

MO640/MC668

Guilherme P. Telles

IC-Unicamp

Avisado está

- Estes slides são incompletos.
- Estes slides contêm erros.

Parte I

Computação com DNA

DNA

- Uma molécula de DNA é formada por duas cadeias (ou fitas) complementares de DNA
- Uma base A emparelha com uma base T e uma base C emparelha com uma base G

```
ACTGAACATTCGTTGAGG  
TGAATTGTAAGCAAGCTCC
```

- Em condições adequadas, as fitas de uma molécula de DNA podem ser separadas em solução e, da mesma forma, moléculas de DNA em fita simples em solução podem se unir por complementaridade.

Computação baseada em moléculas

- Computação baseada em moléculas é uma forma de computação que usa enzimas e moléculas de DNA, RNA ou proteínas.
- Instâncias de problemas são codificadas como moléculas e reações químicas as transformam.
- Ao final das reações novas moléculas codificam a resposta para o problema.

Computadores baseados em DNA

- Armazenam volumes enormes de dados em espaço mínimo: 1 bit por nm^3 em moléculas de DNA contra 1 bit por 10^{12} nm^3 em uma fita de vídeo.
- Realizam um número muito grande de operações, da ordem de 10^{20} operações por segundo,
- Consumem 10 ordens de grandeza menos energia por operação.
- Provavelmente serão usados em conjunto com os computadores baseados em silício, resolvendo problemas cuja solução exija grande esforço computacional.

Experimento de Adleman

- Leonard Adleman, em 1994.
- Demonstrou que a técnica funciona ao resolver um problema complexo usando DNA.
- O problema que Adleman escolheu foi o do caminho Hamiltoniano em grafos.

Intuitivamente

- Dado um mapa de uma cidade em que as mãos de direção das ruas estão indicadas e dados endereços de entregas para um caminhão, o problema é encontrar uma rota que começa no depósito, passa por todos os endereços uma única vez respeitando a mão de direção das ruas e termina na garagem.

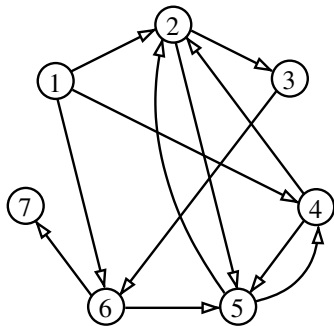
Formalmente

- Um grafo orientado G com vértices distinguíveis v_s e v_t tem um caminho Hamiltoniano se e somente se existe uma seqüência de arestas e_1, e_2, \dots, e_k que começa em v_s , termina em v_t e passa por todos os outros vértices uma única vez.

Exemplo

- Considerando-se $v_s = 1$ e $v_t = 7$, o grafo contém o caminho Hamiltoniano

$(1, 4), (4, 5), (5, 2), (2, 3), (3, 6), (6, 7)$



Dificuldade

- O problema do caminho Hamiltoniano é um problema NP-completo.
- Mesmo instâncias moderadas do problema (por exemplo, grafos com mais de 1000 vértices) demandam um tempo impraticável de computação.
- Mesmo assim, o algoritmo a seguir pode ser usado para resolver o problema.

Algoritmo

1. Gere caminhos no grafo aleatoriamente
 2. Mantenha apenas os caminhos que começam em v_s e terminam em v_t .
 3. Se o grafo tem n vértices, mantenha apenas os caminhos com n vértices.
 4. Mantenha apenas os caminhos que contém os vértices do grafo apenas uma vez.
 5. Se resta algum caminho, responda sim. Senão responda não.
-

Algoritmo

- Não é difícil perceber que esse algoritmo não é muito eficiente: pode ser necessário gerar um número enorme de caminhos no Passo 1 e depois processar todos esses caminhos.
- Para que esse algoritmo funcione em um computador digital, é necessária uma garantia de que, se houver caminho Hamiltoniano no grafo de entrada então esse caminho será gerado no Passo 1.

Idéia de Adleman

- Codificar os vértices e arestas do grafo como moléculas de DNA e através de ligações químicas entre as moléculas, construir todos os caminhos do Passo 1 ao mesmo tempo.
- Alcançou um grau de paralelismo muito maior do que é possível com computadores digitais.
- Todos os passos do algoritmo são feitos “em tubos de ensaio”, usando técnicas da biologia molecular.

Experimento

- Adleman codificou cada vértice do grafo como uma sequência aleatória de DNA com 20 bases.
- Para cada aresta (u, v) do grafo ele criou uma molécula de DNA com o complemento das 10 últimas bases do vértice u e com o complemento das 10 primeiras bases do vértice v .
- Para as arestas da forma $(1, u)$ e $(u, 7)$, todas as bases dos vértices 1 e 7 foram adicionadas.

