Bounds on the number of independent dependencies

Dezső Miklós

Rényi Institute, Budapest

Algebra Across the Borders II
Bolyai Institute, Szeged

Question

How big the number of independent functional dependencies of an n-ary relation schema can be?

Question

How big the number of independent functional dependencies of an n-ary relation schema can be?

Let R be a relational database model, and X denote the set of attributes. We say that (for two subsets of attributes A and B) $A \rightarrow B$, that is, B functionally depends on A, if in the database R the values of the attributes in A uniquely determine the values of the attributes in B.

Definition

A set of functional dependencies $\mathcal{D}$ (in a database R) are called linearly independent if none of them can be (logically) derived from the rest, i.e., the closure of $\mathcal{D} \setminus \{A \to B\}$ does not contain $\{A \to B\}$ for any $\{A \to B\} \in \mathcal{D}$.

Definition

A set of functional dependencies $\mathcal{D}$ *(in a database* R*) are called linearly independent if none of them can be (logically) derived from the rest, i.e., the closure of* $\mathcal{D} \setminus \{A \to B\}$ *does not contain* $\{A \to B\}$ *for any* $\{A \to B\} \in \mathcal{D}$*.*

Clearly (Armstrong axioms)

- if $B \subset A$, then $A \to B$, *reflexivity rule*.

Definition

A set of functional dependencies $\mathcal{D}$ (in a database R) are called linearly independent if none of them can be (logically) derived from the rest, i.e., the closure of $\mathcal{D} \setminus \{A \to B\}$ does not contain $\{A \to B\}$ for any $\{A \to B\} \in \mathcal{D}$.

Clearly (Armstrong axioms)

- if $B \subset A$, then $A \to B$, *reflexivity rule.*
- if $A \to B$, $C \subset X$, then $A \cup C \to B \cup C$, *augmentation rule.*

Definition

A set of functional dependencies $\mathcal{D}$ (in a database R) are called linearly independent if none of them can be (logically) derived from the rest, i.e., the closure of $\mathcal{D} \setminus \{A \to B\}$ does not contain $\{A \to B\}$ for any $\{A \to B\} \in \mathcal{D}$.

Clearly (Armstrong axioms)

- if $B \subset A$, then $A \to B$, *reflexivity rule.*
- if $A \to B$, $C \subset X$, then $A \cup C \to B \cup C$, *augmentation rule.*
- if $A \to B$ and $B \to C$, then $A \to C$, *transitivity rule.*

Lower bound (construction)

Take the maximum number of incomparable subsets of attributes (by Sperner's theorem $\binom{n}{\lfloor n/2 \rfloor}$ for a set of n attributes) and let the whole set of attributes depend on all of them, which therefore give a set of dependencies of cardinality $\binom{n}{\lfloor n/2 \rfloor}$.

Lemma

Assign to a functional dependency $A \to B$ *the set of* $2^{n-|A|} - 2^{n-|A \cup B|}$ *binary vectors* $\mathbf{a} = (a_1, \ldots, a_n)$ *of the form:*

$$a_i = \begin{cases} 1\,, & \text{if } \; i \in A \\ 0 \text{ or } 1\,, & \text{if } \; i \in (B \setminus A) \; \text{ but not all entries } = 1 \\ 0 \text{ or } 1\,, & \text{otherwise.} \end{cases}$$

Then, a set of functional dependencies implies another functional dependency if and only if the binary vectors of the implied functional dependency are all contained in the union of the sets of binary vectors of the given functional dependencies.

Upper bound

Theorem

For every n an upper bound for the maximum number of independent functional dependencies on an n-element set of attributes is $2^n - 1$.

Upper bound

Theorem

For every n an upper bound for the maximum number of independent functional dependencies on an n-element set of attributes is $2^n - 1$.

Proof (sketch). Let $\mathcal{F}$ be a set of independent functional dependencies. Replace each dependency $A \to B$ by $A \to A \cup B$, obtaining $\mathcal{F}'$, $\mathcal{F}'$ is independent as well.
$|\mathcal{F}| = |\mathcal{F}'|$, since the images of dependencies in $\mathcal{F}$ will be different in $\mathcal{F}'$.
Assume that $A \to A \cup B = A \to A \cup C$ for $A \to B$ and $A \to C$ in $\mathcal{F}$. Then $A \cup B = A \cup C$, $C \subset A \cup B$ and therefore $A \to B$ implies $A \to A \cup B$ implies $A \to C$, a contradiction.

