Finite Model Theory

Martin Otto
University of Wales Swansea

Dagstuhl 10-99

Model Theory – Finite Model Theory

model theory:

relating $\left\{ \begin{array}{l} \text{syntax} \\ \text{languages} \\ \text{logics} \end{array} \right\}$ and $\left\{ \begin{array}{l} \text{semantics} \\ \text{structures} \\ \text{models} \end{array} \right\}$

classically distinguished: first-order logic FO

completeness, compactness

study of FO definability / FO model classes

structure theory, classification theory

$\boxed{\text{(first-order) logic} + \text{algebra}}$

classical domain of discourse:

all, finite and infinite, structures

in contrast, for CS relevant context,
often need to reason with

exclusive reference to finite structures

Finite Model Theory

domain of discourse: finite structures

database instances	finite relational structures
transition systems, temporal, causal, ...	finite (decorated) graphs, finite Kripke structures
strings, words	finite, ordered, mon. structures

FO loses distinguished role:
vary logic to suite purpose of application

domain of finite structures: less uniformity (!)
individual discrete structures vs. smooth classes

$\boxed{\text{logic} + \text{discrete mathematics}}$

combinatorics,
algorithms, complexity

---

classically: view structures up to
elementary equivalence $\equiv$

new issue: choice of partition/granularity and logic

Development of FMT

- what survives of classical FO model theory

  — many issues
  — some but not many methods
  — very few theorems

- what is specifically new in FMT

  — logic + complexity:
  descriptive complexity,
  syntactic criteria for tractability

  — logic + combinatorics:
  0–1 laws, limit laws, asymptotics

  — specific logics:
  fixed-point logics,
  finite variable logics,
  ...

FO over Finite Structures

FO too strong

$$\mathfrak{A} \equiv \mathfrak{B} \quad \Leftrightarrow \quad \mathfrak{A} \simeq \mathfrak{B}$$

each individual isomorphism type
characterised by single FO sentence

compactness fails (trivially)

most of its consequences too

FO too weak

fundamental phenomena inexpressible
e.g., connectivity, evenness

inexpressibility results:
games and probabilistic arguments

Fagin's 0–1 Law:

any sentence $\varphi \in$ FO holds true in almost all sufficiently large structures or in almost none

Games and Expressibility

a tool which survives and thrives

$\equiv$ a limit notion, inherently infinitary, trivial in FMT

analyse through truncations (quantifier rank)
gradations (number of variables)

$\rightarrow$ Ehrenfeucht-Fraïssé style games
with infinite runs as limits $\approx_m \longrightarrow \approx$

players **I** and **II** play on structures $\mathfrak{A}$ and $\mathfrak{B}$

I and **II** take turns to put pebbles as markers
establishing correspondence between elements

$\left.\begin{array}{l}\textbf{I}\text{ challenges}\\ \textbf{II}\text{ supports}\end{array}\right\}$ similarity claim $\mathfrak{A} \approx \mathfrak{B}$

II must maintain locally isomorphic patterns
mimicking in $\mathfrak{B}/\mathfrak{A}$ moves **I** makes in $\mathfrak{A}/\mathfrak{B}$

$\mathfrak{A} \approx_m \mathfrak{B}$ iff **II** has strategy for m rounds
$\mathfrak{A} \approx \mathfrak{B}$ iff **II** has winning strategy

Ehrenfeucht-Fraïssé Correspondences

adapting the protocol to capture various logics/equivalences

- **classical**: add new pair of pebbles in each round

  $\simeq_m \longrightarrow \simeq_p$ (m-)partial isomorphism

  $\equiv_m \longrightarrow \equiv_\infty$ FO/L_∞-equivalence

  Ehrenfeucht,Fraïssé,Karp

- **modal**: move one pair of pebbles along edges

  $\sim_m \longrightarrow \sim$ (m-)bisimulation

  $\equiv_m^{\mathrm{ML}} \longrightarrow \equiv_\infty^{\mathrm{ML}}$ ML/ML_∞-equivalence

  Barwise, van Benthem

- **k-variable**: reposition one of k pairs of pebbles

  $\simeq_m^k \longrightarrow \simeq_p^k$ k-local (m-)partial isomorphism

  $\equiv_m^k \longrightarrow \equiv_\infty^k$ FO^k/L_∞^k-equivalence

  Poizat,Barwise,Immerman

- **guarded**: pebbles covered by a positive literal

  $\sim_m^{\mathrm{G}} \longrightarrow \sim^{\mathrm{G}}$ guarded (m-)bisimulation

  $\equiv_m^{\mathrm{GF}} \longrightarrow \equiv_\infty^{\mathrm{GF}}$ GF/GF_∞-equivalence

