The Complexity of Approximating Counting Problems

A.Antonopoulos (N.T.U.A.)
Network Algorithms

June 2011

1. Introduction
   - The Class #P
   - Approximation Schemes

2. Approximate Counting Problems

Basic Definitions

- There are many problems where we want to *count* the **number** of solutions.
- Of course, this is more "difficult" than finding if a solution exists!
- We want to define the class of counting the number of solutions to **NP** problems:

Definition

Let $L \in \mathbf{NP}$, M its associated verifier, and polynomial p the bound on the length of its "*Yes*" certificates.
For a string $x \in \Sigma^*$, define $f(x)$ to be the number of strings y such that $|y| \leq p(|x|)$ and $M(x, y) = 1$.
Functions $f : \Sigma^* \rightarrow \mathbb{N}$ constitute the class $\mathbf{\#P}$.

Basic Definitions

Definition (#P-Completeness)

Function f is said to be **#P**-*complete* if *every* function $g \in$ **#P** can be reduced to f in the following sense:

- There is a polynomial-time function $R : \Sigma^* \to \Sigma^*$ such that, given an instance x of g, produces an instance $R(x)$ of f.
- There is a polynomial-time function $S : \Sigma^* \times \mathbb{Z}^+ \to \mathbb{Z}^+$ such that, given x and $f(R(x))$, computes $g(x)$, i.e.:

$$g(x) = S(x, f(R(x))), \forall x \in \Sigma^*$$

- The solution conting versions of all known **NP**-complete problems are **#P**-complete!

Basic Definitions

Definition

A *Randomized Approximation Scheme* (*RAS*) for a function $f : \Sigma^* \to \mathbb{N}$ is a Probabilistic Turing Machine that takes as input a pair $(x, \varepsilon) \in \Sigma^* \times (0, 1)$ and produces as output an integer random variable Y satisfying the condition:

$$\mathbf{Pr}\left[e^{-\varepsilon} f(x) \leq Y \leq e^{\varepsilon} f(x)\right] \geq \frac{3}{4}$$

A RAS is said to be *fully polynomial* (*FPRAS*) if it runs in time $poly(|x|, \varepsilon^{-1})$.

Counting DNF Solutions

Counting DNF solutions

Let:

$$f : C_1 \vee C_2 \vee \cdots \vee C_m$$

Where $C_i = l_1 \wedge l_2 \wedge \cdots \wedge l_{r_i}$, and l_j is a literal. We assume that each clause is satisfiable.

We want to compute $\#f =$"*the number of satisfying truth assignments of f*".

Counting DNF Solutions

- The idea is to define a r.v. X s.t. $\mathbf{E}[X] = \#f$ (unbiased estimator).
- Let S_i the set of t.a. to $x_1, \dots, x_n$ that satisfy C_i.
- $|S_i| = 2^{n-r_i}$ and $\#f = |\cup_{i=1}^{m} S_i|$.
- Let $c(\tau)$ the number of clauses t.a. τ satisfies.
- Let M be the *multiset* union of S_i's⇒It contains each satisfying t.a. τ, $c(\tau)$ times!
- Pick a satisfying t.a. τ for f with probability $c(\tau)/|M|$.
- Define

$$X(\tau) = \frac{|M|}{c(\tau)}$$

- X can be efficiently sampled:

Counting DNF Solutions

Lemma 1

Random Variable X can be efficiently sampled.

Proof:

- Pick clause: $\textbf{Pr}[\textit{Picking Clause } C_i] = |S_i|/|M|$
- Among the t.a. satisfying the picked clause, choose one at random.
- The probability with which τ is picked is:

$$\sum_{i:\tau \text{ satisfies } C_i} \frac{|S_i|}{|M|} \times \frac{1}{|S_i|} = \frac{c(\tau)}{|M|}$$

□

Counting DNF Solutions

Lemma 2

X is an unbiased estimator for $\#f$.