Consider only set of independent dependencies where for all $(A \to B) \in \mathcal{F}$ we have $A \subset B$. Take the graph G: let the vertices of the graph be all the 2^n subsets of the n attributes and for $A, \ B \subset X$ the edge (A, B) will be present in the G iff $A \to B$ or $B \to A$ is in $\mathcal{F}$. This graph may not contain a cycle, therefore it is a forest, that is it has at most $2^n - 1$ edges, or dependencies. □

Better lower bound

Anything better than $\binom{n}{\lfloor n/2 \rfloor}$ would be good.

Better lower bound

Anything better than $\binom{n}{\lfloor n/2 \rfloor}$ would be good.
Is it possible to add some (at least one more) (independent) dependencies to the set of (independent) dependencies $A \to X$ (for all $|A| = \lfloor n/2 \rfloor$)?

Lemma

(Graham, Sloane) If n is an odd prime number, one can find $\frac{1}{n^2}\binom{n}{\frac{n+3}{2}}$ *subsets* $V_1, V_2, \ldots$ *of size* $\frac{n+3}{2}$ *in the set* $[n] = \{1, 2, \ldots n\}$ *in such a way that* $|V_i \cap V_j| < \frac{n-1}{2}$ *holds.*

Lemma

(Graham, Sloane) If n is an odd prime number, one can find $\frac{1}{n^2}\binom{n}{\frac{n+3}{2}}$ *subsets* $V_1, V_2, \ldots$ *of size* $\frac{n+3}{2}$ *in the set* $[n] = \{1, 2, \ldots n\}$ *in such a way that* $|V_i \cap V_j| < \frac{n-1}{2}$ *holds.*

Proof. Consider the subsets $\{x_1, x_2, \ldots, x_{\frac{n+3}{2}}\}$ of integers satisfying $1 \leq x_i \neq x_j \leq n$ for $i \neq j$ and the equations

$$x_1 + x_2 + \cdots + x_{\frac{n+3}{2}} \equiv a \pmod{n}, \qquad (1)$$

$$x_1 x_2 \cdot \cdots \cdot x_{\frac{n+3}{2}} \equiv b \pmod{n} \qquad (2)$$

for some fixed integers a and b.

The *shadow* of a family $\mathcal{A} \subset \binom{[n]}{\frac{n+1}{2}}$ is

$$\sigma(\mathcal{A}) = \{B : \ |B| = \frac{n-1}{2}, \ \ \exists A \in \mathcal{A} : B \subset A\}.$$

The *shadow* of a family $\mathcal{A} \subset \binom{[n]}{\frac{n+1}{2}}$ is

$$\sigma(\mathcal{A}) = \{B : \ |B| = \frac{n-1}{2}, \ \ \exists A \in \mathcal{A} : B \subset A\}.$$

A pair $\{U_1, U_2\}$ $U_i \in \binom{[n]}{\frac{n+1}{2}}$ is *good* if $|U_1 \cap U_2| = \frac{n-1}{2}$.

The *shadow* of a family $\mathcal{A} \subset \binom{[n]}{\frac{n+1}{2}}$ is

$$\sigma(\mathcal{A}) = \{B : \ |B| = \frac{n-1}{2}, \ \ \exists A \in \mathcal{A} : B \subset A\}.$$

A pair $\{U_1, U_2\}$ $U_i \in \binom{[n]}{\frac{n+1}{2}}$ is *good* if $|U_1 \cap U_2| = \frac{n-1}{2}$.

The family $\mathcal{P} \subset \binom{[n]}{\frac{n+1}{2}}$ is a *chain* if $\mathcal{P} = \mathcal{P}_1 \cup \ldots \cup \mathcal{P}_l$ where the $\mathcal{P}_i$'s are good pairs and $\sigma(\mathcal{P}_i) \cap \sigma(\mathcal{P}_j) = \emptyset$ for $i \neq j$. The weight $w(\mathcal{P})$ of this chain is l.

Lemma

There is a chain $\mathcal{P} \subset \binom{[n]}{\frac{n+1}{2}}$ *chain with weight at least*

$$|\mathcal{P}| = \frac{1}{n^2}\binom{n}{\frac{n+3}{2}}.$$

Lemma

There is a chain $\mathcal{P} \subset \binom{[n]}{\frac{n+1}{2}}$ *chain with weight at least*

$$|\mathcal{P}| = \frac{1}{n^2}\binom{n}{\frac{n+3}{2}}.$$

Proof. Start with the family $\mathcal{V} = \{V_1, V_2, \ldots\}$ defined before. In each V_i choose two different $\frac{n+1}{2}$-element subsets U_{i1} and U_{i2}. □

Lemma

If $\mathcal{P} \subset \binom{[n]}{\frac{n+1}{2}}$ is a chain then the following set of functional dependencies is independent:

$$A \to B, \quad \textit{where} \quad |A| = \frac{n-1}{2}, \quad |B| = \frac{n+1}{2}, \quad A \subset B \in \mathcal{P}, \qquad (3)$$

$$A \to A', \quad \textit{where} \quad |A| = \frac{n-1}{2}, \quad A \notin \sigma(\mathcal{P})$$

$$\textit{and} \quad A' \quad \textit{is arbitrarily chosen so that} \quad A \subset A', \quad |A'| = \frac{n+1}{2}. \quad (4)$$

(In (3) we have all dependencies $A \to B$ given, while in (4) for the remaining A exactly one A' satisfying the conditions.)

