

APPROXIMATE FUNCTIONAL DEPENDENCIES—IMPLICATION PROBLEM REVISITED

Nicolas Fröhlich¹, Matilda Häggblom², Åsa Hirvonen², and Minna Hirvonen¹

¹Leibniz Universität Hannover, Germany and ²University of Helsinki, Finland

Motivation

Data in the real world is rarely perfect. To reason meaningfully about functional dependencies in flawed data, we use approximate dependence atoms.

Dependence atoms in team semantics correspond to functional dependencies, an important and well-studied class of database constraints.

We study the error-tolerant approximate dependence atom, originally defined by Kivinen and Mannila (1995). The approximation allows for some part of the team to contradict the dependence relation.

Väänänen (2017) proposed an axiomatisation for their implication problem, that is, a set of rules that can determine whether a set of atoms implies a given atom.

Unfortunately, we show that at least one rule is missing from the system. Still, the original system is complete for unary atoms.

We also consider the complexity of model checking for approximate dependence: it is NP-complete for disjunctions of two atoms and LOGSPACE-hard for individual atoms.

Approximate dependence

A team $X = \{s_1, \dots, s_n\}$ is a finite set of assignments $s_i: \mathcal{V} \longrightarrow \mathcal{M}$ that map variables to values. For a variable sequence $x = \langle x_1, \dots, x_n \rangle$, we write $s(x)$ for $\langle s(x_1), \dots, s(x_n) \rangle$.

Let p be a real number, $0 \leq p \leq 1$. For finite teams, $X \models =_p(x, y)$ if there is a subteam $Y \subseteq X$, $|Y| \leq p \cdot |X|$, such that for all $s, s' \in X \setminus Y$, $s(x) = s'(x)$ implies $s(y) = s'(y)$.

For example, consider the approximate dependence atom $=_{\frac{1}{2}}(\mathbf{R}, \mathbf{S})$ expressing “most reviewers are assigned at most one submission.”

The team on the right satisfies this constraint, since only one assignment (s_4 or s_7) needs to be removed so that $\mathbf{R}$ determines $\mathbf{S}$.

On the other hand, $=_{\frac{1}{2}}(\mathbf{S}, \mathbf{R})$ does not hold in this team, because there are two pairs of conflicting assignments: s_1, s_3 and s_4, s_6 .

	Reviewer	Submission
s_1	Franz Baader	#01
s_2	Salem Benferhat	#02
s_3	Marco Calautti	#01
s_4	Anselm Haak	#03
s_5	Benny Kimelfeld	#04
s_6	Yasir Mahmood	#03
s_7	Anselm Haak	#05

Main results

Definition

A rational approximate functional dependence atom $=_p(x, y)$ is such that $p \in \mathbb{Q} \cap [0, 1]$. We say that $=_p(x, y)$ is unary if $|x| = |y| = 1$, and constant if $|x| = 0$ and $|y| = 1$.

System A from Väänänen (2017)

- (A1) $=_0(xy, x)$.
- (A2) $=_1(x, y)$.
- (A3) If $=_p(x, yv)$, then $=_p(xu, y)$.
- (A4) If $=_p(x, y)$, then $=_p(xu, yu)$.
- (A5) If $=_p(xu, yv)$, then $=_p(ux, yv)$ and $=_p(xu, vy)$.
- (A6) For $r = \min\{p + q, 1\}$, if $=_p(x, y)$ and $=_q(y, v)$, then $=_r(x, v)$.
- (A7) For $p \leq q \leq 1$, if $=_p(x, y)$, then $=_q(x, y)$.

First, we modify the system to $\mathbf{A}^*$ consisting of rules $A1, A2, A_{34}, A5, A6$ and $A7$, with

$$(A_{34}) \quad \text{If } =_p(x, y), \text{ then } =_p(x, xy).$$

System $\mathbf{A}^*$ is complete for usual dependencies sensitive to repetition of variables, but not for their approximate variant.

