International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 4, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Finding an Optimal Cover for a Kind of Set of Functional Dependencies in Polynomial Time

Xiaoning Peng, Zhijun Xiao

***Abstract* – A smaller cover makes numerous algorithms (e.g., the classic synthesizing algorithm) need less storage space and less run time. Maier proves that the problem of finding an optimal cover (possible fewest attributes) is NP-complete. For a *single-ended* set of functional dependencies (the left side of every functional dependency has a single attribute), it is shown here that the optimal cover of a single-ended set of functional dependencies can be found in polynomial time, using the notion of *mini cover*. A simplified graphical representation (FD-graph) for a set of functional dependencies is introduced. It is provable that the transitive reduction of the FD-graph of a single-ended set of functional dependencies can be transformed into corresponding mini cover and further optimal cover.**

***Keywords* – Functional Dependency, Mini Cover, Optimal Cover, Relational Database, Transitive Reduction.**

I. INTRODUCTION

Data dependencies are crucial integrity constraint in the relational model [1]. Functional dependency (FD) is an important form of data dependencies. Given a set of FDs *F*, the closure of *F*, denoted by F^+, is the set of all FDs that can be inferred from *F*. Given sets of FDs *F* and *H*, *F* is a cover of *H* or *F* and *H* are equivalent if $F^+=H^+$. Given a set of FDs *F*, one is often interested in knowing whether there is a set of FDs *H* with *H* has fewer FDs or fewer attributes than *F* such that $F^+=H^+$. Minimum cover (possible fewest FDs) can be found in polynomial time [2], [3], [4]. Maier proves that the problem of finding an optimal cover (possible fewest attributes) is NP-complete [3]. However, for a single-ended set of FDs, an optimal cover can be found in polynomial time, which is the *focus* of our work.

The optimal cover problem is NP-complete, and the related excellent works are rare. Cotelea [5] attempts to propose a problem decomposition method which shows that the optimal cover problem can be broken down into some smaller problems while these problems seem can be solved easily. However, a counterexample is shown [6] and thus this method is not correct. Mannila et al. [7] deeply reveal the relationship between minimum cover and optimal cover, and introduce that in some cases, the size of minimum cover is much larger than corresponding optimal cover, which shows that the minimum cover should be refined so that it can be much closer to corresponding optimal cover.

The *motivation* to find an optimal cover is:

- Less storage space. The output of numerous algorithms (e.g., the classic synthesizing algorithm [8]) can be refined (less relations that need to be stored).
- Less run time. The efficiency of numerous algorithms depends on the size of a set of FDs. For example, the membership problem [9] can be solved in time $O(\|F\|)$ ($\|F\|$ denotes the number of attributes in *F*) for a set of FDs *F*. Ausiello et al. [4] construct a directed graph (FD-graph) to represent a set of FDs. We simplify this representation for a single-ended set of FDs.

The *process* of finding an optimal cover for a single-ended set of FDs is:

Step 1: It is provable that the optimal cover of a single-ended set of FDs can be found using the mini cover of the single-ended set of FDs.

Step 2: The single-ended set of FDs is represented by the corresponding FD-graph, and then the transitive reduction [10] of the FD-graph is found. Actually, the transitive reduction of the FD-graph can be transformed into corresponding mini cover and further optimal cover.

The *organization* of the article is: In section 2, rules and derivation sequence are introduced. Section 3 proves that the optimal cover of a single-ended cover can be found, using the notion of mini cover and derivation sequence. Section 4 simplifies a graphical representation for a set of FDs to find a mini cover and further an optimal cover, using the notion of transitive reduction. Section 5 shows an algorithm to find an optimal cover for a single-ended set of FDs and its computational complexity, which is the optimal cover of a single-ended set of FDs can be found in polynomial time.

II. DERIVATION SEQUENCE

As usual, capital letters from the beginning of the alphabet, e.g. *A*, *B* and *C*, represent single attributes, capital letters from the end of the alphabet, e.g. *X*, *Y* and *Z*, stand for sets of attributes, and concatenation, e.g. *AB* and *XY*, is used for union.

Given a set of FDs *F*, *F* implies $X\rightarrow Y$ if $X\rightarrow Y\in F^+$. If *X*, *Y*, *Z* and *W* are sets of attributes, then inference rules [11] are:

F1. $X\rightarrow X$.
F2. $X\rightarrow Y$ implies $XZ\rightarrow Y$.
F3. $X\rightarrow Y$ and $X\rightarrow Z$ imply $X\rightarrow YZ$.
F4. $X\rightarrow YZ$ implies $X\rightarrow Y$.
F5. $X\rightarrow Y$ and $Y\rightarrow Z$ imply $X\rightarrow Z$.
F6. $X\rightarrow Y$ and $YZ\rightarrow W$ imply $XZ\rightarrow W$.

