Complexidade de Comunicação

Luciano Silva

Orientador: Yoshiharu Kohayakawa

Departamento de Ciência da Computação,IME–USP

Sumário

Discutiremos a quantificação dos custos de comunicação num modelo de computação distribuída formado por dois processadores. Obteremos um limitante inferior para a complexidade de comunicação e mostraremos alguns resultados de problemas de decisão envolvendo seqüências neste modelo.

1 Protocolos e árvores-protocolo

Sejam $a_1, ..., a_n$ e $b_1, ..., b_m$ as possíveis entradas de dois processadores P_1 e P_2, respectivamente. A tarefa deles será computar um bit c_{ij}, a partir de a_i e b_j, com trocas de mensagens. Para isto impomos um *protocolo* : descrevemos para cada processador sua próxima ação, em função da entrada e dos bits recebidos/enviados. Assumirémos que os processadores seguem o protocolo e desconsideraremos os custos de processamentos locais. A matriz $C = (c_{ij})$ é chamada de *matriz de comunicação* do problema e completamente o descreve. Ambos os processadores conhecem toda a matriz C. O processador P_1 conhece o índice i de uma linha de C e P_2 o índice j de uma coluna. Para computar c_{ij}, P_1 pode, por exemplo, enviar a P_2 o número i. Isto significa $\lceil \log n \rceil$ bits.[1] Este processo é conhecido como *protocolo trivial*.

Qual é a relação entre o protocolo e esta matriz? Primeiramente, o protocolo precisa determinar quem começa. Suponha que P_1 envie o primeiro bit ε_1. Este bit precisa ser determinado pelo índice i, ou seja, as linhas de C precisam ser divididas em duas partes, de acordo com $\varepsilon_1 = 0$ ou 1. A matriz C então fica decomposta em duas submatrizes, C_0 e C_1. Esta decomposição é determinada pelo protocolo, mas ambos os processadores o conhecem. Então, a mensagem de P_1 determina em que submatriz está a sua linha. A próxima mensagem decompõe C_0 e C_1. Se o próximo a enviar é P_2, então ele faz um particionamento das colunas em duas classes; se é P_1, novamente há uma divisão das linhas, e assim por diante.[2] O protocolo pára se os dois processadores reduziram as

[1] Se $m < n$ então é melhor P_2 enviar j para P_1.

[2] Note que duas matrizes num mesmo nível não precisam sofrer a mesma decomposição.

possibilidades a uma submatriz C' e a partir dela eles podem dizer, com certeza, o valor de c_{ij}. Isto implica que C' ou é uma matriz tudo 1 ou tudo 0. Podemos, então, representar graficamente o protocolo como uma árvore binária, onde os nós são submatrizes de C e as folhas são submatrizes tudo 1 ou tudo 0. Se esta árvore for completa e tiver profundidade t, teremos o mesmo número 2^{t-1} de folhas tudo 1 e tudo 0.[3]

Seguindo o protocolo, os dois processadores movem da raiz da árvore para alguma folha. Estando em algum nó, então se os filhos deste nó apareceram por uma decomposição horizontal, quem envia o próximo bit é P_1 e será P_2, caso contrário. O bit será 0 ou 1, se a linha/coluna estiver no ramo esquerdo/direito, respectivamente. Chegando a uma folha, então temos a resposta ao problema de comunicação. A *complexidade de tempo* do protocolo é a profundidade desta árvore. A *complexidade de comunicação* da matriz C é a menor complexidade de tempo de todos os protocolos que resolvem o problema ao qual está associado C, denotada por $\kappa(C)$.

Usando o fato de que o número de folhas tudo 1 é no máximo $2^{\kappa(C)-1}$ e argumentos de Álgebra Linear, podemos provar o seguinte limitante inferior para $\kappa(C)$.

Teorema 1 $\kappa(C) \geq 1 + \log_2(posto(C))$

Em particular, se as linhas (ou colunas) da matriz C formarem um conjunto linearmente independente, então o protocolo trivial é ótimo.

2 Problemas de decisão sobre seqüências

A título de exemplo, aplicando o teorema anterior, podemos obter os seguintes resultados sobre problemas de decisão envolvendo seqüências:

- suponha que os processadores conhecem, cada um, uma seqüência de comprimento n e querem decidir se elas são iguais. Obviamente a matriz associada é $2^n \times 2^n$ unitária. Como seu posto é 2^n, não existe protocolo melhor que o trivial.

- considere um protocolo decidindo se duas seqüências de 0's e 1's de comprimento n são iguais e seja $h > 0$. Então o número de seqüências $a \in \{0,1\}^n$ para as quais o protocolo usa menos que h bits na entrada (a, a) é, no máximo, 2^h.

3 Bibliografia

[1] Lovász, L., *Computational Complexity*, Notas de curso 1994, c. 170pp.

[3] Se a profundidade de uma árvore não completa for t, temos que 2^{t-1} ainda é um limitante superior para o número de folhas tudo 1.