Missing rule

Let $\Sigma = \{=_0(x), =_{\frac{1}{8}}(x, v_1v_2), =_{\frac{2}{8}}(x, v_1v_3), =_{\frac{2}{8}}(v_1, y), =_0(v_2v_3, y)\}$ be a set of approximate atoms. If $X \models \Sigma$ for any team X , then $X \models =_{\frac{5}{16}}(x, y)$.

However, $\mathbf{A}^*$ does not allow us to derive $=_{\frac{5}{16}}(x, y)$ from Σ .

The missing rule illustrates a dependence phenomenon that only turns up in the approximate notion. When no errors are allowed, a tuple yz depends on x exactly when both y and z depend on x , but in the approximate version, the distribution of errors between variables introduces interplay between the variables that is hard to axiomatise.

Unary completeness theorem

Let $\Sigma \cup \{=_p(x, y)\}$ be a finite set of unary and constancy rational approximate dependence atoms. If $\Sigma \models =_p(x, y)$, then $\Sigma \vdash =_p(x, y)$ using the rules from system $\mathbf{A}$ restricted to unary atoms.

Counterexample team

$X:$	x	w	y	z_0
s_0	0	0	0	0
s_1	0	1	1	1
s_2	0	1	2	2
s_3	0	1	2	3
s_4	0	1	2	4

Let $n - 1$ be the least common denominator for the approximations in $\Sigma \cup \{=_p(x, y)\}$. For each approximation q , let q' be such that $q = \frac{q'}{n-1}$. Construct $X = \{s_0, \dots, s_{n-1}\}$ according to the smallest q for which $\Sigma \vdash =_q(x, w)$ by defining $s_i(w) = i$ if $i \leq q'$, and q' otherwise. Additionally, let $s_i(z_0) = i$ for all $i \in \{0, \dots, n - 1\}$.

Complexity results

- The model checking problem for $=_{\frac{3}{13}}(u, v) \vee =_0(x, y)$ is NP-complete.
- The model checking problem for $=_{\frac{1}{2}}(x, y)$ is LOGSPACE-hard.

For disjunctions of two atoms, our first result demonstrates an increase – from NL-complete to NP-complete – in complexity, compared to the non-approximate setting. Indeed, by our second result, even single approximate atoms permit LOGSPACE-hard model checking, while classical dependence atoms are first-order rewritable.

Future work

- ? Is the implication problem for approximate dependence atoms finitely axiomatisable? If not, can the axiomatisation still be presented with a finitely describable rule schema?
- ? The structure consisting of positive rationals together with addition (addition and multiplication) is a monoid (semiring). We could consider dependence atoms where the approximations come from a more general structure such as a monoid or a semiring.
- ? Consider other definitions to capture the approximations, such as absolute value approximations and local variants where the error is in relation to the size of the team restricted to the variables in the atom considered.

Selected references

- Durand A., Hannula M., Kontinen J., Meier A., Virtema J. (2018), Approximation and dependence via multiteam semantics, Ann. Math. Artif. Intell. 83 297–320.
- Hannula, M. (2024), Conditional Independence on Semiring Relations, in: G. Cormode, M. Sheklyan (Eds.), LIPIcs vol 290, 20:1–20:20. 27th International Conference on Database Theory (ICDT '24).
- Kivinen, J., Mannila, H. (1995), Approximate inference of functional dependencies from relations, Theoretical Computer Science 149 (1995) 129–149. Fourth International Conference on Database Theory (ICDT '92).
- Väänänen, J. (2017), The logic of approximate dependence. In: Baškent, C., Moss, L.S., Ramanujam, R. (eds.) Rohit Parikh on Logic, Language and Society. 227–234.

For more:



Acknowledgments

The authors want to thank Jouko Väänänen and the anonymous reviewers for many helpful suggestions. The authors appreciate the project support by a Finnish-German co-operation Grant: DAAD (German Exchange Service, project id 57710940), Research Council of Finland (project id 359650). The first author appreciates funding by the German Research Foundation (DFG) under the grant ME2479/3-1 and project id 511769688. The fourth author appreciates funding by the Magnus Ehrnrooth Foundation.

MAGNUS EHRNROOTHIN
SÄÄTIÖ