This gives the number of dependencies:

$$\binom{n}{\lfloor \frac{n}{2} \rfloor} + \frac{1}{n^2}\binom{n}{\frac{n+3}{2}} = \left(1 + \frac{1}{n^2} + o\left(\frac{1}{n^2}\right)\right)\binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

□

Lemma

Assign to a functional dependency $A \to B$ the set of $2^{n-|A|} - 2^{n-|A \cup B|}$ binary vectors $\mathbf{a} = (a_1, ..., a_n)$ *of the form:*

$$a_i = \begin{cases} 1\,, & \text{if } \; i \in A \\ 0 \text{ or } 1\,, & \text{if } \; i \in (B \setminus A) \;\text{ but not all entries } \; = 1 \\ 0 \text{ or } 1\,, & \text{otherwise.} \end{cases}$$

Then, a set of functional dependencies implies another functional dependency if and only if the binary vectors of the implied functional dependency are all contained in the union of the sets of binary vectors of the given functional dependencies.

Corollary

Assign to a functional dependency $A \rightarrow B$ *the set of all* $\ell = 2^{n-|A|} - 2^{n-|A \cup B|}$ *subsets* $\mathbf{A} = (A_1, \ldots, A_\ell)$ *of the set of attributes satisfying*

$$A \subset A_i \text{ and } A \cup B \not\subset A_i$$

Then, a set of given functional dependencies implies another functional dependency if and only if the above sets **A** *assigned to the implied functional dependency are all assigned to at least one of the given functional dependencies.*

Corollary

Assign to a functional dependency $A \to B$ *the set of all* $\ell = 2^{n-|A|} - 2^{n-|A \cup B|}$ *subsets* $\mathbf{A} = (A_1, \ldots, A_\ell)$ *of the set of attributes satisfying*

$$A \subset A_i \text{ and } A \cup B \not\subset A_i$$

Then, a set of given functional dependencies implies another functional dependency if and only if the above sets **A** *assigned to the implied functional dependency are all assigned to at least one of the given functional dependencies.*

Proof. An immediate consequence of the previous proposition with the vectors a_i from the proposition being characteristic vectors of the sets A_i.

Lemma

We can assign to a set of independent functional dependencies $\{A_i \to B_i\}_{1\leq i\leq m}$ *a set of pairs* $\{F_i, b_i\}_{1\leq i\leq m}$ *where* F_i*'s are subsets of the attributes and* $b_i = b(F_i)$*'s are elements of* F_i*'s and they have the following* weak Sperner property*:*

$$\text{if } F_j \subset F_i \text{ then } b_i \notin F_j$$

Lemma

We can assign to a set of independent functional dependencies $\{A_i \to B_i\}_{1 \leq i \leq m}$ a set of pairs $\{F_i, b_i\}_{1 \leq i \leq m}$ where F_i's are subsets of the attributes and $b_i = b(F_i)$'s are elements of F_i's and they have the following weak Sperner property*:*

$$\textit{if } F_j \subset F_i \textit{ then } b_i \notin F_j$$

Proof. Assign to each of the m independent functional dependencies the sets defined in the above corollary. Since the dependencies are independent, for every $\mathbf{A}_i$ there will be a C_i subset of dependencies from it not contained in any other $\mathbf{A}_j$. That is, these C_i's will have the following properties:

- $A_i \subset C_i$ and $B_i \not\subset C_i$
- if $A_j \subset C_i$ then $B_j \subset C_i$ as well

For each i pick an element (attribute) b_i from B_i which is not in C_i (it must exist, since $B_i \not\subset C_i$, otherwise $A_i \cup B_i$ would be a subset of C_i as well, a contradiction to the choice of C_i's). With this choice we have a set of pairs $\{C_i, b_i\}_{1\leq i\leq m}$ where C_i's are subsets of the attributes and b_i's are *not* elements of C_i's with the following property: if $C_j \subset C_i$ then $b_j \in C_i$ (since $A_j \subset C_j$ and $C_j \subset C_i$ implies that $A_j \subset C_i$, which implies that $B_j \subset C_i$, that is, every element of B_j, in particular b_j are elements of C_i).
Finally, replace in the all of above pairs the C_i's by $F_i = \overline{C_i}$, yielding pairs $\{F_i, b_i\}_{1\leq i\leq m}$ with the required properties. Indeed, since $b_i \notin C_i$, $b_i \in \overline{C_i} = F_i$ and if $\overline{C_i} = F_i \subset F_j = \overline{C_j}$ then $C_j \subset C_i$, therefore $b_j \in C_i$, that is $b_j \notin F_i = \overline{C_i}$.

