IS622 Week 3 HW

Brian Chu | Sept 13, 2015

Exercise 2.5.1: What is the communication cost of each of the following algorithms, as a function of the size of the relations, matrices, or vectors to which they are applied? (a) The matrix-vector multiplication algorithm of Section 2.3.2.

In the matrix-vector multiplication algorithm, each matrix (M) element produces the key-value pair $(i, m_{ij}v_j)$. The communication cost is the total number of key-value pairs, which is *r x c* or *O(rc)* where r and c are the number of rows and columns in the matrix. The communication cost is not affected whether the vector v is stored in main memory or not.

(b) The union algorithm of Section 2.3.6.

If R and S are the two relations, the union algorithm just passes their sum *(R+S)* as input tuples *t* to key-value pairs *(t, t)*. The communication cost is *O(R+S)*.

(c) The aggregation algorithm of Section 2.3.8.

In each grouped tuple *(a,b,c)*, key *a* represents the group. There may be as many as *r* groups in the relation R so the communication cost may be *O(r)*.

(d) The matrix-multiplication algorithm of Section 2.3.10.

If we assume square matrices with *n* rows and columns, the communication cost is the multiplication of these dimensions: *O(nxn)* or *O(n^{2})*.

Exercise 2.6.1: Describe the graphs that model the following problems. (a) The multiplication of an n × n matrix by a vector of length n.

Taken from page 55: "If we multiply n × n matrices M and N to get matrix P , then there are $2n^2$ inputs, m_{ij} and n_{jk}, and there are n^2 outputs p_{ik}. Each output p_{ik} is related to 2n inputs: m_{i1}, m_{i2}, . . . , m_{in} and n_{1k}, n_{3k}, . . . , n_{nk}.

(b) The natural join of R(A,B) and S(B,C), where A, B, and C have domains of sizes a, b, and c, respectively.

The natural join is the intersection of R and S, namely the tuples (a,b) and (b,c) which represent the graph inputs. The output is the corresponding tuple *(a, b, c)*, where *b* is common to R and S.

(c) The grouping and aggregation on the relation R(A,B) where A is the grouping attribute and B is aggregated by the MAX operation. Assume A and B have domains of size a and b, respectively.

The graph inputs are represented by each (a, b) tuple. The graph outputs are $(a, MAX(b))$ and connected based on similar key values of the grouping variable a.