Proof:

$$\mathbf{E}[X] = \sum_{\tau} \mathbf{Pr}[\tau \ is \ picked] \cdot X(\tau) = \sum_{\tau \ satisfies \ f} \frac{c(\tau)}{|M|} \times \frac{|M|}{c(\tau)} = \#f$$

□

Counting DNF Solutions

Lemma 3

If m denotes the number of clauses in f, then:

$$\frac{\sigma(X)}{\mathbf{E}[X]} \leq m - 1$$

Proof:

- Let $\alpha = |M|/m$. Clearly, $\mathbf{E}[X] \geq \alpha$ (1).
- For each satisfying t.a. τ of f: $1 \leq c(\tau) \leq m$. So, $X(\tau) \in [\alpha, m\alpha]$ and $|X(\tau) - \mathbf{E}[X(\tau)]| \leq (m-1)\alpha$.
- So, $\sigma(X) \leq (m-1)\alpha$ (2).
- (1) & (2) prove the lemma!

$\square$

Counting DNF Solutions

Lemma 4

For any $\varepsilon > 0$,

$$\mathbf{Pr}\left[|X_k - \#f| \leq \varepsilon \#f\right] \geq \frac{3}{4}$$

where $k = 4(m-1)^2/\varepsilon^2$.

Proof:

$$\mathbf{Pr}\left[|X_k - \mathbf{E}\left[X_k\right]| \geq \varepsilon \cdot \mathbf{E}\left[X_k\right]\right] \leq \left(\frac{\sigma\left(X_k\right)}{\varepsilon \cdot \mathbf{E}\left[X_k\right]}\right)^2 = \left(\frac{\sigma\left(X\right)}{\varepsilon \sqrt{k} \mathbf{E}\left[X\right]}\right)^2 \leq \frac{1}{4}$$

So finally,

Theorem

The is an FPRAS for the problem of counting DNF solutions!

Basic Definitions

Definition

An *approximation-preserving* reduction from f to g is a probabilistic oracle Turing Machine M that takes as input a pair $(x, \varepsilon) \in \Sigma^* \times (0, 1)$, and satisfies the following conditions:

1. Every oracle call made by M is of the form (w, δ), where w is an instance of g, and $\delta \in (0, 1)$ is an error bound satisfying $\delta^{-1} \leq poly(|x|, \varepsilon^{-1})$.
2. M is a RAS for f whenever its oracle is a RAS for g.
3. M runs in $poly(|x|, \varepsilon^{-1})$.

If such a reduction form f to g exists, we write $f \leq_{AP} g$ (*AP-reducible*).
If $(f \leq_{AP} g) \wedge (g \leq_{AP} f)$, we write $f \equiv_{AP} g$ (*AP-interreducible*).

Essential Counting Problems

#SAT Definition

Instance: A Boolean formula ϕ in CNF.

Output: The number of satisfying assignments to ϕ.

#BIS Definition

Instance: A bipartite graph B.

Output: The number of indepedent sets in B.

Three classes of AP-interreducible problems:

1. The class of counting problems that admit an *FPRAS*.
2. The class of counting problems AP-interreducible with #SAT.
3. The class of counting problems AP-interreducible with #BIS.

Counting Problems that admit an *FPRAS*

- Problems that admit an *FPRAS* despite being **#P**-Complete!

#MATCH Definition

Instance: A Graph G.

Output: The number of matchings (of all sizes) in G.

#DNF Definition

Instance: A Boolean formula ϕ in DNF.

Output: The number of satisfying assignments to ϕ.

Counting problems AP-interreducible with #SAT

Definition

Suppose $f, g : \Sigma^* \to \mathbb{N}$. A *parsimonious reduction* from f to g is a function $p : \Sigma^* \to \Sigma^*$ satisfying:

1. $f(w) = g(p(w)), \forall w \in \Sigma^*$
2. p is computable by a polynomial-time deterministic TM

- Parsimonious reduction preserve the number of solutions.
- A parsimonious reduction is a special instance of an AP-reduction.
- $\#\texttt{SAT}$ is $\#\mathbf{P}$-complete with respect to *AP-reducibility*.
- Zuckerman (1996) proved that there is no *FPRAS* for $\#\texttt{SAT}$ unless $\mathbf{NP} = \mathbf{RP}$.