Definition 1 ([11]). Let *F* be a set of FDs. A sequence *P* of FDs is a derivation sequence on *F* if every FD in *P* either is a member of *F* or follows from previous FDs in *P* by an application of one of the inference rules F1 to F6. *P* is a derivation sequence for $X\rightarrow Y$ if $X\rightarrow Y$ is one of the FDs in *P*.

Example 1. Let $F=\{AB\rightarrow C,\ BC\rightarrow D,\ E\rightarrow AB\}$. The following sequence *P* is a derivation sequence for $CE\rightarrow D$.

1. $AB\rightarrow C$ (given)
2. $BC\rightarrow D$ (given)

3. $AB \to D$ (F6 from 1 and 2)
4. $E \to AB$ (given)
5. $E \to D$ (F5 from 4 and 3)
6. $CE \to D$ (F2 from 5)

Proposition 1 ([8]). Let G be a set of FDs, and let $X \to Y$ be an FD in G. If $V \to W$ is in G^+ and $X \to Y$ is used for some derivation sequence of $V \to W$ from G, then $V \to X$ is in G^+.

If X, Y, Z and W are sets of attributes, and C is a single attribute, then B-axioms [11] are:

B1. $X \to X$.
B2. $X \to YZ$ and $Z \to CW$ imply $X \to YZC$.
B3. $X \to YZ$ implies $X \to Y$.

Definition 2 ([11]). Consider derivation sequence for $X \to Y$ on a set F of FDs using the B-axioms that satisfy the following constraints: (1) The first FD is $X \to X$; (2) The last FD is $X \to Y$; (3) Every FD other than the first and last is either an FD in F or an FD of the form $X \to Z$ that was derived using axiom B2. Such a derivation sequence is called a RAP-derivation sequence, for the order in which the B-axioms are used.

Proposition 2 ([11]). Let F be a set of FDs. If there is a derivation sequence on F for $X \to Y$, then there is a RAP-derivation sequence on F for $X \to Y$.

III. Finding an Optimal Cover for a Single-Ended Set of FDs Using Mini Cover

Given a set of FDs F, we declare notations [12]:

- RS_F: the group (*not set*) of right sides of all FDs in F.
- $||RS_F||$: the number of attributes in RS_F (with repetiti-ons counted).
- LS_F: the group (*not set*) of left sides of all FDs in F.
- $||LS_F||$: the number of attributes in LS_F (with repetiti-ons counted).

Definition 3. A set of FDs F is *single-ended* if the left side of every FD in F has a single attribute.

Definition 4 ([12]). A set of FDs F is *mini* if the right side of every FD has a single attribute in F, where F has the fewest FDs and, within that constraint, the fewest attributes.

Definition 5 ([3]). A set of FDs F is *optimal* if there is no set G with fewer attributes than F such that $F^+=G^+$.

Lemma 1. If F is the mini cover of a single-ended set of FDs G, then the left side of every FD has a single attribute in F.

Proof. Suppose F has an FD $A_1A_2\ldots A_n \to X$ ($n \geq 2$). That is, there is an FD with the left side is not single attribute in F. Since G is a single-ended set of FDs and $F^+=G^+$, there is a derivation sequence for $A_1A_2\ldots A_n \to X$ on G. By Proposition 2, there is a RAP-derivation sequence for $A_1A_2\ldots A_n \to X$ on G. Suppose, in G, there is no FD with the left side is A_k ($1 \leq k \leq n$), then, by Definition 2, we cannot construct a RAP-derivation sequence for $A_1A_2\ldots A_n \to X$ on G. This is a contradiction. Hence, in G, there is an FD with the left side is A_k.

Since $F^+=G^+$, $A_k \to X$ is in F^+. By Definition 4, $A_k \to X$ cannot be derived from $F-\{A_1A_2\ldots A_n \to X\}$, since otherwise in F $A_1A_2\ldots A_{k-1}A_{k+1}\ldots A_n$ would not appear in $A_1A_2\ldots A_n \to X$. Hence, there is a derivation sequence of $A_k \to X$ using $A_1A_2\ldots A_n \to X$. By Proposition 1, $A_k \to A_1A_2\ldots A_n$ must be in F^+, and thus $A_k \to A_1A_2\ldots A_{k-1}A_{k+1}\ldots A_n$ is in F^+. Thus, $A_1A_2\ldots A_n \to X$ can't be in F, since $A_1A_2\ldots A_{k-1}A_{k+1}\ldots A_n$ is superfluous.

