Heuristics, Closing Thoughts

Lecture 26
May 1, 2014

Part I

Heuristics

Coping with Intractability

Question: Many useful/important problems are **NP-Hard** or worse. How does one cope with them?

Coping with Intractability

Question: Many useful/important problems are **NP-Hard** or worse. How does one cope with them?

Some general things that people do.

1. Consider special cases of the problem which may be tractable.
2. Run inefficient algorithms (for example exponential time algorithms for **NP-Hard** problems) augmented with (very) clever heuristics
   1. stop algorithm when time/resources run out
   2. use massive computational power
3. Exploit properties of instances that arise in practice which may be much easier. Give up on hard instances, which is OK.
4. Settle for sub-optimal (aka approximate) solutions, especially for optimization problems

NP and EXP

EXP: all problems that have an exponential time algorithm.

Proposition

NP $\subseteq$ **EXP**.

NP and EXP

EXP: all problems that have an exponential time algorithm.

Proposition

NP $\subseteq$ EXP.

Proof.

Let $\mathbf{X} \in \mathbf{NP}$ with certifier $\mathbf{C}$. To prove $\mathbf{X} \in \mathbf{EXP}$, here is an algorithm for $\mathbf{X}$. Given input $\mathbf{s}$,

1. For every $\mathbf{t}$, with $|\mathbf{t}| \leq \mathbf{p(|s|)}$ run $\mathbf{C(s, t)}$; answer "yes" if any one of these calls returns "yes", otherwise say "no". $\qquad \square$

Every problem in **NP** has a brute-force "try all possibilities" algorithm that runs in exponential time.

Examples

1. **SAT**: try all possible truth assignment to variables.
2. **Independent set**: try all possible subsets of vertices.
3. **Vertex cover**: try all possible subsets of vertices.

Improving brute-force via intelligent backtracking

1. Backtrack search: enumeration with bells and whistles to "heuristically" cut down search space.
2. Works quite well in practice for several problems, especially for small enough problem sizes.

Backtrack Search Algorithm for SAT

Input: CNF Formula φ on **n** variables $\mathbf{x_1}, \ldots, \mathbf{x_n}$ and **m** clauses
Output: Is φ satisfiable or not.

1. Pick a variable $\mathbf{x_i}$
2. φ' is CNF formula obtained by setting $\mathbf{x_i = 0}$ and simplifying
3. Run a simple (heuristic) check on φ': returns "yes", "no" or "not sure"
   1. If "not sure" recursively solve φ'
   2. If φ' is satisfiable, return "yes"
4. φ'' is CNF formula obtained by setting $\mathbf{x_i = 1}$
5. Run simple check on φ'': returns "yes", "no" or "not sure"
   1. If "not sure" recursively solve φ''
   2. If φ'' is satisfiable, return "yes"
6. Return "no"

Certain part of the search space is pruned.

Example
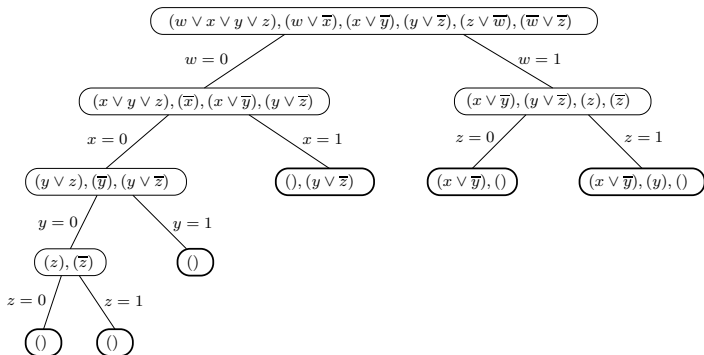

Figure : Backtrack search. Formula is not satisfiable.

Figure taken from Dasgupta etal book.

How do we pick the order of variables?

Backtrack Search Algorithm for SAT

How do we pick the order of variables? Heuristically! Examples:

1. pick variable that occurs in most clauses first
2. pick variable that appears in most size 2 clauses first
3. ...

Backtrack Search Algorithm for SAT

How do we pick the order of variables? Heuristically! Examples:

1. pick variable that occurs in most clauses first
2. pick variable that appears in most size 2 clauses first
3. ...

What are quick tests for Satisfiability?

Backtrack Search Algorithm for SAT

How do we pick the order of variables? Heuristically! Examples:

1. pick variable that occurs in most clauses first
2. pick variable that appears in most size 2 clauses first
3. ...

