LOWERING OF THE ENTROPY IN DECISION MAKING BY USING OF MODERN PREDICTION ALGORITHM

Michal Fraňo
Faculty of Management Science and Informatics, University of Žilina, Slovak Republic

Abstract: *Decision making belongs to the key roles of a manager. Decision making is based on informations. Each information contains some amount of entropy. In strategic management the entropy grows with the time horizon that the decision of the manager has to cover. In order to decrease the entropy of information, prognostic algorithms are used. They refine on the information which is used in decision making. The more exact the information is, the higher is the efficiency of the management.*

Keywords: *Confidence Interval, Decision Making, Entropy, Estimation, Prediction Algorithm.*

1. Introduction

Decision making plays a key role in managerial practice. It is one of the basic roles of a manager in every level of management. The basis for decision making are informations. On grounds of knowledge obtained from information the manager is able to make a decision. The more content abundant the information is, the more distinct and effective decision making is. The decision making of the manager is often influenced by an entropy. The report that the manager keeps at disposition, is often not very exact. The information contained in it is not sufficient for decision making. Under these circumstances the decision making could with quite low probability bring an economical benefit for the enterprise. Consequently it is in the managerial practice substantial to be concerned with the question of increasing the accuracy of the information. As for the area of qualitative and quantitative informations about the future development of any process in the enterprise, the manager can use for increasing of the accuracy of the informations the knowledges from the scientific field of forecasting. It offers nowadays powerful and well developed tools for increasing the accuracy of the informations. The increased information accuracy is necessary for long-term management. The importance of the interconnection between management and prognostics is growing. Especially in strategical decision making that covers longer time horizon, should management as the science cooperate narrowly with prognostics.

2. Entropy in Decision Making

By analyzing of the origin of the entropy in decision making it is necessary to use the knowledge of areas of probability and mathematical statistics. In the process of decision making the manager is able to estimate the development of the future process evolution. Many processes, which can be observed today, are the repetition of processes that occured in the past. On the basis of investigation of these processes the manager can bind set of probability of repeating it. From the fact that it is worked with the probability there, results for the manager in decision making process the necessity to face some uncertainty. This uncertainty is also called entropy. The more precise the manager is able to predict the future process development, the smaller measure of entropy will affect his decision. From this reason the effort of a manager, that decides at higher level of management, will concentrate on obtaining of maximally usable information, which means maximum accurate estimation of future process development. The estimation of future process development can have two forms: an estimation of future value of the process or an interval estimation. The value of the estimation gives information about the value of the process in certain moment in the future. By the interval estimation the interval is estimated that will comprise the future value of the proces with desired probability. This interval is known in mathematical statistics as the confidence interval.

2.1 Entropy in managerial practice

Exact information is exceedingly important in the management. Entropy is definded on the basis of formulas known from the theory of information described by C. Shannon *(in Černý, 1981)* . An important issue is the amount of information I(X) that is indirectly proportional to the probability P(X), with which can the receiver of the information guess the content of the message X.

The amount of information is definded by the formula:

$$I(X) = f\,\frac{1}{P(X)} \tag{1}$$

By substitution of particular functional dependency in the formula (1) is obtained following relation:

$$I(X) = -\log_2 P(X) \tag{2}$$

The amount of information I(X) equals the measure of uncertainty H(X) that is through the message retrieved. The measure of uncertainty H(X) is called the entropy. The unit of the entropy in the theory of information is one bit.

2.2 Decreasing of entropy by suitable estimation of the confidence interval

Confidence intervals have an important position in the theory of statistical estimations. The width of the confidence interval is connected with the estimation accuracy. The shorter the interval, the more accurate the estimation is. If the event occures, the I(X) = 0. If the interval is wide enough, it will include the future value of the process. In this case the information that the future value belongs into this interval, will have the measure of the information I(X) = 0. Generally it is true: the bigger the width of the confidence interval, the smaller the entropy. The more accurate the confidence interval after the process of forecasting is, the higher is the amount of entropy that was retrieved.

The principle noted above is shown by the following example from managerial practice.