Lemma 1 shows that the mini cover of a single-ended set of FDs is also a single-ended set of FDs, and implies the following Corollary 1 (the proof of Corollary 1 is the same as Lemma 1).

Corollary 1. If F is the optimal cover of a single-ended set of FDs G, then the left side of every FD has a single attribute in F.

Theorem 1. If F is the mini cover of a single-ended set of FDs G and a set of FDs H obtained by combining in F all FDs which have identical left sides into an FD using the rule F3, then H is the optimal cover of G.

Proof. By Lemma 1, the left side of every FD in H is unique and has a single attribute. Suppose T is the optimal cover of G, then $T^+=H^+$ and the left side of every FD in T is unique. By Corollary 1, every FD in T has a single attribute on the left side.

Firstly, we prove that the left side of every FD in T must exist in LS_H. Suppose in T there is an FD $A \to X$ with A does not exists in LS_H. Since $T^+=H^+$, there is a derivation sequence for $A \to X$ on H. By Proposition 2, there must have a RAP-derivation sequence for $A \to X$ on H. However, since A does not exists in LS_H and the left side of every FD in H is not only unique but also a single attribute, by Definition 2, we cannot construct a RAP-derivation sequence for $A \to X$ on H. This is a contradiction.

Secondly, we prove that the left side of every FD in H must exist in LS_T. Suppose in H there is an FD $B \to Y$ with B does not exists in LS_T. Since $T^+=H^+$, there is a derivation sequence for $B \to Y$ on T. By Proposition 2, there must have a RAP-derivation sequence for $B \to Y$ on T. However, since B does not exists in LS_T and the left side of every FD in T is not only unique but also a single attribute, by Definition 2, we cannot construct a RAP-derivation sequence for $B \to Y$ on T. This is a contradiction.

From the above, the LS_H and LS_T are identical.

Finally, we prove that $||RS_H||=||RS_T||$. Suppose $||RS_H||\neq||RS_T||$. Since LS_H and LS_T are identical and T is an optimal cover, $||RS_T||<||RS_H||$. We can form $T'$by splitting the right sides of FDs in T into single attributes using rule F4. We can also form F by splitting the right sides of FDs in H into single attributes using rule F4. Since $||RS_T||<||RS_H||$, T'has fewer FDs than F. That is, F is not mini cover. This is a contradiction.

Since LS_H and LS_T are identical and $||RS_H||=||RS_T||$, H is the optimal cover of G.

Corollary 2. Let F be a set of FDs and let G be the mini cover for F. If the left side of every FD in G has a single attribute, then we can obtain an optimal cover for F by combining in G all FDs which have identical left sides into an FD using rule F3.

IV. Transitive Reduction of FD-Graph

Ausiello et al. [4] propose a graphical representation for a set of FDs. We can simplify the representation for a single-ended set of FDs.

Definition 6. Let S be a single-ended set of FDs, then the FD-graph G_S=<V, E> associated with S is the graph such that

(i) for every FD $A \to Y$ with the left side A in S, there is a node in V labeled A;

(ii) for every FD $A \to Y$ in S where $Y = A_1A_2 \ldots A_k$, there are arcs from the node labeled A to the nodes labeled A_1, A_2, …, A_k;

Definition 7. Let S be a set of FDs. The closure of an FD-graph G_S=<V, E> associated with S is the graph G_S^+=<V, E^+>, labeled on the nodes and on the arcs, where the set V is the same as in G_S, while the set $E^+=\{(i, j)|i, j \in V$ and there exists a directed path $<i, j>\}$.

Proposition 3. Let S be a set of FDs and let G_S=<V, E> be the FD-graph associated with S. $A \to B$ is in S^+ if and only if there exists a directed path from A to B in G_S^+.

Definition 7 and Proposition 3 are stemmed from [4]. Proposition 3 shows that all directed graph theory can be completely used for the FD-graph of a single-ended set of FDs.

Definition 8 ([10]). A graph G_t is a transitive reduction of the directed graph G whenever the following two conditions are satisfied:

(i) there is a directed path from vertex u to vertex v in G_t if and only if there is a directed path from u to v in G, and

(ii) there is no graph with fewer arcs than G_t satisfying condition (i).