What are quick tests for Satisfiability?
Depends on known special cases and heuristics. Examples.

1. Obvious test: return "no" if empty clause, "yes" if no clauses left and otherwise "not sure"
2. Run obvious test and in addition if all clauses are of size **2** then run 2-SAT polynomial time algorithm
3. ...

Branch-and-Bound

Intelligent backtracking can be used also for optimization problems.
Consider a minimization problem.
Notation: for instance **I**, **opt(I)** is optimum value on **I**.

P_0 initial instance of given problem.

1. Keep track of the best solution value **B** found so far. Initialize **B** to be crude upper bound on **opt(I)**.
2. Let **P** be a subproblem at some stage of exploration.
3. If **P** is a complete solution, update **B**.
4. Else use a lower bounding heuristic to quickly/efficiently find a lower bound **b** on **opt(P)**.
   1. If **b** $\geq$ **B** then prune **P**
   2. Else explore **P** further by breaking it into subproblems and recurse on them.
5. Output best solution found.

Example: Vertex Cover

Given $G = (V, E)$, find a minimum sized vertex cover in G.

1. Initialize $B = n - 1$.
2. Pick a vertex u. Branch on u: either choose u or discard it.
3. Let b_1 be a lower bound on $G_1 = G - u$.
4. If $1 + b_1 < B$, recursively explore G_1
5. Let b_2 be a lower bound on $G_2 = G - u - N(u)$ where $N(u)$ is the set of neighbors of u.
6. If $|N(u)| + b_2 < B$, recursively explore G_2
7. Output B.

Example: Vertex Cover

Given $G = (V, E)$, find a minimum sized vertex cover in G.

1. Initialize $B = n - 1$.
2. Pick a vertex u. Branch on u: either choose u or discard it.
3. Let b_1 be a lower bound on $G_1 = G - u$.
4. If $1 + b_1 < B$, recursively explore G_1
5. Let b_2 be a lower bound on $G_2 = G - u - N(u)$ where $N(u)$ is the set of neighbors of u.
6. If $|N(u)| + b_2 < B$, recursively explore G_2
7. Output B.

How do we compute a lower bound?
One possibility: solve an LP relaxation.

Local Search

Local Search: a simple and broadly applicable heuristic method

1. Start with some arbitrary solution **s**
2. Let **N(s)** be solutions in the "neighborhood" of **s** obtained from **s** via "local" moves/changes
3. If there is a solution **s'** ∈ **N(s)** that is better than **s**, move to **s'** and continue search with **s'**
4. Else, stop search and output **s**.

Local Search

Main ingredients in local search:

1. Initial solution.
2. Definition of neighborhood of a solution.
3. Efficient algorithm to find a good solution in the neighborhood.

TSP: Given a complete graph $G = (V, E)$ with c_{ij} denoting cost of edge (i, j), compute a Hamiltonian cycle/tour of minimum edge cost.

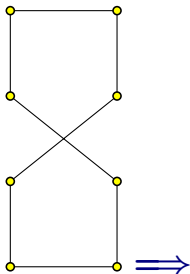# Example: TSP

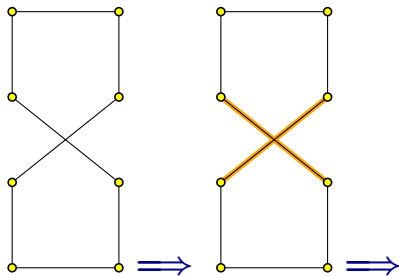TSP: Given a complete graph $G = (V, E)$ with c_{ij} denoting cost of edge (i, j), compute a Hamiltonian cycle/tour of minimum edge cost.

2-change local search:

1. Start with an arbitrary tour s_0
2. For a solution s define s' to be a neighbor if s' can be obtained from s by replacing two edges in s with two other edges.
3. For a solution s at most $O(n^2)$ neighbors and one can try all of them to find an improvement.
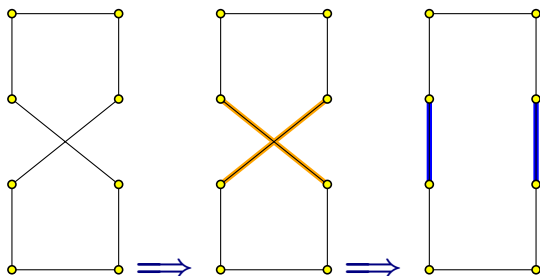
TSP: 2-change example

Figure below shows a bad local optimum for **2**-change heuristic...