Assume a time series of 192 observed values that represent monthly amount of manipulated transportation units from January 1990 to December 2005. In this time period 2 estimations for every value are made. The first estimation is made by a higher manager that has all knowledge about the technology of manipulation with transportation units. The second estimation is made by a scientist having no knowledge about the technology of the process of manipulation of transportation units but having knowledge about forecasting using modern prediction algorithms. The real value and both forecasts are shown in the table 1. The forecast made by the manager based on his knowledge of process technology is called the empiric estimation The second value will be obtained by using a modern prediction algorithm working on the principle of backpropagation neural network.

Table 1

December 1999	December 2005
Real value: 33846	Real value: 30621
Empirical estimation: 41819	Empirical estimation: 29845
Prediction on the basis of neural network: 33499	Prediction on the basis of neural network: 30445

In both cases a 95% confidence interval is desired. Significance level α is set according the formula for the 100(1-α)% confidence interval.

The formulas for enumerating the lower and upper bounds of confidence interval for each algorithm can be taken into consideration. These formulas for older types of prediction algorithms are described in *(Gaynor, Kirkpatrick, 1994)*. The confidence intervals with newer algorithms like GMDH or Backpropagation algorithm can be computed for the short-term forecasting, assuming that the forecast errors are independent and Gaussian distributed. According to *(Da Silva,Moulin, 2000)* three techniques for the computation of confidence interval are available for the multilayered neural network trained by the backpropagation algorithm: (1) error output; (2) resampling; (3) multilinear regression adapted to neural networks.

The error output technique appears to be the most appropriate for the managerial practice taking into consideration its universality and relative simplicity. The procedure known from mathematical statistics can be used that hypothesizes about the suspected type of statistical distribution of the error output and verifies the hypothesis by the χ^2 test at predetermined significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

The hypothesis H_0 presumes that the set of deviations (forecast - real value) has the Gaussian distribution. This hypothesis will be verified at the significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

In order to achieve proper results, the statistical set should have at least 50 values. The values of monthly standard deviations for the years 2001 - 2005 will be examined. The mean value of this time series μ is -731.62, the standard deviation σ is 2649.43 and the range is 11642.

After having sorted the members of the statistical set into 8 classes, the test criteria T is enumerated according the formula for the χ^2 test:

$$T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{3}$$

where k is the number of classes, n is the summary frequency, n_i class frequency, p_i is the class probability. The random value T for $n \to \infty$ has under H_0 the distribution $\chi^2(\nu)$, where ν is the number of degrees of freedom. $\nu = k - r - 1$, where r is the number of unknown parametres.

The hypothesis is rejected, when T belongs to the critical area W_α:

$$W_\alpha = \{ T: T > \chi_\alpha^{\ 2}(\nu)\} \tag{4}$$

For the mean value μ = - 731.62 and for the standard deviation σ = 2649.43 the table 2, enumerating the test criteria T, is obtained.

Table 2

Interval				$\Phi(x_i)$	p_i	$n*p_i$	T_i
lower bound	upper bound	x_i	n_i				
(-6740, -5055⟩		-5897.5	4	0.0256	0.0256	1.535971	3.952833
(-5055, -3370⟩		-4212.5	5	0.094453	0.068853	4.131209	0.182706
(-3370, -1685⟩		-2527.5	13	0.248938	0.154485	9.269072	1.501749
(-1685, 0⟩		-842.5	14	0.483308	0.234371	14.06226	0.000276
(0, 1685⟩		842.5	14	0.723789	0.24048	14.42881	0.012744
(1685, 3370⟩		2527.5	6	0.890674	0.166885	10.0131	1.608393
(3370, 5055⟩		4212.5	3	0.968987	0.078313	4.698773	0.614166
(5055, 6740)		5897.5	1	0.993827	0.02484	1.490415	0.161369
			n = 60			**T =**	**8.034237**

Number of the degrees of freedom is $\nu = 8 - 2 - 1 = 5$.

