} \to \{H\}\ \}$
> R_2 with Σ_2 is not in BCNF.
> $\{A, B\} \to \{D\}$ violates BCNF condition of R_2 with Σ_2.
>   Compute $\{A, B\}^+ = \{A, B, D\}$ with Σ_2. Decompose R_2 into the following two fragments.
>   * $R_3 = \{A, B, D\}$ with
> $\Sigma_3 = \{\ \{A, B\} \to \{D\}\ \}$
> R_3 with Σ_3 is in BCNF.
>   * $R_4 = \{A, B, F, G, H\}$ with
> $\Sigma_4 = \{\ \{A, B, F\} \to \{G\},\ \ \{F\} \to \{H\}\ \}$
> R_4 with Σ_4 is not in BCNF.
> $\{F\} \to \{H\}$ violates BCNF condition of R_4 with Σ_4.
>   Compute $\{F\}^+ = \{F, H\}$ with Σ_4. Decompose R_4 into the following two fragments. (... *see next page*)

- $R_5 = \{F, H\}$ with
  $\Sigma_5 = \{ \{F\} \to \{H\} \}$
  R_5 with Σ_5 is in BCNF.
- $R_6 = \{A, B, F, G\}$ with
  $\Sigma_6 = \{ \{A, B, F\} \to \{G\} \}$
  R_6 with Σ_6 is in BCNF.

The result is the decomposition into fragments $\delta = \{R_1, R_3, R_5, R_6\}$. We will write this as the following decomposition.

$$\{ \{A, C, E\},\ \{A, B, D\},\ \{F, H\},\ \{A, B, F, G\} \}$$

Note: For the projection, you may get more functional dependencies but they should be *equivalent* to the given projection. We show only the *minimal cover* of the projection.

(b) Is the result dependency preserving?

Comments:
Yes.
From the above decomposition, let us find the union of all the corresponding projections and denote it with $\Sigma_\cup$.

$$\{ \{A\} \to \{C, E\},\ \{C, E\} \to \{A\},\ \{A, B\} \to \{D\},\ \{F\} \to \{H\},\ \{A, B, F\} \to \{G\} \}$$

All of the functional dependencies in $\Sigma_\cup$ above can be easily derived from Σ. What we need to check is whether the functional dependencies in Σ can be derived from $\Sigma_\cup$.

- $\{A\} \to \{C, E\}$ — *is in* $\Sigma_\cup$
- $\{A, B\} \to \{D\}$ — *is in* $\Sigma_\cup$
- $\{F\} \to \{H\}$ — *is in* $\Sigma_\cup$
- $\{C, E\} \to \{A\}$ — *is in* $\Sigma_\cup$
- $\{B, C, E\} \to \{D\}$ — $\{B, C, E\}^+ = \{A, B, C, D, E\} \supseteq \{D\}$
- $\{A, B, F\} \to \{D, G\}$ — $\{A, B, F\}^+ = \{A, B, C, D, E, F, G, H\} \supseteq \{D, G\}$
- $\{B, C, E, F\} \to \{G\}$ — $\{B, C, E, F\}^+ = \{A, B, C, D, E, F, G, H\} \supseteq \{G\}$

Again, note that all attribute closures should be computed with $\Sigma_\cup$ and not Σ.

(c) From the decomposition and the mapping of the attributes and the letters, can you figure out the mapping of fragments to the entity/relationship sets as mentioned in the text description?

Comments:
First, let us map the attributes back following the answer from "Tutorial: Functional Dependencies".

- $\{A, C, E\}$ becomes {`name`, `cultivar`, `region`}
  ⇒ `beans` entity set.

- $\{A, B, D\}$ becomes {`name`, `dname`, `price`}

  $\Rightarrow$ `drink` weak entity set with attribute from `beans`.
- $\{F, H\}$ becomes {`bname`, `address`}

  $\Rightarrow$ `branch` entity set.
- $\{A, B, F, G\}$ becomes {`name`, `dname`, `bname`, `qty`}

  $\Rightarrow$ `sells` relationship set.

Notice how the keys in the fragments (according to the projection) corresponds to the identifying attributes.

Comments:

It is always good to connect concepts learnt in the later part of the course with concepts taught in the earlier part of the course. Here we see that normalization (through BCNF decomposition) results in the same tables as entity-relationship modelling.

There can still be differences as we still have choices on which identifying attributes become the primary keys. Additionally, normalization may miss out on some entity sets if they do not contribute non-trivial functional dependencies (e.g., an entity set where all attributes are identifying attributes).

This is because normalization only deals with contraints in the form of functional dependencies. On the other hand, entity-relationship modelling may take into account other constraints.

We should also note that **projection** is an error-prone process. We advise not to skip any steps when computing the projection. Otherwise, some functional dependencies that may be derived could be omitted by mistake. The 5 steps may be tedious, but they guarantee a correct result. Taking shortcuts does not guarantee correctness in all cases.

5 steps to compute projection of R with Σ onto X:

1. Find all subsets of attributes of X.
2. For each subset X', compute the closure (i.e., $\varphi_1 := \texttt{AttrClose}(X', \Sigma)$).
3. Keep only the relevant attributes (i.e., $\varphi_2 := \varphi_1 \cap X$).
4. Remove attributes that do not contribute new information (i.e., $\varphi_3 := \varphi_2 - X'$).
5. If φ_3 is not empty, form a functional dependency $X' \rightarrow \varphi_3$.

References

[1] S. Bressan and B. Catania. *Introduction to Database Systems*. McGraw-Hill Education, 2006. ISBN: 9780071246507.

[2] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. 2nd ed. Prentice Hall Press, 2008. ISBN: 9780131873254.

[3] Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. 2nd. USA: McGraw-Hill, Inc., 2000. ISBN: 0072440422.