Monadic Second Order Logic

Zdeněk Dvořák

September 14, 2015

1 Courcelle's theorem

A formula in *Monadic Second Order Logic* (MSOL) may contain:

- variables v_1, v_2, … for vertices
- variables e_1, e_2, … for edges
- variables V_1, V_2, … for sets of vertices
- variables E_1, E_2, … for sets of edges
- equality $=$ for vertices and edges
- incidence predicate $i(v_1, e_1)$, …—is v_1 incident with e_1?
- set membership predicates $v_1 \in V_1$, $e_1 \in E_1$, …
- logical conjunctions $\vee$, $\neg$
- existential quantifier $\exists$ applied to any kind of variable
- anything else expressible from these, for example
  - universal quantifier $(\forall X)\psi \equiv \neg(\exists X)\neg\psi$
  - "and" conjuction $\psi_1 \wedge \psi_2 \equiv \neg(\neg\psi_1 \vee \neg\psi_2)$
  - non-equality $v_1 \neq v_2 \equiv \neg(v_1 = v_2)$
  - adjacency $e(v_1, v_2) \equiv v_1 \neq v_2 \wedge (\exists e_1)\, i(v_1, e_1) \wedge i(v_2, e_1)$
  - implication $\psi \Rightarrow \varphi \equiv \neg\psi \vee \varphi$
  - subset relation $V_1 \subseteq V_2 \equiv (\forall v_1)\, v_1 \in V_1 \Rightarrow v_1 \in V_2$
  - equality for sets $V_1 = V_2 \equiv V_1 \subseteq V_2 \wedge V_2 \subseteq V_1$

If φ is a formula in MSOL without free variables and G is a graph, we write $G \models \varphi$ if G satisfies the formula φ.

Formally: A *variable assignment* in G for a formula φ is a function σ such that

- for each free vertex variable v_i, $\sigma(v_i)$ is a vertex of G,
- for each free edge variable e_i, $\sigma(e_i)$ is an edge of G,
- for each free vertex set variable V_i, $\sigma(V_i)$ is a set of vertices of G, and
- for each free edge set variable E_i, $\sigma(E_i)$ is a set of edges of G.

We write $G, \sigma \models \varphi$ when

- $\varphi \equiv \alpha = \beta$ for some variables α, β such that $\sigma(\alpha) = \sigma(\beta)$
- $\varphi \equiv i(v_1, e_1)$ for some variables v_1, e_1 such that $\sigma(v_1)$ is incident with $\sigma(e_1)$
- $\varphi \equiv \alpha \in A$ for vertex or edge variable α and vertex set or edge set variable A such that $\sigma(\alpha)$ is an element of $\sigma(A)$.
- $\varphi \equiv \psi_1 \vee \psi_2$ and $G, \sigma \models \psi_1$ or $G, \sigma \models \psi_2$
- $\varphi \equiv \neg\psi$ and the claim $G, \sigma \models \psi$ does not hold
- $\varphi \equiv (\exists\alpha)\, \psi$ and there exists a variable assignment σ' in G for ψ that is obtained from σ by assigning a value to the variable α and satisfies $G, \sigma' \models \psi$

Then, $G \models \varphi$ is a shorthand for $G, \sigma \models \varphi$, where σ is the null function.

Many interesting graph properties can be expressed in MSOL. For example,

- V_1 is an independent set in G:
$$I(V_1) \equiv (\forall v_1, v_2)\, v_1 \in V_1 \wedge v_2 \in V_1 \Rightarrow \neg e(v_1, v_2).$$
- G is 3-colorable if and only if
$$G \models (\exists V_1, V_2, V_3)\, I(V_1) \wedge I(V_2) \wedge I(V_3) \wedge (\forall v_1)\, v_1 \in V_1 \vee v_1 \in V_2 \vee v_1 \in V_3.$$
- G has a perfect matching if and only if
$$G \models (\exists E_1)\,(\forall v_1)\,(\exists e_1)\, e_1 \in E_1 \wedge i(v_1, e_1) \wedge [(\forall e_2)\,(e_2 \in E_1 \wedge i(v_1, e_2)) \Rightarrow e_1 = e_2]$$

This makes the following algorithmic result very interesting.