Theorem 2. Let S be a single-ended set of FDs and let G be the FD-graph of S. If M is the mini cover of S, then the FD-graph of M is the transitive reduction of G.

Proof. Suppose the FD-graph of M is not the transitive reduction of G, then the FD-graph of M has superfluous arcs. That is, by Proposition 3, M has redundant FDs. This is a contradiction.

Theorem 2 shows that if G_t is the transitive reduction of G, by Definition 6, we can transforms G_t into the mini cover of S. By Theorem 1, we also can transform G_t into the optimal cover of S.

ALGORITHM 1: OPTIMIZE(S)

Input: A single-ended set of FDs S.

Output: An optimal cover O of S.

Step 1. Find the transitive reduction G_t of the FD-graph of S;

Step 2. Transform G_t into the optimal cover O of S;

V. Computational Complexity

Definition 6, Theorem 1 and Theorem 2 show that the optimal cover of a single-ended set of FDs S can be found using ALGORITHM 1.

Let n denote the number of vertexes of the FD-graph of S and e denote the number of arcs of the FD-graph of S, then, in ALGORITHM 1, Step 1 can be finished in time $O(n^a)$ ($a \geq 2$) [10]. It is apparent that Step 2 can be finished in time $O(n+e)$. Thus, ALGORITHM 1 can be finished in time $O(n^a+n+e)$ ($a \geq 2$).

VI. Conclusion

We have proved that the optimal cover of a single-ended set of FDs can be found, using the notion of mini cover. We have simplified a graphical representation for FDs, and transformed the transitive reduction of the FD-graph of a single-ended set of FDs into corresponding mini cover and further optimal cover. We have shown here that the optimal cover of a single-ended set of FDs can be found in polynomial time.

Acknowledgment

This work is supported by the constructing program of the key discipline in Huaihua University.

References

[1] E. F. Codd, “A relational model of data for large shared data banks”, Communication of the ACM, vol. 13, no. 6, pp. 377–387, 1970.
[2] R. C. Shock, “Computing the minimum cover of functional dependencies”, Information Processing Letters vol. 22, no. 3, pp. 157–159, 1986.
[3] D. Maier, “Minimum covers in the relational database model”, J. ACM, vol. 27, no. 4, pp. 664–674, 1980.
[4] A. D. Giorgio Ausiello, D. Sacca, “Graph algorithm for functional dependency manipulation”, J. ACM, vol. 30, no. 4, pp. 752–766, 1983.
[5] V. Cotelea, “Problem decomposition method to compute an optimal cover for a set of functional dependencies”, Database Systems Journal, vol. 2, no. 4, pp. 17–30, 2011.
[6] X. Peng, Z. Xiao, “Comments on ’problem decomposition method to compute an optimal cover for a set of functional dependencies’”, Database Systems Journal, vol. 4, no. 3, pp. 50–51, 2013.
[7] H. Mannila, K. J. Raiha, “On the relationship of minimum and optimal covers for s set of functional dependencies”, Acta Informatica vol. 20, no. 2, pp. 143–158, 1983.
[8] P. A. Bernstein, “Synthesizing third normal form relations from functional dependencies”, ACM Transactions on Database Systems, vol. 1, no. 4, pp. 277–298, 1976.
[9] C. Beeri, P. A. Bernstein, “Computational problems related to the design of normal form relational schemas”, ACM Transactions on Database Systems, vol. 4, no. 1, pp. 30–59, 1979.
[10] M. R. G. A. V. Aho, J. D. Ullman, “The transitive reduction of a direct graph”, SIAM J. Comput. vol. 1, no. 2, pp. 131–137, 1972.
[11] D. Maier, The Theory of Relational Databases, Computer Science Press, Rockville, Md., 1983, pp. 44-55.
[12] X. Peng, Z. Xiao, “Optimal covers in the relational database model”, submitted for publication.

Author’s Profile

Xiaoning Peng
was born in October, 1962. He is professor at CS Department, the University of Huaihua, Hunan, where he directs the Key Discipline of Computer Applications Technology and Image Processing and Graphics Applications Laboratory. His present research interests include computer vision, pattern recognition, database theory and digital image processing and their applications. He is a member of the CCF and a member of the Hunan Computer Federation, and Hunan CEF. Email: hhpxn@163.com

Zhijun Xiao
received the bachelor’s degree in computer science and technology from Huaihua University in June 2013. He is a member of the Key Discipline of Computer Applications Technology and Image Processing and Graphics Applications Laboratory. His research interests include the P versus NP problem and database management. Email: pxnxzj@outlook.com