  Andreka,van Benthem,Nemeti

Relevant Logics and Equivalences

$\mathbf{FO}^{k}$: the k-variable fragment $\mathrm{FO}^k \subseteq \mathrm{FO}$
FO with variables $x_1, \dots, x_k$ (re-usable)

ML: the modal fragment $\mathrm{ML} \subseteq \mathrm{FO}^2$

$$[\Box\varphi](x) \equiv \forall y(Exy \rightarrow \varphi(y))$$
$$[\Diamond\varphi](x) \equiv \exists y(Exy \wedge \varphi(y))$$

GF: the guarded fragment $\mathrm{GF} \subseteq \mathrm{FO}$

$$\forall \boldsymbol{y}.\alpha\varphi(\boldsymbol{x},\boldsymbol{y}) \equiv \forall \boldsymbol{y}\Big(\alpha(\boldsymbol{x},\boldsymbol{y}) \rightarrow \varphi(\boldsymbol{x},\boldsymbol{y})\Big)$$
$$\exists \boldsymbol{y}.\alpha\varphi(\boldsymbol{x},\boldsymbol{y}) \equiv \exists \boldsymbol{y}\Big(\alpha(\boldsymbol{x},\boldsymbol{y}) \wedge \varphi(\boldsymbol{x},\boldsymbol{y})\Big)$$

guards α: atoms in all variables $\boldsymbol{x}, \boldsymbol{y}$.

$\mathbf{L}^{k}_{\infty}$, $\mathbf{ML}_{\infty}$, $\mathbf{GF}_{\infty}$: closed w.r.t. infinite $\bigwedge \bigvee$

$$\mathrm{ML} \bullet\!\longrightarrow \mathrm{ML}_\infty$$
$$\mathrm{GF} \bullet\!\longrightarrow \mathrm{GF}_\infty$$
$$\mathrm{FO}^k \bullet\!\longrightarrow \mathrm{L}^k_\infty$$

Limit behaviour of the games in FMT

over finite (or ω-saturated) structures:

$$\bigcap_m \approx_m \;=\; \approx \;=\; \equiv_{\mathcal{L}} \;=\; \equiv_{\mathcal{L}_\infty}$$

Finite Variable Logics

FO too weak, $\equiv_{\mathsf{FO}}$ trivial, L_∞ defines everything

in contrast: $\mathbf{L}^\omega_\infty = \bigcup_k \mathbf{L}^k_\infty \;\subsetneq\; \mathbf{L}_\infty$

infinitary finite-variable logic

- important as a framework
  for the study of expressibility
- non-trivial gradation FO^k — L^k_∞
- FO^k algorithmically natural fragment of FO

e.g. model-checking (combined complexity)

FO^k Ptime-complete
FO Pspace-complete Vardi

Preservation / Characterisation Thms

relating semantic closure properties to syntactic form

classically, for $\varphi \in \mathrm{FO}$: (failing in FMT: Tait)

Mod(φ) closed under extensions $\Leftrightarrow$ $\varphi \equiv \varphi'$ for some existential φ'

similarly: monotonicity — positivity
$\simeq^2$-invariance — FO^2

- classical proofs involve compactness *and* passage through infinite structures
- interest: syntactic adequacy, normal forms
  recursive syntax for non-recursive classes

classically *and* in FMT, for $\varphi \in \mathrm{FO}$:

Mod(φ) closed under bisimulation $\Leftrightarrow$ $\varphi \equiv \varphi'$ for some modal formula φ'

classically: van Benthem
FMT, entirely new proof: Rosen

similarly for $\varphi \in \mathrm{MSO}$ and L_μ

Janin,Walukiewicz
remains open for FMT

Descriptive Complexity

measuring logics against complexity

Capturing (Gurevich, following Chandra,Harel)

matching $\begin{cases} \text{complexity class } \mathcal{C} \\ \text{logic L} \end{cases}$

$\mathcal{C}$: MC problems in data complexity $\mathcal{C}$

L: MC problems expressible in L

$\mathsf{L} \equiv \mathcal{C}$, L captures $\mathcal{C}$:

- model-checking for L is in $\mathcal{C}$ (effectively)
- any model-checking problem in $\mathcal{C}$ is expressible in L