The critical values of the $\chi^2(\nu)$ distribution are obtained according the statistical tables:

for $\alpha = 0.05$ is $\chi_{0,05}{}^2(5) = 11.1$. $T < \chi_{0,05}{}^2(5)$, because 8.034 < 11.1. At the significance level 0.05 there is not possible to reject the hypothesis H_0. It appears from this that H_0 is accepted at the significance level $\alpha = 0.05$.

At the significance level $\alpha = 0.01$ is $\chi_{0,01}{}^2(5) = 15.1$. Test criteria does not belong to the critical area W_α, and the hypothesis H_o is also accepted at the significance level $\alpha = 0.01$.

The formula for enumeration of lower bound and upper bound by the 100(1-α)% confidence interval by known parameter of standard deviation σ is following:

$$\bar{x} - k_a \frac{s}{\sqrt{n}} \leq m \leq \bar{x} + k_a \frac{s}{\sqrt{n}} \tag{5}$$

where k_a is the critical value used by Gaussian probability distribution.

The standard deviation s is defined by the formula

$$s = \sqrt{D(X)}\,, \tag{6}$$

where $D(X)$ is called dispersion of the random value X.

The manager continually empirically estimated the values for January 1999 to November 1999. By this empirical estimation the standard deviation s = 10989.64.

By continual estimating of the monthly values of the year 1999 made by the backpropagation prediction algorithm the standard deviation s = 3239.63.

The formula for lower and upper bounds of the confidence interval for empirical estimation is:

$$\bar{x} - k_{0.05}\frac{s}{\sqrt{n}} \le m \le \bar{x} + k_{0.05}\frac{s}{\sqrt{n}} \tag{7}$$

$$33846 - 1.96\frac{10989.64}{\sqrt{11}} \le m \le 33846 + 1.96\frac{10989.64}{\sqrt{11}}$$

$$27352.27 \le m \le 40339.73$$

For the estimation based on neural network following lower and upper bounds of the confidence interval after substitution following formulas are obtained:

$$33846 - 1.96\frac{3239.63}{\sqrt{11}} \le m \le 33846 + 1.96\frac{3239.63}{\sqrt{11}}$$

$$31931.72 \le m \le 35760.28$$

By the empirical estimation the estimated value fits with the 95% probability the interval $\langle 27352.27\,,\,40339.73\rangle$.

By the estimation made by the prediction algorithm working on the principle of neural network the estimated value fits with 95% probability the interval $\langle 31931.72\,,\,35760.28\rangle$.

Using the modern prediction algorithm the accuracy of estimation was increased in a significant way.

The question is, with what probability would the empirical estimation belong by given standard deviation s in december 1999 to the shorter interval? The shorter interval is the interval obtained by the prediction algorithm.

The mean value in the formula (5) is 33846. Lower and upper bounds of the more accurate confidence interval are: 31931.72 and 35760.28.

The formula (5) after substitution is following:

$$33846 - k_{?}\frac{10989.64}{\sqrt{11}} = 31931.72$$

$$k_{?} = 0.5777$$

The value α = 0.57 is found for the critical value k_a = 0.5777 according the table of critical values of Gaussian probability distribution.

In the case of empirical estimation the narrower interval is (1-0.57)%, that means 43% confidence interval for the mean value of the process. Using the prediction algorithm changes the probability, with which the mean value of the process fits the narrower interval, from 43% to 95%. It is 52% increase of the reliability.

In the same way the increase of the reliability for year 2005 is enumerated. The process for year 2005 was much smoother and thank this characteristic it was possible to estimate better its future development. The mean value of the process in december 2005 is 30621. For the empirical estimation enumerated on the basis of 11 values of the year 2005 the standard deviation σ = 2408.6.