Counting problems AP-interreducible with #SAT

Definition (Counting Versions of NP-Complete Problems)

If $A : \Sigma^* \rightarrow \{0, 1\}$ some decision problem in **NP**.
It is known that:

$$A(x) = 1 \Leftrightarrow (\exists y, |y| = p(|x|) : R(x, y) = 1)$$

for a polynomial-time computable predicate R.
The **counting problem** $\#A : \Sigma^* \rightarrow \mathbb{N}$, corresponding to A, is defined by:

$$\#A(x) = |\{y : |y| = p(|x|) \wedge R(x, y)\}|$$

Counting problems AP-interreducible with #SAT

Theorem

Let A be an **NP***-complete decision problem. Then, the corresponding counting problem* $\#A$ *is* **#P***-complete with respect to AP-reducibility.*

Proof:

- $\#A \in \#P$
- Also, #SAT is AP-reducible to $\#A$: #SAT can be approximated by PTM M equipped with an oracle for the *decision* problem of SAT.
- This oracle can be replaced by an *approximate counting oracle* (RAS) for $\#A$.
- Thus, M consists an approximation-preserving reduction from #SAT to $\#A$. □

Counting problems AP-interreducible with #SAT

#LARGEIS Definition

Instance: A positive integer m and a graph G in which *every* indepedent set has size at most m.
Output: The number of size-m indepedent sets in G.

Corollary

$\#LARGEIS \equiv_{AP} \#SAT$

#IS Definition

Instance: A graph G.
Output: The number of indepedent sets (of all sizes) in G.

Theorem

$\#IS \equiv_{AP} \#SAT$

Counting problems AP-interreducible with #BIS

$\#P_4$-COL Definition

Instance: A graph G.
Output: The number of P_4 colourings of G, where P_4 is the path of length 3.

#DOWNSETS Definition

Instance: A partially ordered set $(X, \preceq)$.
Output: The number of downsets in $(X, \preceq)$.

#1P1NSAT Definition

Instance: A CNF Boolean formula ϕ, with at most one unnegated literal per clause, and at most one negated literal.
Output: The number of satisfying assignments to ϕ.

Counting problems AP-interreducible with #BIS

#BEACHCONFIGS Definition

Instance: A graph G.
Output: The number of *Beach Configurations* in G (P_4^* colourings of G, where P_4^* is the path of length 3 with loops on all four vertices).

Theorem

The problems #BIS, $\#P_4$-COL, #DOWNSETS, #1P1NSAT, #BEACHCONFIGS are all AP-interreducible.

- Very easily:
  #BIS $\equiv_{AP}$ $\#P_4$-COL
  #DOWNSETS $\equiv_{AP}$ #1P1NSAT
- We can also show the reduction:
  #BIS $\leq_{AP}$ #BEACHCONFIGS $\leq_{AP}$ #DOWNSETS $\leq_{AP}$ #BIS

Counting problems AP-interreducible with #BIS

Lemma

$\#BIS \equiv_{AP} \#P_4\text{-}COL$

Proof:

These problems are essentially the same:

A graph G is P_4-colourable $\Leftrightarrow$ is Bipartite

Two of the colours point out the IS!

Conversely, an IS in a (connected) bipartite graph arises from one of the two P_4 colourings!

□

Counting problems AP-interreducible with #BIS

Lemma

#DOWNSETS $\equiv_{AP}$ *#1P1NSAT*

Proof:

The first is a *restricted* case of the second, in which:

1. All clauses have two literals $(x \Rightarrow y)$
2. There are **no** cyclic chains of implications:
   $x_0 \Rightarrow x_1 \Rightarrow \cdots \Rightarrow x_{\ell-1} \Rightarrow x_0$.

-Given an instance of #1P1NSAT, any forced variables (1) may be removed by substituting TRUE or FALSE.
-Any set of ℓ variables forming a cycle (2) may be replaced by a single one.

