Automatizované řešení úloh s omezeními

Martin Kot

Katedra informatiky, FEI,
Vysoká škola báňská – Technická universita Ostrava
17. listopadu 15, Ostrava-Poruba 708 33
Česká republika

26. září 2013

Sítě omezení

Síť omezení

- **Síť omezení** (constraint network) se skládá množiny proměnných $X = \{x_1, \dots, x_n\}$ s vlastními doménami $D = \{D_1, \dots, D_n\}$ (obsahujícími možné hodnoty pro každou proměnnou $D_i = \{v_1, \dots, v_k\}$) a množiny omezení $C = \{C_1, \dots, C_t\}$. Formálně tedy jde o trojici $\mathcal{R} = (X, D, C)$.
- Omezení C_i je relace R_i definovaná na podmnožině proměnných $S_i, S_i \subseteq X$.
- S_i se nazývá **rozsah** (scope) relace R_i
- Relace určují vzájemné povolené přiřazení hodnot proměnným
- Pro $S_i = \{x_{i_1}, \dots, x_{i_r}\}$ je $R_i \subseteq D_{i_1} \times \dots \times D_{i_r}$
- Omezení tedy můžeme chápat jako dvojici $C_i = (S_i, R_i)$
- Pozn. - Při jasném rozsahu používáme jen R_i nebo rozsah uvedeme v indexu R_{S_i}, často taky bez závorek R_{xyz}
- Množina rozsahů $S = \{S_1, \dots, S_t\}$ - **schéma sítě** (network scheme)

Síť omezení

- Předpokládá se jen jedno omezení nad každým $S_i \in S$ (při relačních omezeních)
- Arita omezení = kardinalita rozsahu
- Binární síť omezení (binary constraint network) - má pouze unární a binární omezení
- Příklad - formalizace n dam na šachovnici
  - Slidy Riny Dechter, chapter 2, slides 2-3
  - Sloupce jsou proměnné $x_1, \ldots, x_n$
  - Možné řádky jsou domény $D_i = \{1, \ldots, n\}$
  - Omezení jsou binární - vyjadřují, že dámy se nesmí ohrožovat vyjmenováním povolených dvojic řádků pro dvojice sloupců
  - Pro 4 dámy: $\mathcal{R} = (X, D, C)$, $X = \{x_1, x_2, x_3, x_4\}$, $\forall i D_i = \{1, 2, 3, 4\}$, $C_1 = R_{12}$, $C_2 = R_{13}$, $C_3 = R_{14}$, $C_4 = R_{23}$, $C_5 = R_{24}$, $C_6 = R_{34}$

Instanciace sítě

- **Instanciace** (instantiation) **podmnožiny proměnných** - celé podmnožině proměnných se přiřadí hodnoty z příslušných domén
- Formálně je instanciace množiny $\{x_{i_1}, \ldots, x_{i_k}\}$ $n$tice dvojic $((x_{i_1}, a_{i_1}), \ldots, (x_{i_k}, a_{i_k}))$, kde (x, a) reprezentuje přiřazení $a \in D_x$ proměnné x
- Pozn. Používá se taky $(x_1 = a_1, \ldots, x_i = a_i)$ nebo $\bar{a} = (a_1, \ldots, a_i)$

Splnění omezení

- Instanciace $\bar{a}$ **splňuje omezení** (S, R) právě tehdy, když je definována nad všemi proměnnými z S a prvky $\bar{a}$ asociované s S jsou v relaci R
- Příklad: uvažujme $R_{xyz} = \{(1,1,1),(1,0,1),(0,0,0)\}$. Potom $\bar{a} = ((x,1),(y,1),(z,1),(t,0))$ splňuje R_{xyz} ale $\bar{a} = ((x,1),(y,0),(z,0),(t,0))$ ne, protože $(1,0,0) \notin R_{xyz}$.

