Supply Chains and Transfer Prices Optimization Using Apache Hadoop Infrastructure and IBM ILOG CPLEX Solver

Dmitry Lakhvich
Bauman Moscow State Technical University
Baumanskaya 2-ya, 5, Russia, Moscow, 105005
e-mail: dmitry.lakhvich@optimalmngrnt.com

Abstract. Article presents the model, mathematical formulation and heuristical algorithm of supply chains and transfer prices problem in case of multinational companies. It demonstrates a solution to this problem based on Apache Hadoop cluster and IBM CPLEX Solver. The presented solution can be applied to problems that can not be solved on a single supercomputer, it divides the original problem into many subtasks that are calculated on a set of computers in the cluster at the same time. Also, the solution space is formed that can be further analyzed and used for other purposes.

Key words: Apache Hadoop, Big Data, Supply chains, transfer price, taxes optimization, bilinear programming, heuristic algorithm, optimal transfer price, multinational company.

1. Main problem

Supply chain and transfer price optimization using Apache Hadoop infrastructure and IBM ILOG CPLEX solver.

Multinational corporations, which produce real things always wanted and want to reduce costs and increase profits of parent company. It is possible in many different ways and there exists a lot of mathematical models. Vidal and Goetschalckx (1999) showed a model of maximizing the profits of multinational corporations after taxes from all subsidiaries as well as after a corporate tax. The assumptions of the model are the following:

- All internal suppliers, plants and distribution centers are considered to be subsidiaries of the US parent company and are supposed to be actively involved in manufacturing, selling, shipping, and servicing activities. As a consequence, the deferral principle applies, that is, the income is not taxed to the parent company until the dividends are received by the shareholders.

- Each internal supplier, plant, and distribution center is assumed to be taxed on their local-source income, except for the parent company, which is subject to tax on its worldwide income. The company attempts to maximize the total profit after tax in all the countries where it operates, but no income is
remitted as a dividend to the shareholders, and therefore the parent company is only taxed on its local income. Further refinements of the model may consider withholding taxes and the reception of dividends from subsidiaries.

- The structure of the organization and its degree of vertical integration allow for the centralization of decisions about the transfer prices.
- Customer demand may or may not be completely satisfied.
- All transfer price variables have a lower and an upper bound, which reflect feasible markups for production costs and a profit margin, or for possible discounts from market prices.
- As a requirement for the justification of transfer prices to tax authorities, all the transfer prices from a given origin and for a given component (or finished product) must be the same for all the destinations. The transfer price, of course, does not include the transportation cost, which is modeled separately. This can be thought of as the transfer price being determined based on production costs, and therefore there is no reason to allow it to be different for different buyers. The allocation of the transportation cost to the origin or destination is a decision variable that is considered independently of the transfer price. If transfer prices are allowed to be different, the problem is easier to solve.
- Import duties are paid by the importing country, based on the FOB value of the transferred products, or based on the CIF value, as appropriate. No export duties are considered.

As it is shown in model, transfer prices play important role in multinational company profit. The inclusion of transfer prices in the model makes it not linear but a bilinear programming problem. New heuristic algorithm for solving this problem was also presented by the authors. The heuristic makes an attempt to linearize the original problem and simplifies the way of searching good supply chain and transfer prices for company.

We’ve found that this algorithm is not effective and that current Big Data technologies let do it quickly and inexpensively. We have chosen Apache Hadoop and IBM ILOG CPLEX for our solution, and made a simple modification of original algorithm. Our solution has the property of scalability and is designed to solve largescale problems.

The obtained results are the solution space (set of local optima of objection function) and can be analyzed a large number of other ways, as well as visualized. In any case, on the basis of the results obtained, this algorithm and solution could be adapted to other models and fields such as problems arising in the oil and gas industry.
2. Mathematical formalization of problem

The model is formalized as a non-convex optimization problem with a linear objective function, a set of linear constraints, and a set of bilinear equalities $P(x; t; v; p)$. The problem has the following general structure:

$$P(x; t; v; p) \rightarrow \max.$$ 

Subject to

$$c^T x + d^T v + x^T A t + x^T B p = f_r; \quad r = 1, m;$$
$$C x \leq b;$$
$$T^I t \leq T^u;$$
$$0 \leq p \leq 1;$$
$$x \geq 0, t \geq 0, v \geq 0.$$

Where $A_r$, $B_r$ ($r = 1, m$) and $C$ are the matrices of coefficients; $b$ the right-hand side vector of related constraints; $c_r$ ($r = 1, m$), $d_r$ ($r = 1, m$), the vectors of coefficients; $f_r$ ($r = 1, m$), the fixed costs at internal suppliers, plants, and DCs; $p$ the vector of proportions to allocate transportation costs; $t$ vector of transfer prices; $T^I; T^u$ the LB and UBs on transfer prices; $v$ vector of profit and loss variables; $x$ is the vector of material flows.

3. Infrastructure and problem solving

We use standard Apache Hadoop cluster where each data-node has its own IBM ILOG CPLEX solver. At each data-node the condition is imposed, that it has enough memory to solve the problem in the linearized form and for all iterations discussed below. We also use Java opencsv library for parsing csv data-files.

