P versus NP

Matt Valeriote

McMaster University

23 January, 2008

Propositional Formulas

Definition

Propositional variables are variables that are allowed to take on the values True (T) or False (F) and can be combined using logical connectives ($\wedge$, $\vee$, $\neg$) to produce complicated statements (propositional formulas).

Example

$$\Phi = x_1 \wedge (x_3 \vee x_4) \wedge \neg x_4 \wedge (\neg x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_3 \vee x_4)$$

Problem:

Is there a way to assign truth values to the propositional variables x_1, x_2, x_3, and x_4 so that the formula Φ is true? In other words, is the formula Φ satisfiable?

Satisfiability of
$x_1 \wedge (x_3 \vee x_4) \wedge \neg x_4 \wedge (\neg x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_3 \vee x_4)$

One attempt:

If we assign the value T to the variables x_1 and x_3 and F to x_2 and x_4 then the truth value of Φ is F, or False, since the conjunct $(\neg x_1 \vee x_2)$ is False under this assignment.

Truth Tables

- One way to settle this, and similar problems, is to try all possible truth assignments to see if any satisfy Φ.
- Since Φ has four propositional variables, and each variable can take on one of two possible values, then there are $2^4 = 16$ different truth assignments for the variables of Φ.
- In general, if a propositional formula Θ has n propositional variables, then there will be 2^n possible truth assignments to try out.

Φ is not satisfiable

Φ's truth table

- The following table contains all 16 possible truth assignments that can be made for the variables of Φ, along with the truth value that Φ has for each assignment:

x_1	x_2	x_3	x_4	Φ
F	F	F	F	F
F	F	F	T	F
F	F	T	F	F
F	F	T	T	F
F	T	F	F	F
F	T	F	T	F
F	T	T	F	F
F	T	T	T	F

x_1	x_2	x_3	x_4	Φ
T	F	F	F	F
T	F	F	T	F
T	F	T	F	F
T	F	T	T	F
T	T	F	F	F
T	T	F	T	F
T	T	T	F	F
T	T	T	T	F

- We can quickly see that no assignment satisfies Φ and thus, Φ is not satisfiable.

SAT

Definition

The satisfiability problem (SAT) is the problem of determining, given a propositional formula Θ, whether there is some truth assignment for the propositional variables in Θ that satisfies Θ.

Some features of SAT

- The truth table method can be employed to solve any instance of SAT and so, in principle, SAT is a solvable (decidable) problem.
- This method is very inefficient, since the time required to use it grows exponentially in the "size" of the formula.
- It can be quickly checked if a given truth assignment satisfies a formula.
- No efficient method is known to solve SAT

A Meta-Problem

Is there an efficient procedure to solve all instances of SAT?

A Similar Sort of Problem

A Housing Problem

- Suppose that you are organizing housing accommodations for a group of four hundred university students.
- Only one hundred of the students will receive places in the dormitory.
- The Dean has provided you with a list of pairs of incompatible students, and requested that no pair from this list appear in your final choice.

Note

- It is easy to check if a given choice of one hundred students proposed by a coworker is satisfactory, however,
- The task of generating such a list from scratch seems to be hard.
- The total number of possibilities is greater than the number of atoms in the known universe.
- Thus no future civilization could ever hope to build a supercomputer capable of solving the problem by brute force.

Your Exam Schedule

A Recent Email from the Registrar

. . . using the flexibility provided by the new software, the university has been able to make significant improvements in the quality of the examination schedule. Almost all of the student, instructor and room constraints have been satisfied . . . .

Student Criteria	Dec. 2005	Dec. 2006
Number of students with 2 exams at the same time	40	0
Number of students with 2 consecutive exams	1951	315
Number of students with 3 consecutive exams	48	38
Number of students with 3 exams on same day	16	5

A Common Feature

Note

These problems have the following common feature:

It is easy and quick to check whether a proposed solution is correct.

Definition

Call a problem (or class of problems) Quickly Checkable if there is a "fast" method for checking whether a proposed solution of the problem is correct.

Note

More formally, problems with this property are said to be solvable in **Non-Deterministic Polynomial Time** (or to lie in NP).

Computational Complexity

A Measure of Complexity

- A way to measure the complexity of a problem is by determining the number of "steps" required to solve it.
- If the steps in question are purely mechanical, then the complexity can also be measured by the amount of time or the number of steps needed to solve it on a computer.
- If the number of steps required to solve instances of a problem can be bounded by some polynomial in the size of the instance, then the problem is said to be solvable in Polynomial Time, or to lie in P.

Examples

- To multiply two n-digit numbers takes approximately n^2-steps.
- Differentiating the nth degree polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ can be done it at most $n(\log_{10}(n)\log_{10}(M))$ steps, where M is the largest coefficient of $p(x)$.

The Travelling Salesman Problem

The Travelling Salesman Problem

Problem

A salesman is required to visit n cities and needs to find the shortest path that passes through each city exactly once and then ends up where it started. This problem is known as the Travelling Salesman Problem.