TSP: 2-change example

Figure below shows a bad local optimum for **2**-change heuristic...

TSP: 2-change example

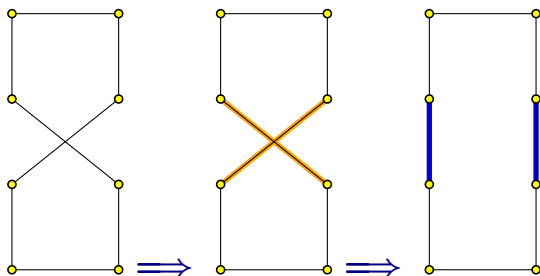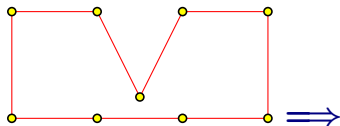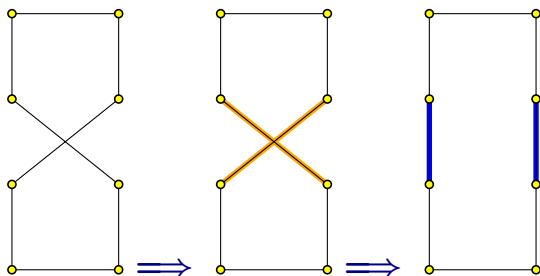Figure below shows a bad local optimum for **2**-change heuristic...

TSP: 3-change example

3-change local search: swap **3** edges out.

Neighborhood of **s** has now increased to a size of $\mathbf{\Omega(n^3)}$
Can define **k**-change heuristic where **k** edges are swapped out.
Increases neighborhood size and makes each local improvement step
less efficient.

TSP: 3-change example

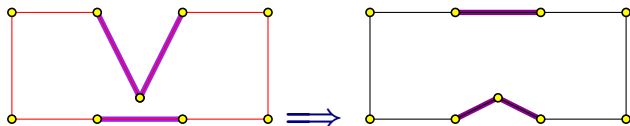3-change local search: swap **3** edges out.

Neighborhood of **s** has now increased to a size of $\Omega(n^3)$
Can define **k**-change heuristic where **k** edges are swapped out.
Increases neighborhood size and makes each local improvement step
less efficient.

Local Search Variants

Local search terminates with a local optimum which may be far from a global optimum. Many variants to improve plain local search.

1. **Randomization and restarts**. Initial solution may strongly influence the quality of the final solution. Try many random initial solutions.

2. **Simulated annealing** is a general method where one allows the algorithm to move to worse solutions with some probability. At the beginning this is done more aggressively and then slowly the algorithm converges to plain local search. Controlled by a parameter called "temperature".

3. **Tabu search**. Store already visited solutions and do not visit them again (they are "taboo").

Heuristics

Several other heuristics used in practice.

1. Heuristics for solving integer linear programs such as cutting planes, branch-and-cut etc are quite effective.
2. Heuristics to solve SAT (SAT-solvers) have gained prominence in recent years
3. Genetic algorithms
4. ...

Heuristics design is somewhat ad hoc and depends heavily on the problem and the instances that are of interest. Rigorous analysis is sometimes possible.

Approximation algorithms

Consider the following *optimization* problems:

1. **Max Knapsack**: Given knapsack of capacity **W**, **n** items each with a value and weight, pack the knapsack with the most profitable subset of items whose weight does not exceed the knapsack capacity.

2. **Min Vertex Cover:** given a graph $G = (V, E)$ find the minimum cardinality vertex cover.

3. **Min Set Cover:** given Set Cover instance, find the smallest number of sets that cover all elements in the universe.

4. **Max Independent Set:** given graph $G = (V, E)$ find maximum independent set.

5. **Min Traveling Salesman Tour:** given a directed graph **G** with edge costs, find minimum length/cost Hamiltonian cycle in **G**.

Solving one in polynomial time implies solving all the others.

Approximation algorithms

However, the problems behave very differently if one wants to solve them *approximately*.

Informal definition: An approximation algorithm for an optimization problem is an efficient (polynomial-time) algorithm that *guarantees* for every instance a solution of some given quality when compared to an optimal solution.

Some known approximation results

1. **Knapsack**: For every fixed $\epsilon > 0$ there is a polynomial time algorithm that guarantees a solution of quality $(1 - \epsilon)$ times the best solution for the given instance. Hence can get a **0.99**-approximation efficiently.

2. **Min Vertex Cover**: There is a polynomial time algorithm that guarantees a solution of cost at most **2** times the cost of an optimum solution.