L provides recursive syntax for $\mathcal{C}$ with recursive compilation into $\mathcal{C}$

Examples of Capturing

Büchi, Elgot, Trakhtenbrot

$\text{Fin.Aut.} \equiv \text{MSO}$ (strings/words)

Fagin

$\text{NP} \equiv \Sigma^1_1$ (all structures)

Immerman, Vardi

$\text{Ptime} \equiv \text{LFP}$ (ordered structures)

Abiteboul, Vardi, Vianu

$\text{Pspace} \equiv \text{PFP}$ (ordered structures)

LFP and PFP: FO plus relational recursion
least and partial fixed-point logic

Fixed-Point Recursion

augmenting FO with recursion mechanisms

FO static and local,
e.g. fails to define transitive closures

natural levels of recursion,
compatible with finite-variable gradation

in the spectrum $\mathrm{FO} \ \bullet\!\!-\!\!-\!\!-\!\!\overset{\mathrm{LFP}}{\bullet}\!\!-\!\!\longrightarrow \ \mathrm{L}^{\omega}_{\infty}$

analogous to $\mathrm{ML} \ \bullet\!\!-\!\!-\!\!-\!\!\overset{\mathrm{L}_{\mu}}{\bullet}\!\!-\!\!\longrightarrow \ \mathrm{ML}_{\infty}$

fixed-point operators

with $\varphi(X, \boldsymbol{x})$ associate operator

$$\varphi^{\mathfrak{A}} \colon P \longmapsto \left\{ \boldsymbol{a} : \mathfrak{A} \models \varphi[P, \boldsymbol{a}] \right\}$$

and iteration stages

$$\begin{aligned} \varphi^{0}[\mathfrak{A}] &= \emptyset \\ \varphi^{n+1}[\mathfrak{A}] &= \varphi^{\mathfrak{A}}\Big[\varphi^{n}[\mathfrak{A}]\Big] \end{aligned}$$

Partial Fixed Point — Pspace

$$\left[\mathrm{PFP}_{X,\boldsymbol{x}}\varphi\right]^{\mathfrak{A}} = \begin{cases} \varphi^n[\mathfrak{A}] & \text{if } \varphi^{n+1}[\mathfrak{A}] = \varphi^n[\mathfrak{A}] \\ \emptyset & \text{if nontrivially periodic} \end{cases}$$

Least Fixed Point — Ptime

φ *positive* in X $\Rightarrow$ $\varphi^{\mathfrak{A}}$ monotone operation
with a least fixed point

$$\left[\mathrm{LFP}_{X,\boldsymbol{x}}\varphi\right]^{\mathfrak{A}} = \bigcup_n \varphi^n[\mathfrak{A}]$$

Inflationary Fixed Point — Ptime

For arbitrary φ iterate $\varphi^*(X,\boldsymbol{x}) = X\boldsymbol{x} \vee \varphi(x,\boldsymbol{x})$

$$\left[\mathrm{IFP}_{X,\boldsymbol{x}}\varphi\right]^{\mathfrak{A}} = \left[\mathrm{PFP}_{X,\boldsymbol{x}}\varphi^*\right]^{\mathfrak{A}}$$

Example: transitive closure as a LFP:

$$\mathrm{LFP}_{X,xy}\big(Exy \vee \exists z(Xxz \wedge Ezy)\big)$$

Fixed-Point Logics LFP, IFP, PFP

Expressiveness and Complexity

LFP, IFP, PFP:

FO + closure under corresponding fixed points
cf. L_μ: closure of ML under LFP

syntactic/semantic inclusions

$$\mathrm{FO} \subsetneq \mathrm{LFP} \subseteq \mathrm{IFP} \subseteq \mathrm{PFP} \subsetneq \mathrm{L}^\omega_\infty$$

$\rightarrow$ compatibility with finite-variable gradation
governed by finite-variable equivalences

$\varphi(X, \boldsymbol{x})$ in FO: iteration stages φ^n
definable in FO

similarly for L^ω_∞, in fact L^k_∞ closed
w.r.t. parameter-free fixed points

Fixed-Point Logics LFP, IFP, PFP

Expressiveness and Complexity

Immerman

LFP normal form: one LFP-application suffices

no alternation hierarchy in LFP (unlike L_μ)

Gurevich,Shelah

LFP $\equiv$ IFP over all (finite!) structures

using Moschovakis' Stage Comparison Thm

Immerman,Vardi / Abiteboul,Vardi,Vianu

$$\left.\begin{array}{rcl} \text{Ptime} & \equiv & \text{LFP} \\ \text{Pspace} & \equiv & \text{PFP} \end{array}\right\} \text{ over ordered structures}$$

Abiteboul,Vianu

LFP $\neq$ PFP (all) $\Leftrightarrow$ Ptime $\neq$ Pspace

The Role of Order

NO capturing results below NP on all structures

- need order for coding and simulation
- order yields canonical input representation (!)