Konzistentní částečná instanciace

- Částečná instanciace je **konzistentní**, když splňuje omezení, v jejichž rozsahu není žádná neinstanciovaná proměnná
- Projekci $\bar{a}$ na podmnožinu rozsahu S_i označujeme $\bar{a}[S_i]$ nebo $\pi_{S_i}(\bar{a})$

Řešení

- **Řešení** (solution) sítě $\mathcal{R} = (X, D, C)$, kde $X = \{x_1, \ldots, x_n\}$ je instanciace všech proměnných, která splňuje všechna omezení
- **Relace řešení** $sol(\mathcal{R})$ (ozn. také ρ_X) je definována jako $sol(\mathcal{R}) = \{a = (a_1, \ldots, a_n) \mid a_i \in D_i, \forall S_i$ ve schématu R je $\bar{a}[S_i] \in R_i\}$
- Říkáme, že síť omezení **vyjadřuje** (express) nebo **reprezentuje** (represents) relaci všech svých řešení.
- Pro síť $\mathcal{R}$ nad X a $A \subseteq X$ je $sol(A)$ nebo ρ_A množina konzistentních instanciací nad A
- Příklad: 4 dámy na šachovnici (slidy Riny Dechter, chapter 2, slide 3) - Uvažujme $Y = \{x_1, x_2, x_3\}$, instanciaci $\bar{a} = ((x_1, 1), (x_2, 4), (x_3, 2))$ na obr. (a). Ta je konzistentní, protože $\bar{a}[\{x_1, x_2\}] = (1, 4)$ a $(1, 4) \in R_{12}$ atd. Ale $\bar{a}$ není součástí žádného řešení. 4 dámy na šachovnici reprezentují $\rho_{1234} = \{(2, 4, 1, 3), (3, 1, 4, 2)\}$.

Graf omezení

- Jedna z reprezentací sítě omezení je prostřednictvím **grafu omezení** ((primal) constraint graph)
- Uzly reprezentují proměnné, hrany spojují proměnné které jsou svázány některým z omezení
- Pro binární omezení - chybějící hrana znamená úplnou relaci mezi příslušnými proměnnými
- Dává smysl pro binární i jiná omezení

Hypergraf, duální graf

- **Hypergraf** je struktura $\mathcal{H} = (V, S)$, V je množina vrcholů, $S = \{S_1, \ldots, S_l\}$ množina podmnožin množiny vrcholů ($S_i \subseteq V$) nazývaných hyperhrany
- V hypergrafu omezení opět vrcholy reprezentují proměnné a hyperhrany (kreslené jako oblasti) reprezentují omezení
- Je přesnější reprezentací nebinárních omezení
- **Duální graf omezení**
  - Hyperhrana původního grafu (tedy rozsah omezení) je reprezentována vrcholem
  - Hrany spojují vrcholy s neprázdným průnikem hyperhran (tedy sdílející některou z proměnných)
  - Hrany jsou ohodnoceny sdílenou proměnnou

Duální problém

- **Duální problém** (dual problem) vychází z duálního grafu
- Omezení tvoří proměnné nazývané **c-proměnné** (c-variables)
- Domény proměnných jsou tvořeny možnými kombinacemi hodnot vynucenými odpovídajícími omezeními
- Omezení mezi c-proměnnými vynucují, že jimi sdíleným (původním) proměnným musí být přiřazené stejné hodnoty
- Jakoukoliv síť omezení takto je možno převést na binární a řešit ji technikami pro binární

Numerická omezení

- **Numerická omezení** (numeric constraints) vyjadřují omezení aritmetickými výrazy
- Příklad: Pro dámy na šachovnici můžeme dát omezení $\forall i, j : x_i \neq x_j \wedge |x_i - x_j| \neq |i - j|$
- Příklad: Můžeme uvažovat proměnné s doménami, které jsou podmnožiny přirozených čísel. Omezení ve tvaru konjunkce lineárních nerovnic $ax_i - bx_j = c$, $ax_i - bx_j < c$, $ax_i - bx_j \leq c$. Jde o speciální případ tzv. celočíselného lineárního programování.
- Příklad: SEND + MORE = MONEY z prvního cvičení

Booleovská omezení

- Když jsou proměnné dvouhodnotové, využívá se často jazyk výrokové (nebo také booleovské) logiky
- Výrokové proměnné nabývají hodnot $\{true, false\}$ nebo $\{1, 0\}$ a označujeme je $P, Q, R, \ldots$
- Literály jsou P (P je true) nebo $\neg P$ (P je false)
- Klauzule - disjunkce literálů
- Operace rezoluce mezi $\alpha \vee Q$ a $\beta \vee \neg Q$ dává $\alpha \vee \beta$
- Konjunktivní normální forma (také označována jako teorie) - množina klauzulí označující jejich konjunkci
- Model nebo řešení teorie ϕ je přiřazení pravdivostních hodnot proměnným tak, aby všechny klauzule byly pravdivé
- SAT - má formule model? (je příslušný problém konzistentní?)