3. **Min Set Cover**: There is a polynomial time algorithm that guarantees a solution of cost at most $(\ln n + 1)$ times the cost of an optimal solution.

4. **Max Independent Set**: Unless $\mathbf{P = NP}$, for any fixed $\epsilon > 0$, no polynomial time algorithm can give a $n^{1-\epsilon}$ relative approximation . Here n is number of vertices in the graph.

5. **Min TSP**: No polynomial factor relative approximation possible.

Approximation algorithms

1. Although **NP-Complete** problems are all equivalent with respect to polynomial-time solvability they behave quite differently under approximation (in both theory and practice).

2. Approximation is a useful lens to examine **NP-Complete** problems more closely.

3. Approximation also useful for problems that we can solve efficiently:

    1. We may have other constraints such a space (streaming problems) or time (need linear time or less for very large problems)

    2. Data may be uncertain (online and stochastic problems).

Example: Vertex Cover

```
Greedy(G):
        Initialize S to be ∅
        While there are edges in G do
            Let v be a vertex with maximum degree
            S ← S ∪ {v}
            G ← G − v
        endWhile
        Output S
```

Example: Vertex Cover

```
Greedy(G):
      Initialize S to be ∅
      While there are edges in G do
          Let v be a vertex with maximum degree
          S ← S ∪ {v}
          G ← G − v
      endWhile
      Output S
```

Theorem

$|S| \leq (\ln n + 1)\textbf{OPT}$ *where* **OPT** *is the value of an optimum set. Here* **n** *is number of nodes in* **G**.

Theorem

There is an infinite family of graphs where the solution **S** *output by Greedy is* $\Omega(\ln n)\textbf{OPT}$.

MatchingHeuristic(**G**):
```
        Find a maximal matching M in G
        S is the set of end points of edges in M
        Output S
```

Example: Vertex Cover

```
MatchingHeuristic(G):
        Find a maximal matching M in G
        S is the set of end points of edges in M
        Output S
```

Lemma

$OPT \geq |M|$.

Lemma

S *is a feasible vertex cover.*

Analysis: $|S| = 2|M| \leq 2OPT$. Algorithm is a **2**-approximation.

Vertex Cover: LP Relaxation based approach

Write (weighted) vertex cover problem as an integer linear program

$$
\begin{array}{ll}
\text{Minimize} & \sum_{v \in V} w_v x_v \\
\text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\
& x_v \in \{0, 1\} \quad \text{for each } v \in V
\end{array}
$$

Vertex Cover: LP Relaxation based approach

Write (weighted) vertex cover problem as an integer linear program

$$
\begin{array}{ll}
\text{Minimize} & \sum_{v \in V} w_v x_v \\
\text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\
& x_v \in \{0, 1\} \quad \text{for each } v \in V
\end{array}
$$

Relax integer program to a linear program

$$
\begin{array}{ll}
\text{Minimize} & \sum_{v \in V} w_v x_v \\
\text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\
& x_v \geq 0 \quad \text{for each } v \in V
\end{array}
$$

Vertex Cover: LP Relaxation based approach

Write (weighted) vertex cover problem as an integer linear program

$$
\begin{array}{lll}
\text{Minimize} & \sum_{v \in V} w_v x_v & \\
\text{subject to} & x_u + x_v \geq 1 & \text{for each } uv \in E \\
& x_v \in \{0, 1\} & \text{for each } v \in V
\end{array}
$$

Relax integer program to a linear program

$$
\begin{array}{lll}
\text{Minimize} & \sum_{v \in V} w_v x_v & \\
\text{subject to} & x_u + x_v \geq 1 & \text{for each } uv \in E \\
& x_v \geq 0 & \text{for each } v \in V
\end{array}
$$

Can solve linear program in polynomial time.
Let x^* be an optimum solution to the linear program.

Lemma

$OPT \geq \sum_v w_v x_v^*$.

Vertex Cover: Rounding fractional solution

LP Relaxation

$$\begin{array}{ll}
\text{Minimize} & \sum_{v \in V} w_v x_v \\
\text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\
& x_v \geq 0 \qquad \text{for each } v \in V
\end{array}$$

Let x^* be an optimum solution to the linear program.
Rounding: $S = \{v \mid x_v^* \geq 1/2\}$. Output S.