□

A Logical Characterisation #BIS and its relatives

- A counting problem is identified with a sentence ϕ in FO Logic, the objects being counted with models of ϕ.
- Standard Definitions:
  - *Vocabulary*: $\sigma = \{\widetilde{R}_0, \ldots \widetilde{R}_{k-1}\}$
  - $\widetilde{R}_i$'s are relation symbol of arities $r_0, \ldots, r_{k-1}$
  - *Structure* $\mathbf{A} = (A, R_0, \ldots, R_{k-1})$ over σ consists a universe A
  - Each relation $R_i \subseteq A^{r_i}$ is an interpretation of $\widetilde{R}_i$.
- We present counting problems as *structures* over suitable vocabularies:

Example

An instance of #IS is a graph which can regarded as a structure $\mathbf{A} = (A, \sim)$, where A is the vertex set, and "$\sim$" is the symmetric binary relation of adjacency.

A Logical Characterisation #BIS and its relatives

- The objects to be counted are represented as sequences of relations $\mathbf{T} = (T_1, \ldots, T_{r-1})$ and first-order variables $\mathbf{z} = (z_0, \ldots, z_{m-1}$.

Definition

A counting problem f (from structures over σ to $\mathbb{N}$) is in the class $\#\mathcal{FO}$ if it can be expressed as:

$$f(\mathbf{A}) = |\{(\mathbf{T}, \mathbf{z}) : \mathbf{A} \models \phi(\mathbf{z}, \mathbf{T})\}|$$

where ϕ is a FO formula with relation symbols from $\sigma \cup \mathbf{T}$ and (free) variables from $\mathbf{z}$.

A Logical Characterisation #BIS and its relatives

Example

If we encode an IS as a unary relation I, then #IS:

$$f_{IS}(\mathbf{A}) = |\{(I) : \mathbf{A} \models \forall x, y : x \sim y \Rightarrow \neg I(x) \vee \neg I(y)\}|$$

- #IS is in the subclass $\#\Pi_1 \subseteq \#\mathcal{FO}$ (since the formula contains only universal quantification).
- In general, we have a (strict) hierarchy of subclasses:

$$\#\Sigma_0 = \#\Pi_0 \subset \#\Sigma_1 \subset \#\Pi_1 \subset \#\Sigma_2 \subset \#\Pi_2 = \#\mathcal{FO} = \#\mathbf{P}$$

- All functions in $\#\Sigma_1$ admit an *FPRAS*!
- All AP-interreducible problems we saw are in the (*syntactically* restricted) subclass $\#RH\Pi_1 \subseteq \#\Pi_1$:

A Logical Characterisation #BIS and its relatives

Definition

A counting problem f is in the class $\#RH\Pi_1$ if it can be expressed in the form:

$$f(\mathbf{A}) = |\{(\mathbf{T}, \mathbf{z}) : \mathbf{A} \models \forall \mathbf{y} : \psi(\mathbf{y}, \mathbf{z}, \mathbf{T})\}|$$

where ψ is an *unquantified* CNF formula in which each clause has at most one occurence of an unnegated relation symbol from **T**, and at most one occurence of a negated relation symbol from **T**.

- "RH" stands for "Restricted Horn"
- The restriction on clauses of ψ applies only to terms involving symbols from **T**.

A Logical Characterisation #BIS and its relatives

Example

An instance of #DOWNSETS can be expressed as a structure $\mathbf{A} = (A, \preceq)$.

Then, #DOWNSETS $\in \#RH\Pi_1$, since the number of downsets may be expressed as:

$$f_{DS}(\mathbf{A}) = |\{(D) : \mathbf{A} \models \forall x, y \in A : D(x) \wedge (y \preceq x) \Rightarrow D(y)\}|$$

Theorem

The problems #BIS, $\#P_4$-COL, #DOWNSETS, #1P1NSAT, #BEACHCONFIGS are all complete for $\#RH\Pi_1$, with respect to AP-reducibility!

References

- *The presentation is based on:*
  Martin E. Dyer, Leslie Ann Goldberg, Catherine S. Greenhill, Mark Jerrum: **The Relative Complexity of Approximate Counting Problems**, Algorithmica 38(3): 471-500 (2003)
- *Also used:*
  Approximation Algorithms, V.V.Vazirani, Springer 2001

Thank You!