For predictions made using the prediction algorithm based on the principle of neural networks the standard deviation is s = 1688.4. The formula (5) is following:

for empirical estimation:

$$30621 - 1.96\frac{2408.6}{\sqrt{11}} \leq m \leq 30621 + 1.96\frac{2408.6}{\sqrt{11}}$$

$$29197.77 \leq m \leq 32044.23$$

For the backpropagation neural network it is obtained:

$$30621 - 1.96\frac{1688.4}{\sqrt{11}} \leq m \leq 30621 + 1.96\frac{1688.4}{\sqrt{11}}$$

$$29623.34 \leq m \leq 31618.66$$

The empirical estimation fits with a 95% probability the interval $\langle 29197.77\,, 32044.23 \rangle$, the estimation based on the prediction algorithm fits with the 95% probability the interval $\langle 29623.34\,, 31618.66 \rangle$.

By using of modern prediction algorithm it was possible to increase the accuracy of the estimation, but not so enormous, like it was in the year 1999. It is the result of much smoother process in 2005 that was easier to predict.

With what accuracy would the empirical estimation fit by given standard deviation s in december 2005 the shorter interval, which was obtained by the prediction algorithm?

$s_{empirical} = 2408.6$, the true value, that comes out in the formula (5) as the mean value, is 30621. Lower and upper-bound of the more exact interval are: 29623.34 a 31618.66.

By using of the formula 5 is obtained:

$$30621 - k_? \frac{2408.6}{\sqrt{11}} = 29623.34$$

$k_? = 1.373$

The value α = 0.17 is found for the critical value k_a = 1.373 according the table of critical value of Gaussian probability distribution.

The narrower interval in the case of empirical estimation is the (1-0.17)% confidence interval, that means 83% confidence interval, for the mean value of the process. The probability that the mean value of the process fits the narrower interval, in the year 2005 from 83% to 95% by using of a modern prediction algorithm. It is an 12% increase of reliability.

Another question is to be answered at this moment: How does the increased reliability affect the entropy?

The basic formula (2) for the enumeration of the entropy is

$$H(X) = -\log_2 P(X)$$

The probability, that the mean value of the estimation fits the narrower 95% reliability interval, will be used in the formula (2).

In the case of empirical estimation in the year 1999 this probality reaches 43%. In the case of the prediction algorithm the probability is 95%.

$$H_{empiciral}(X) = -\log_2 0.43 = 1.2176$$

$$H_{prediction}(X) = -\log_2 0.95 = 0.074$$

By usage of the prediction algorithm the change the entropy is 1.1436 bit for year 1999.

In 2005 are the values of the entropy following:

$$H_{empirical}(X) = -\log_2 0.83 = 0.26882$$

$$H_{prediction}(X) = -\log_2 0.95 = 0.074$$

and the usage of the prediction algorithm reduced the entropy by 0.19482 bit.

In both cases the use of prediction algorithm changes positively the entropy and the manager can use more accurate information. This helps him to make a more effective decision.

2.3 Decreasing of entropy and its relation to the economic benefit

The use of prediction algorithm and from it resulting change of entropy would have no practical sense for the manager, if it didn´t bring any improvement in economic results of the enterprise. It would stay only in the position of speculations and abstact terms, words and numbers that would say very few.

One of the basic criteria in decision making is the profitability. How does the change of entropy influence the economical result of the enterprise will be shown on an example from the managerial practice. Values from this example were also used by calculations in the subhead 2.2.

The manager administrates an enterprise that offers transportation services. The costs related to one unit of the service are 30 €. There would be no extra costs, if the estimation was the same as the real value. In this case only the costs necessary for prestation of demanded amount of service would be present. By reason that the estimations have always some deviation, also other two kinds of expenses will be present.

Overestimated forecast causes too big potential ready for the prestation of the service. This potential will not be used. It can be taken as one fifth of the costs of the prestation of the service. Extra costs will be in this case enumerated according the formula (forecast - real value) * 6 €.

In the case of underestimated forecast there will not be sufficient capacity ready for the work. It may cause delays of prestation of the service and loss of the trust of customers. It will be assumed, that the extra costs caused by the underestimated forecast will be one fourth of the cost for prestation of one service unit. Extra costs will be calculated according to the formula (real value - forecast) * 7.50 €.

Table 3 shows real values, forecasts on the basis of empirical estimation and forecasts made by using of prediction algorithm. This table shows also the extra costs in Euro caused by these estimations.