Vertex Cover: Rounding fractional solution

LP Relaxation

$$\begin{array}{ll}
\text{Minimize} & \sum_{v \in V} w_v x_v \\
\text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\
& x_v \geq 0 \qquad\quad \text{for each } v \in V
\end{array}$$

Let x^* be an optimum solution to the linear program.
Rounding: $S = \{v \mid x_v^* \geq 1/2\}$. Output S.

Lemma

S is a feasible vertex cover for the given graph.

Vertex Cover: Rounding fractional solution

LP Relaxation

$$\begin{array}{ll}\text{Minimize} & \sum_{v \in V} w_v x_v \\ \text{subject to} & x_u + x_v \geq 1 \quad \text{for each } uv \in E \\ & x_v \geq 0 \quad\quad \text{for each } v \in V \end{array}$$

Let x^* be an optimum solution to the linear program.
Rounding: $S = \{v \mid x_v^* \geq 1/2\}$. Output S.

Lemma
S *is a feasible vertex cover for the given graph.*

Lemma
$w(S) \leq 2 \sum_v w_v x_v^* \leq 2OPT$.

Set Cover and Vertex Cover

Theorem

Greedy gives $(\ln n + 1)$-approximation for Set Cover where n is number of elements.

Theorem

Unless $\mathbf{P} = \mathbf{NP}$ no $(\ln n + \epsilon)$-approximation for Set Cover.

2-approximation is best known for Vertex Cover.

Theorem

Unless $\mathbf{P} = \mathbf{NP}$ no $\mathbf{1.36}$-approximation for Vertex Cover.

Conjecture: Unless $\mathbf{P} = \mathbf{NP}$ no $(2 - \epsilon)$-approximation for Vertex Cover for any fixed $\epsilon > 0$.

Part II

Closing Thoughts

Topics I wish I had time for

1. Data structures
2. Lower bounds on sorting and related problems
3. More on the use of randomization in algorithms
4. Algorithms for string processing, compression, coding
5. More on heuristics and applications

Theoretical Computer Science

1. Algorithms: find efficient ways to solve particular problems or broad category of problems
2. Computational Complexity: understand nature of computation — classification of problems into classes (**P**, **NP**, **co-NP**) and their relationships, limits of computation.
3. Logic, Languages and Formal Methods

Form the foundations for computer "science"

The Computational Lens

The Algorithm: Idiom of Modern Science by Bernard Chazelle
http://www.cs.princeton.edu/chazelle/pubs/algorithm.html

Computation has gained ground as *fundamental* artifact in mathematics and science.

1. nature of proofs, **P** vs **NP**, complexity, ...
2. quantum computation and information
3. computational biology and the biological processes, ...

Standard question in math and sciences: Is there a *solution/algorithm*?

The Computational Lens

The Algorithm: Idiom of Modern Science by Bernard Chazelle
http://www.cs.princeton.edu/chazelle/pubs/algorithm.html

Computation has gained ground as *fundamental* artifact in mathematics and science.

1. nature of proofs, **P** vs **NP**, complexity, ...
2. quantum computation and information
3. computational biology and the biological processes, ...

Standard question in math and sciences: Is there a *solution/algorithm*?
New: Is there an *efficient* solution/algorithm?

Related theory courses

1. Graduate algorithms (every year, offered this fall)
2. Computational complexity (every year, offered next spring)
3. Randomized algorithms (every other year)
4. Approximation algorithms (every other year)
5. Advanced data structures (every once in a while)
6. Cryptography and related topics (almost every year, offered this fall)
7. Algorithmic game theory, combinatorial optimization, computational geometry, logic and formal methods, coding theory, information theory, graph theory, combinatorics, ...

Technological and sociological changes

- Scale of data due to the web, digitization of knowledge, people
- Mobility, cloud, and distributed data
- Social and economic networks/activity in online/virtual fora
- Privacy and security
- Possibility of quantum computation: new model and algorithms
- Biology/DNA as digital information/computation
- End of Moore's law?
- Increasing integration of sensing and computational devices

Impact on Algorithms

- Scale of data implies efficiency of algorithms very important
  - Sub-linear time and data streaming (approximation) algorithms via sampling, estimation, ...
  - Various forms of parallel algorithms (Map-Reduce etc)
- Large data availability implies statistical machine learning can be very effective so many new applications
- Algorithmic aspects of strategic behavior of people: algorithmic game theory, sponsored search auctions, influence propagation in social networks, ...,
- Distributed computing, cyber-physical computing
- Interdisciplinary work: quatum information and computation, computational biology and chemistry

And...

Questions?

And...

Questions?

Final Exam: Friday May 9th, 7–10pm

Thanks!

Notes

Notes