4. Initial variables

On the basis of a mathematical formalization of the model, we determine the total number of unique sequences and form all the variables of the problem. Based on this sequence, formed by the matrix, wherein:

- the same type according to the destination variable are placed close;
- all quadratic elements (products) should be placed after all linear elements.

Scalars: $\text{Nvar}$ — number of one-dim variables; $\text{Npro}$ — number of products; $\text{Ivartp}$ — initial index for transfer prices; $\text{Qvartp}$ — number of transfer prices; $M$ — number of constrains.
Initial matrix:
- Matrix has size \((N_{\text{var}} + N_{\text{pro}} + 2) \times (M + 5)\).
- First line contains scalar values \((N_{\text{var}}, N_{\text{pro}}, \ldots)\) and indexes of transfer prices for product.
- Next line contains indexed of variables and after them indexes of another variables for product.
- Next \(M\) lines first \(N_{\text{var}} + N_{\text{pro}}\) rows contains numerical coefficient for constrains. Each \(N_{\text{var}} + N_{\text{pro}} + 1\) row contains binary operator of constrain. And \(N_{\text{var}} + N_{\text{pro}} + 2\) right side value of constrain.
- \(M + 3\) line contains coefficients for objective function.
- \(M + 4\) line contains lover bound for each variable and \(M + 5\) upper variable.

5. Process workflow

Workflow of BLP problem solving consists of next phases:
1. Generation of initial transfer prices grid.
2. Loading initial transfer prices in HDFS.
5. Sorting the results.

In the first phase, we generate sets of values of transfer prices that satisfy the constrains. The resulting sets are combination of sets of all possible combinations of upper and lower bound of the transfer price for each nomenclature. For problems in which there are a large number of transfer prices original value can be replaced by value of bit field, where 0 — means the lower bound of transfer price, and 1 means the upper bound. So 64 prices are encoded in one 64 bit number. In next step we upload these sets into HDFS and start Apache Hadoop streaming job. Mapper and a reducer are executables that read the input from stdin (line by line) and emit the output to stdout. As the mapper task runs, it converts its inputs into lines and feed the lines to the stdin of the process. In our solution each line is a set of initial transfer prices. The output of each map is a key/value pair. In our case a key is hash value from the received transfer prices, and as the value — local maxima coordinates and value of the objective function. Reducer just sorts the results. The algorithm of the phase will be described in the next chapter. When job is finished HDFS contains sorted solution space for further analyses.

6. Workflow for map job and data-node

Each node has initial data of initial matrix. The mapper data is set of initial transfer prices. Each map uses IBM ILOG CPLEX for linear programing (LP) problems solving.
As mentioned initially, the original problem is the problem of bilinear programming (BLP).

To parallelize the solution to the original problem we are using the following method presented by Vidal and Goetschalckx (1999). Map takes the initial transfer prices from the stream, fixes them in bilinear programing problem then the problem becomes linear in flows. Next we get set of flows and fix it, now this problem is linear in transfer prices. Thus, we can iterate by successively fixing one set of variables and solving the remaining LP problem for the other set. The process can be terminated when the change in the objective function value is negligible. Then it calculates hash of resulting transfer prices and makes it as a key of map result. For result it uses coordinates of local maxima and value of objective function. The workflow for each data-node is shown on Fig. 1.

Figure 1. Workflow on Apache Hadoop datanode
We do not use internal QP or BLP CPLEX solver because it makes the attempts to parallelize the problem meaningless and loses the local maxima that will not allow further analysis of the solution space.

7. Conclusion

This approach allows solving of optimization problems and maximizing profits for the large companies, which in turn means more profit that can be directed towards dividends to shareholders or reinvested into the company.

All this is achieved through the intelligent use of modern technologies and algorithms that allow splitting the original task into subtasks that can be solved independently. The solution can also be used in various departments of the company in shaping the future strategy of conduct in the global market. An example of further research is the search of unprofitable products.

Further development of this solution will be enhanced with the addition of dynamic components (exchange rates, etc.)

Reference


Расчет трансфертных цен и цепочек поставок крупных мультинациональных компаний при помощи ILOG IBM CPLEX и инфраструктуры Apache Hadoop

Д. С. Лахвич
Московский государственный технический университет им. Н. Э. Баумана
105005, Москва, ул. 2-я Бауманская, 5
e-mail: dmitry.lakhvich@optimalmgmnt.com

Аннотация. В статье рассмотрена модель оптимизации цепочек поставок и трансфертных цен для крупной мультинациональной компании. Исходная задача максимизации общей прибыли основной компании представлена в формализованном виде — задаче билинейного программирования. Расмотрен способ линеаризации данной задачи, а также распараллеливания решения на множество узлов кластера Apache Hadoop. Предложенное решение позволяет решать задачи для компаний обладающих большим числом товарных позиций, а также большим числом дочерних предприятий.

Ключевые слова. Apache Hadoop, Big Data, цепочки поставок, трансфертные цены, оптимизация налогов, билинейное программирование, эвристический алгоритм, мультинациональная компания, транснациональная корпорация.

Литература

Автор: Лахвич Дмитрий Сергеевич — аспирант МГТУ им. Н. Э. Баумана