Table 3

Month	Real value	Empirical estimation	Extra costs	Prediction algorithm	Extra costs
I.99	23069	36271	79212	28612	33258
II.99	24989	39732	88458	29521	27192
III.99	28995	43989	89964	29916	5526
IV.99	28215	40300	72510	31505	19740
V.99	32361	43989	69768	30601	13200
VI.99	32206	40470	49584	32497	1746
VII.99	27816	41819	84018	32040	25344
VIII.99	33438	41819	50286	30987	18382.5
IX.99	33071	40470	44394	33711	3840
X.99	38340	41819	20874	33749	34432.5
XI.99	35057	40470	32478	37055	11988
XII.99	33846	41819	47838	33499	2602.5
Total:			**729384**		**197252**
I.05	29547	29347	1500	31998	14706
II.05	28588	26764	13680	31552	17784
III.05	31593	29845	13110	32663	6420
IV.05	33729	28818	36832.5	33416	2347.5
V.05	32962	29347	27112.5	33531	3414
VI.05	30975	28818	16177.5	33925	17700
VII.05	29263	29347	504	30622	8154
VIII.05	30815	29845	7275	30785	225
IX.05	30810	28818	14940	31644	5004
X.05	31576	29347	16717.5	33494	11508
XI.05	31199	28818	17857.5	31470	1626
XII.05	30621	29845	5820	30445	1320
Total:			**171526.5**		**90208.5**

From the table 3 it is evident that the extra costs by using neural networks are high too. But if they are compared with the costs risen by the empirical estimation, the savings are evident. The decision of the manager that is made on the basis of scientifical prediction algorithm brings besides the decreasing of the entropy also a significant decreasing of the costs. The differences between both kinds of extra costs were lower in the year 2005. It was caused by other character of the process that was in the year 2005 relatively smooth and easily predictable.

3. Conclusion

Entropy as the measure of information uncertainty affects negatively the economical results of the enterprise. If there is a possibility to reduce the entropy by using some sophisticated scientific methods, it is advisable to do it. It brings positive effects in the process of managerial decision making. Management as an interdisciplinary scientific area has today a lot of possibilities to incorporate the research results and tools from other scientific areas. The fusion of management and prognostics brings in managerial practice an competitional advantage in the form of better planning and from that resulting lowering of the costs. The advanced prediction algorithms help the manager in a significant way to reduce the entropy in his decision making and to increase the effectiveness of decisions.

References:

[1] ČERNÝ, J.: Entropy and information in cybernetics, ALFA Bratislava, 1981.

[2] GAYNOR, P., KIRKPATRICK, R.: Introduction to time-series modeling and forecasting in business and economics. McGraw-Hill, 1994.

[3] KOSKO, B.: Neural Networks and Fuzzy Systems, Prentice Hall, 1992.

[4] Magazine TREND, number 4, year 2003, pp. 39.

[5] MADARÁSZ, L.: Stochastic processes and theory of informations, Alfa, Bratislava, 1984.

[6] MARČEK, M., PANČÍKOVÁ, L., MARČEK, D.: Ekonometrics and soft computing, EDIS ŽU in Žilina, 2008, ISBN 978-80-8070-746-0.

[7] MONTGOMERY, D., JOHNSON, L., GARDINER, J.: Forecasting and Time Series Analysis, 1990, McGraw-Hill.

[8] DA SILVA, A.P., MOULIN, L.S.: Confidence intervals for neural network based short-term loadforecasting, Power Systems, IEEE Transactions on, Volume 15, Issue 4, 2000, pp. 1191-1196.

[9] STUCHLÝ, J.: Mathematics IV, Alfa, Bratislava, 1982.

[10] ŠNOREK, M.: Neural networks and neurocomputers, ČVUT Praha, 2004.

[11] VAJDA, I.: Theory of information and statistical decision making, Alfa, Bratislava, 1982.

Contact address:

Ing. Michal Fraňo
University of Žilina
Faculty of Management Science and Informatics
Slovak Republic
Email: michal_frano@yahoo.com