Maximum Entropy Method and Underdetermined Systems Applied to Computer Network Topology and Routing

MILAN TUBA
Megatrend University Belgrade
Faculty of Computer Science
Bulevar umetnosti 29, 11070 Novi Beograd
SERBIA
tuba@ieee.org

Abstract: The maximum entropy method (MEM) is a relatively new technique for solving underdetermined systems. It has been successfully applied in many different areas. All methods for solving underdetermined systems introduce some additional, artificial constraints. The advantage of the maximum entropy method is that it uses the most natural additional constraint: one that does not introduce any new, arbitrary and unwarranted information. One important property of entropy maximization is that it favors uniform distribution. Network design and analysis almost always involve underdetermined systems, especially when routing policy has to be determined. The number of possible routings grows with the factorial of the number of the nodes in the network and the number of possible topologies is exponential in the number of links. The number of constraints is typically polynomial in the number of nodes in the network. That makes the network design problem a good candidate for the maximum entropy method application. It is intuitively clear that an optimal network should not have overloaded or underutilized links. The hope is that the maximum entropy constraint will give a starting topology and routing with smoothly distributed traffic that would lead to the solution that is closer to the optimal. The problem is computationally feasible and with proper identification and selection of certain parameters the method gives reasonable topology and routing. It is possible to apply MEM if we start our analysis with totally interconnected network of n nodes. Some lines will be dropped later in the process of improving utilization or reducing the cost. To apply the maximum entropy method we have to decide what will be the variables of the system. Some combination of required traffic values can be used for that if we remember that for MEM application we do not need to start with probabilities, but an arbitrary set of numbers which can be normalized. Additional parameters are introduced which allow the control of optimization process. Philosophical discussions about the real meaning of the maximum entropy method are interesting, but since the method was successfully applied in many areas, for any new area the most important criterion is not how well can we explain the relation between the MEM and that area, but how useful are the results we get by applying the method.

Key–Words: Maximum entropy method, Computer network routing, Optimization, Modeling

1 Introduction

The network design problem is an old, but also an unsolvable one. It is a very interesting problem because it has great practical value and since it is untractable, heuristics and suboptimal solutions are used for decades. It is an open problem and since unique best solution can not be found, every new approach is promising in the sense that solution obtained can be better then previous ones, at least in some cases. The maximum entropy method has been used recently in many different areas for solving underdetermined systems. An analysis of both, network design problem and maximum entropy method, is given here with the argument that maximum entropy method can be a reasonable way to approach the network design problem.

2 Maximum Entropy Method

The maximum entropy method was recently used with great success in many different areas where underdetermined systems are involved. It is most frequently used in chemistry [1], but also in many other very diverse areas: character recognition [2], data analysis [3], image processing [4], [5], economy [6]. Theoretical developments also continue [7].

The basic idea is to get a unique solution from the underdetermined system by introducing the additional constraint that the entropy function should be maximized. The other methods that were used for solving underdetermined systems use the same technique: they introduce additional, artificial constraints that make the number of constraints equal to the number of unknowns. The difference is that the maximum entropy method introduces the most natural additional constraint: one that does not introduce any new, arbitrary and unwarranted information. It uses only the information that is given and makes no assumptions about missing information.

Before we go to the formal definition of the maximum entropy principle, it is interesting to mention that besides very pragmatic uses (like here) there were very extensive philosophical discussions about the real meaning of this principle. The predecessor of the maximum entropy principle is the principle of insufficient reason (James Bernoulli: *"Ars Conjectandi,"* 1713). It states that in the absence of any information (knowledge), all outcomes should be considered equally possible. It was involved in the discussions about prior probabilities (probabilities of one event, state of the knowledge) and relative frequencies. Relative frequencies become predominant and some useful works from Laplace and Bayes were criticized. Shannon's works on information theory opened a new opportunity for revitalization of the principle of insufficient reason, this time as a more sophisticated maximum entropy principle that was introduced by Jaynes.