MC algorithm must yield same result on (different encodings of) isomorphic inputs

genericity/$\simeq$-invariance : non-recursive

with order: canonical representations up to $\simeq$
as $(A, <, \dots) \simeq (\{1, 2, \dots, n\}, <, \dots)$

implicit canonisation

Is there a logic for Ptime?

NB: "no" $\Rightarrow$ Ptime $\neq$ NP

representation problem:

"Ptime" $=$ Ptime$/\simeq$ (!)

all partial successes rely on
efficient canonisation / reduction to ordered case

Canonisation towards Capturing Results

canonisation w.r.t. $\simeq$ and w.l.o.g. for graphs:

$$\mathsf{H}\colon \mathfrak{G} \longmapsto \mathsf{H}(\mathfrak{G}) = \big(\{1,\ldots,|G|\}, E\big)$$

such that
- $\mathsf{H}(\mathfrak{G}) \simeq \mathfrak{G}$
- $\mathfrak{G} \simeq \mathfrak{G}' \;\Leftrightarrow\; \mathsf{H}(\mathfrak{G}) = \mathsf{H}(\mathfrak{G}')$

IF/WHERE Ptime canonisation H available:

$$\mathsf{Ptime} \;\equiv\; \mathsf{Ptime} \circ \mathsf{H}$$

Ptime/$\simeq$	no semantic constraints all polyn. clocked alg.

IF/WHERE definable Ptime canonisation H available:

i.e., ordered copy of $\mathsf{H}(\mathfrak{G})$ L-definably interpretable over $\mathfrak{G}$, LFP $\subseteq$ L $\subseteq$ Ptime

$$\mathsf{Ptime} \equiv \mathsf{Ptime} \circ \mathsf{H} \equiv \mathsf{LFP}\big(\mathsf{H}\big) \equiv \mathsf{L}$$

by the Immerman-Vardi-Thm

Capturing Some of Ptime

orthogonal approaches:

- restricted classes of unordered structures
- stronger closure properties

Examples

Grohe, Mariño

Ptime captured by IFP + counting
on large natural classes of graphs:
on planar graphs / S-embeddable graphs

definable canonisation

O_

Capturing results on all structures for fragments
$\mathrm{Ptime}/\!\approx\ \subseteq\ \mathrm{Ptime}/\!\simeq$ for $\approx$ rougher than $\simeq$:
modal and 2-variable fragments of Ptime

canonisation, inversion of game-invariants

Bisimulation-Invariant Ptime

Ptime$/\!\sim$

all Ptime MC problems
that respect bisimulation

natural canonisation w.r.t. bisimulation

w.l.o.g. consider structures $\mathfrak{W} = (W, E, P_i)$

$$\mathrm{I}(\mathfrak{W}, w) = \bigl(E^*[w]\,/\!\sim, E^{\sim}, P_i\,/\!\sim, [w]^{\sim}\bigr)$$

succinct quotient representation
w.r.t. bisimulation game

LFP-definable ordering of quotients

inductive refinement of $\sim_m$-classes

lexicographically order $\sim_{m+1}$-classes
w.r.t. to incidence with $\sim_m$-classes

$\Rightarrow$

$$\text{Ptime}/\!\sim \;\equiv\; \text{LFP}(\mathrm{I}) \quad \equiv \quad \mathsf{L}_\mu^\omega := \bigcup_k \mathsf{L}_\mu^k$$

L_μ^k μ-calculus over the induced systems on k-tuples of states

$$\mathrm{ML} \;\subsetneq\; \mathsf{L}_\mu \;\subsetneq\; \mathsf{L}_\mu^\omega \;\subsetneq\; \mathrm{ML}_\infty$$

Canonisation and Game Analysis

- succinct quotient w.r.t. game yields ordered complete invariants I^k w.r.t. $\simeq^k/\equiv^k$

  NOT a representative of its class

- canonisation requires model reconstruction:

  possible in Ptime for $\simeq^2$ and $\simeq^2_{\mathrm{ctg}}$
  impossible for $k = 3$ (Grohe)