Theorem 1 (Courcelle). *For any MSOL formula φ without free variables and for any $k \geq 0$, there exists a linear-time algorithm deciding whether a graph G of tree-width at most k satisfies $G \models \varphi$.*

All the natural variants of this claim are true as well.

Theorem 2 (Courcelle). *For any MSOL formula φ with one free variable V_1 and for any $k \geq 0$, there exist polynomial-time algorithms that given a graph G of tree-width at most k*

- *find the largest set $V_1 \subseteq V(G)$ satisfying φ, and*
- *determine the number of sets $V_1 \subseteq V(G)$ satisfying φ.*

2 Limits of MSOL

Ehrenfeucht-Fraïssé game: Two players, Spoiler and Duplicator. Given two graphs A_1 and A_2, and integer $n \geq 1$. For $k = 1, \ldots, n$,

- Spoiler chooses $i_k \in \{1, 2\}$, and selects a vertex/edge/set of vertices/set of edges X_{k,i_k} in A_{i_k}.
- Duplicator chooses $i_k \in \{1, 2\}$, and selects a vertex/edge/set of vertices/set of edges (the same type as Spoiler) $X_{k,3-i_k}$ in A_{3-i_k}.

Duplicator wins if there are the same relations among the corresponding selected objects, i.e., for $1 \leq k, t \leq n$,

- If $X_{k,1}$ is a vertex/edge and $X_{t,1}$ is a vertex/edge, then $X_{k,1} = X_{t,1}$ if and only if $X_{k,2} = X_{t,2}$.
- If $X_{k,1}$ is a vertex and $X_{t,1}$ is an edge, then $X_{k,1}$ is incident with $X_{t,1}$ if and only if $X_{k,2}$ is incident with $X_{t,2}$.
- If $X_{k,1}$ is a vertex/edge and $X_{t,1}$ is a set of vertices/edges, then $X_{k,1} \in X_{t,1}$ if and only if $X_{k,2} \in X_{t,2}$.

Lemma 3. *Let P be a graph property. Suppose that for every $n \geq 1$, there exist graphs $A_{n,1}$ and $A_{n,2}$ such that $A_{n,1}$ has the property P, $A_{n,2}$ does not have the property P, and Duplicator has a winning strategy for the Ehrenfeucht-Fraïssé game for n and the graphs $A_{n,1}$ and $A_{n,2}$. Then P cannot be expressed in MSOL.*

Proof. Suppose for a contradiction that a graph G has the property P if and only if $G \models \varphi$ for some MSOL formula φ without free variables. We can assume that φ is in prenex normal form, that is,

$$\varphi = (Q_1\alpha_1)(Q_2\alpha_2)\dots(Q_n\alpha_n)\psi(\alpha_1,\dots,\alpha_n)$$

for some existential or universal quantifiers $Q_1, \dots, Q_n$, variables $\alpha_1, \dots, \alpha_n$ and a quantifier-free MSOL formula ψ.

We are going to describe Spoiler's winning strategy in n-round Ehrenfeucht-Fraïssé game for graphs $A_{n,1}$ and $A_{n,2}$. Consider the situation after first k rounds. For $i \in \{1,2\}$, let $\sigma_{k,i}$ be the variable assignment in $A_{n,i}$ such that for $j = 1,\dots,k$, we have $\sigma_{k,i}(\alpha_j) = X_{j,i}$. We maintain the invariants that

$$A_{n,1}, \sigma_{k,1} \models (Q_{k+1}\alpha_{k+1})(Q_{k+2}\alpha_{k+2})\dots(Q_n\alpha_n)\psi(\alpha_1,\dots,\alpha_n) \tag{1}$$

and

$$A_{n,2}, \sigma_{k,2} \not\models (Q_{k+1}\alpha_{k+1})(Q_{k+2}\alpha_{k+2})\dots(Q_n\alpha_n)\psi(\alpha_1,\dots,\alpha_n). \tag{2}$$

For $k = 0$, these invariants hold since $A_{n,1}$ has the property P and $A_{n,2}$ does not have the property P.