2.1 Definition of the MEM

Let us now give the formal definition of the maximum entropy method:

Suppose that for a discrete random variable X we know the values $x_1,\ x_2, ...,\ x_n$ that it can take, but we do not know the corresponding probabilities $p_1,\ p_2, ...,\ p_n$. We also know expected values for $k < n-1$ functions of X (for example, the first k moments):

$$E[\,f_r(X)] \;=\; m_r \qquad r = 1, 2, ..., k. \tag{1}$$

In fact, we do not need to know the values $x_1,\ x_2,\ ...,\ x_n$, or analytical expressions for functions $f_r,\ r = 1, 2,\ ..., k$. It is sufficient to know the values $f_r(X_i);\ r = 1, 2,\ ..., k;\ \ i = 1, 2,\ ..., n$. Also, we do not have to start with probabilities $p_1,\ p_2,\ ...,\ p_n$. We can start with any set of numbers $t_1,\ t_2,\ ...,\ t_n$. Then we introduce

$$p_i \;=\; \frac{t_i}{\sum t_j}$$

This gives us (together with $\sum p_i = 1$) $k+1 < n$ constraints for n unknown variables $p_1,\ p_2,\ ...,\ p_n$. This system is underdetermined and has an infinite number of solutions. We want to find the unique solution that maximizes the entropy of the system. That is the best solution, in the sense that it uses only the information given. It is neutral to the missing information (it does not introduce any hidden assumptions). This additional constraint can be expressed as:

Maximize the entropy function

$$H(p_1,\ p_2,\ ...,\ p_n) \;=\; -\,K\,\sum_{i=1}^{n} p_i\, ln(p_i). \tag{2}$$

If we select $K = 1$, entropy will be expressed in natural units (rather than in bits).

2.2 Solution

We will use the method of Lagrange multipliers. This will not guarantee us that probabilities are non-negative. We introduce the substitution $p_i = e^{-q_i}$. This gives us a stronger constraint than the one that we wanted: all probabilities are now positive definite (none of them can be zero). Our problem now is to maximize

$$H(q_1,\ q_2, ...,\ q_n) \;=\; \sum_{i=1}^{n} q_i\, e^{-q_i} \tag{3}$$

under the conditions

$$\sum_{i=1}^{n} e^{-q_i} \;=\; 1 \tag{4}$$

$$\sum_{i=1}^{n} e^{-q_i} f_r(x_i) \;=\; m_r\ , \qquad r = 1, 2, ..., k \tag{5}$$

When we introduce Lagrange multipliers $\lambda,\ \mu_1,\ \mu_2, ...,\ \mu_k$, we get the function:

$$F(q_1,\ q_2, ...,\ q_n) \;= \tag{6}$$

$$\sum_{i=1}^{n} q_i\, e^{-q_i} \;+\; \lambda \sum_{i=1}^{n} e^{-q_i} \;+\; \sum_{r=1}^{k} \mu_r \sum_{i=1}^{n} e^{-q_i}\ f_r(x_i)$$

All partial derivatives should be zero:

$$\frac{\delta F}{\delta q_i} \;=\; e^{-q_i}\,[1 - q_i - \lambda - \sum_{r=1}^{k} \mu_r\ f_r(x_i)] \;=\; 0\ , \tag{7}$$
$$i = 1, 2, ..., n$$

e^{-q_i} is never zero, so we get

$$q_i \;=\; 1 - \lambda - \sum_{r=1}^{k} \mu_r\ f_r(x_i)\ , \qquad i = 1, 2, ..., n \tag{8}$$

The problem is now solved: (4), (5), and (8) give $n+k+1$ equations for $n+k+1$ unknown variables $p_1,\ p_2, ...,\ p_n,\ \mu_1,\ \mu_2, ...,\ \mu_k,\ \lambda$. The system should have unique solution, but it is not linear and some numeric method have to be used.