- get evaluation of LFP and PFP in terms of quotients w.r.t. $\simeq^k$:

  $$\mathrm{LFP} \equiv \bigcup_k \mathrm{Ptime}(\mathrm{I}^k)$$

  $$\mathrm{PFP} \equiv \bigcup_k \mathrm{Pspace}(\mathrm{I}^k)$$

  $\rightarrow$ Abiteboul-Vianu-Theorem:
  $\mathrm{PFP} = \mathrm{LFP} \Leftrightarrow \mathrm{Pspace} = \mathrm{Ptime}$

Some other Model Checking Issues

Vardi

w.r.t. combined complexity:

$$\left.\begin{array}{c}\mathsf{MC(FO)}\\ \mathsf{MC(FO}^k)\end{array}\right\}\quad \text{complete for}\quad \left\{\begin{array}{c}\mathsf{Pspace}\\ \mathsf{Ptime}\end{array}\right.\quad (k \geqslant 2)$$

indeed, $\mathsf{MC(FO}^2)$, like $\mathsf{MC(ML)}$, in $\mathcal{O}(|\mathfrak{A}| \cdot |\varphi|)$

Frick, Grohe

FO has linear time data complexity
on locally tree-decomposable graphs

Gottlob, Grädel, Veith

linear time data complexity for Datalog Lite:

a guarded version of stratified Datalog
allowing (guarded) universally quantified literals

comprising CTL and alternation-free L_μ

Model Theory for Specific Domains
logical modelling

to encode problem instances as structures

$$\text{determine} \quad \left\{ \begin{array}{l} \text{basic class of structures} \\ \textbf{notion of equivalence} \end{array} \right.$$

$$\text{to match} \quad \left\{ \begin{array}{l} \text{the right objects} \\ \text{the right } \textbf{level of abstraction} \end{array} \right.$$

$\Uparrow$ the semantic domain

$\Downarrow$ the logic

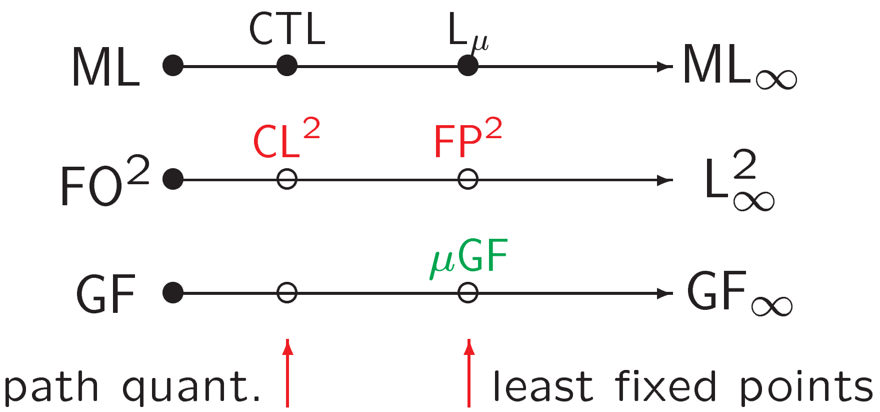
chosen to respect prescribed equivalence
and to meet expressiveness requirements

choice of logic subordinate to choice of semantic domain

e.g. modal / 2-variable / guarded domains

Issues in Algorithmic Model Theory for Specific Domains

SAT: comparative analysis



guarded domain inherits modal robustness
unlike 2-variable domain

Grädel,O_,Rosen / Grädel,Walukiewicz / ...

Model Checking Complexity

- Ptime fragments
- syntactic criteria for tractability

Eliminability of LFP-Recursion

collapse and boundedness

- where least fixed points collapse
  $\rightarrow$ variants of McColm's conjectures
- which least fixed points collapse
  (decision problem)

Examples

Boundedness $\begin{cases} \text{undecidable over FO}^2 \\ \text{decidable over ML} \end{cases}$

Kolaitis,O_

Decidability of the bottom levels
of the alternation hierarchy in L_μ

O_ / Wilke

in particular have a decision procedure for the

recursion-free fragment of L_μ
corresponding to $\mathrm{ML} \subseteq \mathrm{L}_\mu$

Model Theory for Specific Domains

logical modelling

adapt the model theoretic toolkit
to semantic requirements

typically one does
not want to deal with

- all finite structures/graphs/ . . .

nor necessarily with

- finite structures up to isomorphism

isolate new relevant domains
e.g. guarded domain

study domains comparatively
e.g. modal vs. 2-variable