Suppose that we already managed to play the first k rounds while preserving the invariant, $0 \le k \le n-1$. If $Q_{k+1} = \exists$, then Spoiler sets $i_{k+1} = 1$ and chooses the value $X_{k+1,1}$ for α_{k+1} in $A_{n,1}$ so that

$$A_{n,1}, \sigma_{k+1,1} \models (Q_{k+2}\alpha_{k+2})\dots(Q_n\alpha_n)\psi(\alpha_1,\dots,\alpha_n).$$

Since (2) holds, Duplicator cannot choose any value $X_{k+1,2}$ for α_{k+1} in $A_{n,2}$ to violate the second invariant.

Similarly, if $Q_{k+1} = \forall$, then Spoiler sets $i_{k+1} = 2$ and chooses the value $X_{k+1,2}$ for α_{k+1} in $A_{n,2}$ so that

$$A_{n,2}, \sigma_{k+1,2} \not\models (Q_{k+2}\alpha_{k+2})\dots(Q_n\alpha_n)\psi(\alpha_1,\dots,\alpha_n).$$

Since (1) holds, Duplicator cannot choose any value $X_{k+1,1}$ for α_{k+1} in $A_{n,1}$ to violate the first invariant.

Hence, Spoiler can ensure that the invariants (1) and (2) hold for $k = n$, that is,

$$A_{n,1}, \sigma_{n,1} \models \psi(\alpha_1,\dots,\alpha_n)$$

and

$$A_{n,2}, \sigma_{n,2} \not\models \psi(\alpha_1,\dots,\alpha_n).$$

However, this implies that ψ contains some relation which is true in the variable assignment $\sigma_{n,1}$ and false in the variable assignment $\sigma_{n,2}$, or vice versa, and thus Spolier wins the game. This is a contradiction. □

Let us remark that the reverse implication (if there exists $n \geq 1$ such that Spoiler wins any n-round game with graphs A_1 satisfying P and A_2 not satisfying P, then P can be expressed in MSOL) is true as well, but somewhat harder to prove and much less useful.

Example: the property "the graph has even number of vertices" cannot be expressed in MSOL.

Proof. Let $A_{n,1}$ consist of 2^n isolated vertices and $A_{n,2}$ of $2^n + 1$ isolated vertices. Hence, in the game, no edges or sets of edges will be chosen. Treat the chosen vertices as single-vertex sets. For $k = 0, \ldots, n$, ensure that after the k-th step, the following invariant holds:

- For all disjoint sets $K, L \subseteq \{1, \ldots, k\}$, the sets

  $$\bigcap_{i \in K} X_{i,1} \setminus \bigcup_{i \in L} X_{i,1}$$

  and

  $$\bigcap_{i \in K} X_{i,2} \setminus \bigcup_{i \in L} X_{i,2}$$

  either have the same size, or they both have size at least 2^{n-k}.

This ensures that in the end, if $X_{k,1}$ is a vertex, then

- when $X_{t,1}$ is a set, then $X_{k,1} \in X_{t,1}$ if and only if $X_{k,2} \in X_{t,2}$,
- when $X_{t,1}$ is a vertex, then $X_{k,1} = X_{t,1}$ if and only if $X_{k,2} = X_{t,2}$,

since the sets $X_{k,1} \setminus X_{t,1}$ and $X_{k,2} \setminus X_{t,2}$ are either both empty or both have at least one element. $\square$

3 Exercises

1. ($\star$) Write a formula in MSOL with one free vertex set variable V_1 which is true exactly when V_1 is a dominating set.

2. ($\star\star$) Write a formula in MSOL which is true exactly for Hamiltonian graphs.

3. ($\star\star\star$) Write a formula in MSOL which is true exactly for planar graphs.

4. ($\star\star\star$) Prove that the property "G has the same number of vertices of degree 0 and 1" is not expressible in MSOL.