To make the calculations easier, we introduce the partition function:

$$Z(\mu_1,\ \mu_2, ...,\ \mu_k) = \sum_{i=1}^{n} p_i\ e^{-\lambda} = \sum_{i=1}^{n} e^{-\lambda - q_i} \quad (9)$$

$$Z(\mu_1,\ \mu_2, ...,\ \mu_k)\ =\ \frac{1}{e}\sum_{i=1}^{n} e^{\sum_{r=1}^{k}\ \mu_r\ f_r(x_i)} \quad (10)$$

It is easy to see that

$$\lambda\ =\ -\ lnZ(\mu_1,\ \mu_2, ...,\ \mu_k) \quad (11)$$

$$m_r\ =\ \frac{\delta}{\delta\mu_r} lnZ(\mu_1,\ \mu_2, ...,\ \mu_k) \quad (12)$$

or

$$m_r\ =\ \sum_{i=1}^{n}[m_r - f_r(x_i)]e^{\sum_{j=1}^{k}\ \mu_j\ f_j(x_i)} = 0\ , \quad (13)$$
$$r = 1, 2, ..., k$$

(13) gives k equations for k unknown variables $\mu_1,\ \mu_2, ...,\ \mu_k$. When we solve it, from (11) we get λ, and then from (8) $q_1,\ q_2, ...,\ q_n$, and finally, from $p_i = e^{-q_i}$ we get $p_1,\ p_2, ...,\ p_n$.

We can introduce $t_j = e^{\mu_j},\ j = 1, 2, ...,\ k$. Then (11) and (13) become:

$$\lambda\ =\ 1 - ln[\sum_{i=1}^{n} \Pi_{j=1}^{k} t_j^{f_j(x_i)}] \quad (14)$$

$$\sum_{i=1}^{n}[m_r - f_r(x_i)]\Pi_{j=1}^{k} t_j^{f_j(x_i)}\ =\ 0, \quad (15)$$
$$r = 1, 2, ..., k$$

There is a standard algorithm to solve this system. However, the function that we need to minimize is not convex even in the simplest case when there is only one constraint: expected value. The standard Newton-Rapson procedure will not work. But the Jacobian matrix for the system we are solving is symmetric and positive definite. This gives us a scalar potential function which is strictly convex and whose minimum is easy to find. The use of the second order Taylor expansion is recommended. However, after much experience with the algorithm, we consider that it is not even worth trying to find the exact value for α that determines how far to go along a certain direction, let alone inverting the Jacobian matrix every time.

2.3 Selection Principle

The previous model has constraints $p_i > 0,\ i = 1, 2, ..., n$. This may be too strong since the probabilities need only to be nonnegative. To make $p_i \geq 0$, we can introduce $p_i = q_i^2$ instead of $p_i = e^{-q_i}$, which we had before.

In practice, we have to decide which p_i will be zero. We can do it in advance and consider a model that has only $n - m$ probabilities (if m probabilities are selected to be zero). If we select too many probabilities to be zero, the system may become overdetermined.

3 Network Design Problem

Computer networks consist of computers, called nodes, and communication lines, called links, that interconnect them. All data that is exchanged among nodes is divided into packets. Destination address is added to messages and these packets are sent to neighboring computers that send to their neighbors and so on, until the message reaches its destination.

The network design problem is:

- For given locations of nodes, traffic matrix (offered traffic for each pair of nodes) and cost matrix (cost to transfer a message for each pair of nodes)
- With performance constraints: reliability, delay (time that a message spend in the network), throughput
- Find values for variables: topology (which nodes will be connected directly with a line and which will have to communicate indirectly, using other nodes as intermediate stations), line capacities (how much traffic will each link be able to carry), flow assignment - routing (which paths messages between any pair of nodes will follow)
- Minimize the cost (of building and maintaining the whole network).