Remark

(Sali, Živković) Let over $\{1, 2, \ldots n\}$ *for every* $k = 1, 2, \ldots, n-1$ *a maximum Sperner family on the set* $\{1, 2, \ldots k\}$ *be taken and add to each of these sets the element* $k+1$. *Let this extra element* $k+1$ *be the special element of each of the sets of this level. Therefore, there will be no set from a level containing another set from the same level, neither will a set from a lower* (l) *level contain another one from a higher* (k) *level. If a higher* (k) *level set contains a lower level one, it's designated element* $k+1$ *will not be in the contained set.*

Remark

(Sali, Živković) Let over $\{1, 2, \dots n\}$ *for every* $k = 1, 2, \dots, n-1$ *a maximum Sperner family on the set* $\{1, 2, \dots k\}$ *be taken and add to each of these sets the element* $k+1$*. Let this extra element* $k+1$ *be the special element of each of the sets of this level. Therefore, there will be no set from a level containing another set from the same level, neither will a set from a lower (l) level contain another one from a higher (k) level. If a higher (k) level set contains a lower level one, it's designated element* $k+1$ *will not be in the contained set.*

This construction gives

$$\sum_{k=1}^{n-1} \binom{k}{\lfloor \frac{k}{2} \rfloor} = \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + \frac{c_n}{n} + o\left(\frac{1}{n}\right)\right)$$

sets.

Lemma

The following statements are equivalent.
(i) One can find an element $b(F) \in F$ *for every member* $F \in \mathcal{F}$ *such that* $F, G \in \mathcal{F}, F \subset G, F \neq G$ *implies* $b(G) \notin F$.
(ii) For every positive integer r *and every choice of distinct members* $G, F_1, \ldots, F_r$ *of* $\mathcal{F}$ *the equality* $\cup_{i=1}^{r} F_i = G$ *cannot hold.*

Theorem

Let $\mathcal{F}$ be a family of distinct subsets of $[n]$ such that $\cup_{i=1}^{r} F_i = G$ cannot hold for any positive integer r and distinct members $G, F_1, \ldots, F_r$ of $\mathcal{F}$. Then

$$|\mathcal{F}| \leq \begin{cases} \binom{n}{\frac{n}{2}} + (2^n/(n+1)) & \text{for even } n\text{'s} \\ \left(1 + \frac{1}{n+1}\right)\binom{n}{\lfloor \frac{n}{2} \rfloor} + (2^n/(n+1)) & \text{for odd } n\text{'s} \end{cases}$$

Theorem

Let $\mathcal{F}$ be a family of distinct subsets of $[n]$ such that $\cup_{i=1}^{r} F_i = G$ cannot hold for any positive integer r and distinct members $G, F_1, \ldots, F_r$ of $\mathcal{F}$. Then

$$|\mathcal{F}| \leq \begin{cases} \binom{n}{\frac{n}{2}} + \left(2^n/(n+1)\right) & \text{for even } n\text{'s} \\ \left(1 + \frac{1}{n+1}\right)\binom{n}{\lfloor \frac{n}{2} \rfloor} + \left(2^n/(n+1)\right) & \text{for odd } n\text{'s} \end{cases}$$

Remark

Kleitman much earlier proved: if $\mathcal{F}$ is a family of distinct subsets of $[n]$ such that $F \cup H = G$ cannot hold for any 3 distinct members, then

$$|\mathcal{F}| \leq \begin{cases} \binom{n}{\frac{n}{2}} + \left(2^n/(n+1)\right) + O\binom{n}{n/4} & \text{for even } n\text{'s} \\ \left(1 + \frac{1}{n+1}\right)\binom{n}{\lfloor \frac{n}{2} \rfloor} + \left(2^n/(n+1)\right) + O\binom{n}{n/4} & \text{for odd } n\text{'s} \end{cases}.$$

Summary

The maximum number of independent functional dependencies over a set of n attributes is between

$$(1 + \frac{1}{n^2})\binom{n}{\lfloor \frac{n}{2} \rfloor}$$

and

$$(1 + \frac{c}{\sqrt{n}})\binom{n}{\lfloor \frac{n}{2} \rfloor}$$

with $c = \sqrt{\frac{\pi}{2}}$.