Binární sítě omezení

- Základním konceptem práce s omezeními je vyvozování nových omezení
- Nová omezení můžou být zpřísněním původních nebo mohou omezovat proměnné, které dříve omezené nebyly
- Příklad: Z $x \geq y$, $y \geq z$ lze vyvodit $x \geq z$
- Odvozené omezení je redundantní vzhledem k síti - jeho odstranění neovlivní množinu řešení
- **Skládání** (composition): pro dané binární (popř. unární) omezení R_{xy}, R_{yz} složení $R_{xy} \cdot R_{yz}$ generuje binární relaci $R_{xz} = \{(a, b) \mid a \in D_x, b \in D_z, \exists c \in D_y$ tž. $(a, c) \in R_{xy}$ a $(c, b) \in R_{yz}\}$
- Skládání lze popsat také jako násobení booleovských matic

Projekční síť

- Ne každá obecná relace může být reprezentována sadou binárních omezení (relací s n proměnnými a k hodnotami v doménách je 2^{k^n}, binárních sítí omezení jen $2^{k^2 n^2}$)
- Ale můžeme takovou relaci aproximovat binárně, například **projekční sítí** (projection network)
- Projekční síť relace ρ je projekce ρ na každou dvojici jejích proměnných
- Formálně, když ρ je relace nad $X = \{x_1, \ldots, x_n\}$, její projekční síť $P(\rho)$ je definována jako $\mathcal{P} = (X, D, P)$ kde $D = \{D_i \mid 1 \leq i \leq n\}$, $D_i = \pi_i(\rho)$,$P = \{P_{ij} \mid 1 \leq i, j \leq n\}$, $P_{ij} = \pi_{x_i, x_j}(\rho)$
- Řešení sítě $P(\rho)$ vždy obsahuje ρ
- Neexistuje binární síť $\mathcal{R}'$ taková, že $\rho \subseteq sol(R') \subset sol(P(\rho))$
- Když relace nemůže být reprezentována projekční sítí, nemůže být reprezentována žádnou binární sítí

Minimální síť

- Pro dvě binární sítě $\mathcal{R}, \mathcal{R}'$ na stejné množině proměnných $x_1, \ldots, x_n$ říkáme, že $\mathcal{R}'$ je alespoň tak striktní (tight) jako $\mathcal{R}$ právě tehdy, když $\forall i, j : R'_{ij} \subseteq R_{ij}$
- Průnik sítí $\mathcal{R}, \mathcal{R}'$, označovaný $\mathcal{R} \cap \mathcal{R}'$ je binární síť získaná vzájemným průnikem všech vzájemně si odpovídajících omezení
- Když jsou sítě $\mathcal{R}, \mathcal{R}'$ ekvivalentní, je s nimi ekvivalentní i $\mathcal{R} \cap \mathcal{R}'$ a je minimálně tak striktní jako obě tyto sítě
- Existuje částečné uspořádání sítí podle striktnosti
- **Minimální síť** je průnikem všech ekvivalentních sítí
- Minimální síť je ekvivalentní se sítěmi, ze kterých průnikem vznikla a je alespoň tak striktní, jako každá z nich
- Formálně: Nechť $\{\mathcal{R}_1, \ldots, \mathcal{R}_l\}$ je množina všech sítí ekvivalentních s R_0 a nechť $\rho = sol(\mathcal{R}_0)$. Potom minimální síť M pro $\mathcal{R}_0$ nebo ρ je definována jako $M(R_0) = M(\rho) = \bigcap_{i=1}^{l} R_i$

Minimální síť

- Minimální síť je identická s projekční sítí množiny řešení minimální sítě. Tedy pro každou binární síť $\mathcal{R}$ tž. $\rho = sol(\mathcal{R})$ platí $M(\rho) = P(\rho)$
- Binární omezení v minimální síti označujeme M_{ij}
- Unární omezení jsou redukované domény
- Když nějaká dvojice hodnot proměnných vyhovuje omezení v minimální síti, potom se vyskytuje v alespoň jednom řešení