Other formulations of the problem are: minimize delay for the given cost or maximize throughput for given cost and delay. It has been shown that all these problems are similar and that the same techniques can be applied. Different aspects of the network design problem, particularly routing and link capacity were investigated [8], [9], [10]. More recent results are in [11] and [12] and the latest survey [13].

This problem is intractable if full and exact solution is required. Networks can have many hundreds of nodes (computers). Fortunately, experience has shown that network design can be done hierarchically and still be near optimal. An example is a network for a country. First, we can decide where to put trunks between major cities, then connect small cities to nearest major cities, then make local networks inside the cities. This approach allows us to work with networks of at most 50 nodes at a time. This is a great help, but the problem is still intractable.

The network design problem, that was for many decades investigated with emphasis on wide area networks, is recently revitalized with application to mobile ad hoc networks [14], [15], [16].

4 Suitability of the MEM for the Network Design Problem

Network design and analysis almost always involve underdetermined systems, especially when routing policy has to be determined. The number of possible routings grows with the factorial of the number of the nodes in the networks and the number of possible topologies is exponential in the number of links. The number of constraints (such as "everything that goes in must go out" for each node that is neither source nor sink) is typically polynomial in the number of nodes in the network.

The problem of network design is to find a topology, routing and capacity assignment such that cost or delay is minimized. Once the topology and the routing are decided, there are exact methods for capacity assignment that will minimize delay or cost. However, there are very few theoretical results on how to select topology and routing. Most of the algorithms that are used today are heuristic and many of them do not even have intuitive justification other then "easy to calculate" or "only simple thing we can do." Here is presented an attempt to use the maximum entropy method to select (initial) topology and routing. The problem of selecting topology and routing is (almost) always an underdetermined one. To solve it, most methods that are currently used introduce new, artificial constraints. These artificial constraints do not have any justification other than that they make the number of unknowns and the number of constraints equal. The maximum entropy method has the nice property that it solves underdetermined systems without introducing any new, unwarranted information. The other advantage of the maximum entropy method is that it makes things as equal as possible. It is intuitively appealing that a network should not have "hot spots," i.e. traffic should be distributed as equally as possible along all lines. The same goal can be attained by using some other function that has maximum when all variables are equal. One very simple such function is the product of all variables. The product function expression seems simpler then entropy function expression which involves logarithms, but when we remember that we need partial derivatives we see that entropy function is better since it separates variables.

It is possible to apply MEM if we start our analysis with totally interconnected network of n nodes. Some lines will be dropped later in the process of improving utilization or reducing the cost. To apply the maximum entropy method we have to decide what will be the variables of the system. Some combination of required traffic values can be used for that if we remember that for MEM application we do not need to start with probabilities, but an arbitrary set of numbers which can be normalized.

There is no theoretical explanation of which of the heuristics used so far are better and under what circumstances. It is worthwhile to try to find new solutions for the network design problem since it is a very difficult one and far from being solved. Computer networks are very numerous today and carry a lot of traffic and reducing a cost for only 1% would be a significant result. Philosophical discussions about the real meaning of the maximum entropy method are interesting, but since method was successfully applied in many areas, for any new area the most important criterion is not how well can we explain the relation between the MEM and that area, but how useful are the results we get by applying the method.

5 Variables and Constraints

Let us consider given traffic matrix $t_{i,j}$, line capacity C and total traffic T.

The network design problem has to be fitted to the model described in the previous section. It is possible to apply MEM if analysis is started with totally interconnected network of n nodes. Initial feasible routing is then trivial. Some lines will be dropped later in the process of improving utilization or reducing the cost.

To apply the maximum entropy method, it has to be decided what will be the variables of the system. Some combination of the required traffic values can be used for that, since for the MEM application it is not necessary to start with the probabilities, but with an arbitrary set of numbers which can be normalized.

It may be desirable to have as variables the traffic along different lines; that is what should be made as equal as possible. However, these variables are too coarse. From them the routing can not be determined. The more serious problem is that there are no natural constraints on these variables.