Codificação

- Para codificar o grafo anterior pode-se construir as seguintes moléculas:

Vértices		Arestas		
		(1,2)	gacttacagt	aggttaacct cgttttccgg
		(1,4)	gacttacagt	aggttaacct cgacgctcgc
		(1,6)	gacttacagt	aggttaacct tcatacctccc
1	ctgaatgtca tccaattgga	(2,3)	ttgtactgga	cctccatagc
2	gcaaaaggcc aacatgacct	(2,5)	ttgtactgga	tagtggtgcg
3	ggaggtatcg ctacaggtga	(3,6)	gatgtccact	tcatacctccc
4	gctgcgagcg tctgagacgc	(4,2)	agactctgcg	cgttttccgg
5	atcaccacgc tcacgatgat	(4,5)	agactctgcg	tagtggtgcg
6	agtaggaggg caatcacgaa	(5,2)	agtgtacta	cgttttccgg
7	tgtcaagcaa tcacttgacg	(5,4)	agtgtacta	cgacgctcgc
		(6,5)	gttagtgctt	tagtggtgcg
		(6,7)	gttagtgctt	acagttcggt agtgaactgc

Experimento

- Uma certa quantidade de moléculas representando vértices e arestas foi misturada em solução.
- Depois, as condições da solução foram alteradas para que as fitas se unissem.
- As uniões entre as moléculas no tubo de ensaios são aleatórias e o resultado é que todos os caminhos orientados no grafo são formados com alta probabilidade, por conta da forma das moléculas.

Caminhos como moléculas

1 4 5
ctgaatgtca tccaattgga gctgcgagcg tctgagacgc atcaccacgc tcacgatgat
gacttacagt aggttaacct cgacgctcgc agactctgcg tagtggtgcg
(1,4) (4,5)

1 6 7
ctgaatgtca tccaattgga agtaggaggg caatcacgaa tgtcaagcaa tcaactgacg
gacttacagt aggttaacct tcatactccc gttagtgtctt acagttcgtt agtgaactgc
(1,6) (6,7)

4 5 4 2
gctgcgagcg tctgagacgc atcaccacgc tcacgatgat gctgcgagcg tctgagacgc gcaaaaggcc aacatgacct
gactctgcg tagtggtgcg agtgctacta cgacgctcgc agactctgcg cgttttccgg
(4,5) (5,4) (4,2)

Filtragens

- Depois, usando PCR a partir dos vértices inicial e final, filtrou apenas os caminhos que começam e terminam nesses vértices.
- Depois apenas as moléculas de fita dupla com tamanho $20n$ são filtradas e mantidas.
- Depois as moléculas em que todos os vértices aparecem são mantidas.
- Se sobrar alguma molécula, a resposta é sim.

Experimento

- O procedimento de Adleman levou uma semana.
- Mas a reação que gerou todos os caminhos codificados como moléculas de DNA levou menos de um segundo.
- Para a reação ele usou em torno de 3×10^{13} cópias de cada molécula representando vértices e arestas.
- Este volume de moléculas foi escolhido para garantir alta probabilidade de que, se houvesse um caminho Hamiltoniano no grafo então ele seria codificado pelas moléculas.
- O tamanho das moléculas que codificam vértices e arestas foi escolhido para que a probabilidade de que arestas e vértices incorretos se unissem fosse pequena e para que as moléculas formadas fossem estáveis à temperatura ambiente.

Experimento

- Mesmo tendo usado um grafo pequeno, Adleman demonstrou a possibilidade de usar moléculas para computar problemas combinatórios que são difíceis de resolver usando computadores baseados em silício.
- A técnica é sensível a alguns tipos de falhas, como a formação de soluções que não são factíveis ou a perda de soluções factíveis durante o processamento *in vitro*.

Extensões

- Desde o trabalho de Adleman, alguns outros problemas foram resolvidos usando moléculas com intervenção humana no processamento.
- A técnica desenvolveu-se de forma a construir um computador baseado em DNA autônomo, isto é, um conjunto de enzimas que quando misturado às moléculas de DNA que codificam o problema que se quer resolver, desencadeiam uma série de reações que resultam em cadeias de DNA que codificam a solução do problema sem intervenção humana.
- Outros trabalhos exploram os limites teóricos da computação com DNA e mostram como contornar possíveis erros experimentais neste tipo de computação.

- Este conjunto de avanços aponta na direção da construção de um computador programável autônomo para resolver problemas gerais que não podem ser resolvidos em tempo aceitável pela tecnologia atual (como por exemplo, a quebra de algoritmos criptográficos).