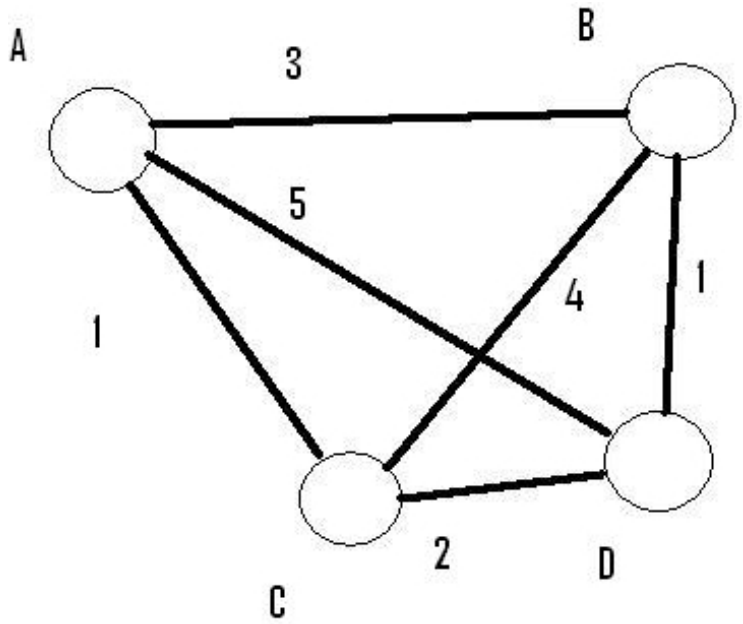
Example

The contest on the previous page is an example of the travelling salesman problem. It involves travelling between 33 cities and was solved by several people in 1962.

The Travelling Salesman Problem

A Simpler Example

Find the shortest path through the four cities labelled A, B, C, and D where the distances between each city are given in the table below.



	A	*B*	*C*	*D*
A	0	3	1	5
B		0	4	1
C			0	2
D				0

Answer: Length 7, through A, B, D, C, A in that order.

How much time does it take to solve the TSP?

Note

One way to solve the Travelling Salesman Problem for a route having n cities is to try every possible path.

How many paths are there through n cities?

A standard counting argument shows that there are

$$[(n-1)\cdot(n-2)\cdot(n-3)\cdots 3\cdot 2\cdot 1]/2 = \frac{(n-1)!}{2}$$

different paths to try.

How much time does it take to solve the TSP?

The following table shows, for some values of n, the number of possible paths through n cities, and the number of seconds and years it would take the world's fastest computer to check every path.

n	$\frac{(n-1)!}{2}$	number of seconds	number of years
4	3	1	0
10	181440	1	0
20	10^{17}	$10,000$	< 1
30	10^{31}	10^{19}	3×10^{11}
50	3×10^{52}	10^{40}	3×10^{32}
3038	10^{9300}	10^{9288}	10^{9280}

A Million Dollar Problem

Conclusion

To find the shortest path by trying all possible paths will take too much time, even for small numbers of cities.

Facts

- *Nobody has found a fast method for solving the Travelling Salesman Problem or the types of problems mentioned at the start of this talk.*
- *These kinds of problems are said to be Quickly Checkable (or to lie in NP).*
- *Nobody has proven that a fast method doesn't exist to solve problems in NP.*
- *Experts refer to this unresolved question as the P = NP problem. (Here P denotes the class of problems that can be quickly solved by a computer in "polynomial-time".)*
- *If SAT can be shown to lie in P then all problems in NP can be quickly solved (i.e., also lie in P).*

How Long Does it take to Factor a Number?

Factoring

Another problem that seems to be hard is that of factoring a number.

Fact

Every natural number can be factored as a product of prime numbers.

Examples

- $24 = 2 \times 2 \times 2 \times 3$.
- $11833575 = 3 \times 5 \times 5 \times 13 \times 53 \times 229$.
- $20804093 = 1087 \times 19139$

Problem

Given an n-digit number, find its prime factors.

A Big Number

Fact

Using the equivalent of a single 2.2 GHz computer it would have taken approximately 75 years to factor this 200 digit number. In fact last year, using 80 fast parallel computers, it took 12 months.

$$
\begin{aligned}
& 27997833911221327870829467638722601621070446786955428537560009929326128400107609345671052955360856061822351910951365788637105954482006576775098580557613579098734950144178863178946295187237869221823983 \\
= \; & (353246193440277012127260497819846436867119740019762502364930346877612125367942320005854795652808834921996073433034145576787281815213538140930474018546\!7) \\
& \times (792586995447833303334708584148005968773797585736421996073433034145576787281815213538140930474018546\!7)
\end{aligned}
$$

How Long Does it take to Factor a Number?

The following table show how long it would take to factor big numbers using today's most powerful computers and most efficient factoring programs.

Number of Digits	Number of Years
200	1
230	$215,000$
300	$342,000,000$
500	10^{15}

Encryption

Facts

- *Many internet transactions use a method called RSA encryption to encode sensitive information, such as credit card numbers.*
- *RSA encryption is a kind of coding method known as a public key encryption system.*
- *The security of the RSA system relies on the fact that it is difficult to factor big numbers.*
- *If someone were to find a fast way to factor large numbers then most of the security systems used by governments and computer systems would become useless.*

Conclusions

Conclusions

- An important feature of a problem is its computational complexity.
- In general, it is difficult to determine the computational complexity of a given problem.
- The P = NP problem is one of the most important unsolved problems in mathematics.
- A lot of research is being conducted on the P=NP problem and related questions.
- Solving one of these problems could have a significant impact on technology and society.