This forces us to select as variables of the system something finer: the traffic of a particular message type (message types are distinguished by the source and destination for a message) on a particular line.

ISSN: 1790-5109

ISBN: 978-960-474-107-6

The number of different message types is $n(n-1)$ (from each node to every other node, except itself). The number of different lines is also $n(n-1)$. Assumption is that each pair of nodes is connected with two lines, one in each direction. If full duplex is considered, the number of lines is one half of the previous case. However, the same model has to be retained since routing has to be determined. The only difference will be the fact that offered load matrix is symmetric.

There is a variable for each pair (message-type, line) so the total number of variables is $n^2(n-1)^2$. Each variable can be marked with four indices of the form $T_{message;line}$ where *message* and *line* are represented each with two indices: *source* and *destination*. The final form of the variables is $T_{MS,MD;LS,LD}$.

Constraints that enforce feasible routing can be determined as follows. For each node there is an equation for each message type. The total number of equations is then $n^2(n-1)$, plus the equation that establishes that the sum of all probabilities is equal to 1. In this case, the last condition is equivalent to the requirement that the total network traffic is equal to some given constant within a certain range.

The equations will express the following conditions: for each transit node the flow-in is equal to the flow-out for each message type separately. For the source nodes and the sink nodes, equation is balanced by the required load for particular message type.

All the coefficients on the left side of this system will be 0, 1 or -1. The right side of the system is the offered load. This system represents constraints from the general MEM model.

The matrix for this system is large, but fortunately very sparse. In each equation only $2(n-1)$ coefficients are different from zero (the number of input and output lines for one node). Some of the coefficients on the right side of the system are not zero, but neglecting that, the total number of the coefficients that are different from zero is $2n^2(n-1)^2$ (the number of equations times the non-zero coefficients in each equation). The total number of the elements in the system is $n^4(n-1)^3$ (the number of equations times the number of variables). The density of the matrix is then calculated as $\frac{2}{n^2(n-1)}$. The density approaches zero with the cube of the number of nodes, which means that is inappropriate or impossible to keep such a matrix in the memory (even though the computation would be faster). It will be necessary to define a function that will, based on the indices, compute coefficients of the matrix.

Table 1 lists memory requirements and matrix density for different size networks, assuming optimistic four bytes per variable, neglecting first equation and right side of the system.

Nod.	Variables	Equat.	Memory	Density
3	36	18	2.5 KB	11.11%
5	400	100	152.6 KB	2.00%
10	8100	900	27.8 MB	0.22%
20	144400	7600	4.2 GB	0.03%

Table 1: Memory requirements

We implemented an algorithm for calculating matrix values, rather than keeping them in the memory.

The previous model makes it possible to use the MEM for the network design problem. It uses feasible routing as a constraint and it gives a solution, but not the one that we hoped for. The reason to use the MEM was its tendency towards equalization, but in that model the variables were not those that had to be equalized but those that were necessary for the feasible routing.

It is possible to modify the MEM model in any desired way and to guide the process.

6 Conclusion

The network design problem is a candidate for the maximum entropy method application since the routing problem is an underdetermined one. It is intuitively clear that an optimal network should not have overloaded or underutilized links. The hope is that the maximum entropy constraint will give a starting topology and routing with smoothly distributed traffic that would lead to the solution that is closer to the optimal. The problem is computationally feasible and it seems that, with proper identification and selection of certain parameters, the method will give reasonable topology and routing.

The criterion for use of the maximum entropy method in the network design will be how useful results can we get from it. If nothing else, we can use the property of the maximum entropy method that it tries to make the probabilities as equal as possible for given constraints. Without trying to prove it now formally, it seems natural that a network should not have “hot spots”, lines that carry much more traffic that other lines. The maximum entropy method can help to smooth traffic along all lines.

A combination of theory and experiments should be applied in this research. Known theoretical properties of networks set initial computer experiments which, in turn, point to some new properties that may be theoretically proven.

Acknowledgments: The research was supported by the Ministry of Science, Republic of Serbia, Project No. 144007

References:

[1] Ding YS, Zhang TL, Gu Q, et al., Using Maximum Entropy Model to Predict Protein Secondary Structure with Single Sequence *Protein and Peptide Letters* Volume 16, Issue 5, pp. 552-560, 2009

[2] Xuan Wang, Lu Li, Lin Yao, Anwar, W., A Maximum Entropy Approach to Chinese Pin Yin-To-Character Conversion. *2006 IEEE International Conference on Systems, Man, and Cybernetics*, 2006, Taipei, Taiwan

[3] Teh, Chee Siong Lim, Chee Peng, A probabilistic SOM-KMER model for intelligent data analysis, *WSEAS Transactions on Systems* Vol. 5, no. 4, pp. 825-832. Apr. 2006

[4] Zhengmao Ye, Habib Mohamadian1, Yongmao Ye, Practical Approaches on Enhancement and Segmentation of Trimulus Color Image with Information Theory Based Quantitative Measuring, *WSEAS Transactions on Signal Processing*, Issue 1, Volume 4, January 2008, pp. 12-20

[5] Heric, Dusan; Zazula, Damjan, Reconstruction of Object Contours Using Directional Wavelet Transform, *WSEAS Transactions on Computers*. Vol. 4, no. 10, pp. 1305-1312. Oct. 2005

[6] Ciavolino E, Dahlgaard JJ, Simultaneous Equation Model based on the generalized maximum entropy for studying the effect of management factors on enterprise performance, *Journal of Applied Statistics* Volume 36, Issue 7, pp. 801-815, 2009

[7] Aladdin Shamilov, Generalized entropy optimization problems and the existence of their solutions, *Physica A: Statistical Mechanics and its Applications*, Volume 382, Issue 2, August 2007, pp. 465-472

[8] Tuba, Milan: Cost Function for Communication Links in Computer Networks, *Bulletins for Applied Mathematics (BAM)*, LXXIII-1028/94, pp. 115-122, Budapest, 1994

[9] Tuba, Milan: A Mathematical Model for Routing Comparison in Computer Networks, *Bulletins for Applied Mathematics (BAM)*, LXXXVI-A 1565/98, Arad, July 1998, pp. 493-503

[10] Tuba, Milan: Parameters for the Internet Optimization on the Local Level, *Applied & Computing Mathematics*, Vol. II, pp. 139-142, Kosice, 1997

[11] Tuba, Milan: Relation between Static and Dynamic Optimization in Computer Network Routing, *Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases*, WSEAS Press 2009, pp. 484-489

[12] Tuba, Milan: Computer Network Routing Based on Imprecise Routing Tables, *WSEAS Transactions on Communications*, Issue 4, Volume 8, April 2009, pp. 384-393

[13] Abd-El-Barr M: Topological network design: A survey, *Journal of Network and Computer Applications*, Volume 32, Issue 3, pp. 501-509, 2009

[14] Karavetsios, P., Economides, A.: Performance Comparison of Distributed Routing Algorithms in Ad Hoc Mobile Networks, *WSEAS Transactions on Communications*, Vol. 3, Issue 1, 2004, pp. 317-321

[15] Sokullu, R., Karaca, O.: Comparative Performance Study of ADMR and ODMPR in the Context of Mobile Ad Hoc Networks and Wireless Sensor Networks, *International Journal of Communications*, Issue 1, Volume 2, 2008, pp. 45-53

[16] Kumar, D., Bhuvaneswaran, R.: ALRP: Scalability Study of Ant Based Local Repair Routing Protocol for Mobile Ad Hoc Networks, *WSEAS Transactions on Computer Research*, Vol. 3, Issue 4, Apr 2008, pp. 224-233