AN AUTOMATED DATA PROCESSING SYSTEM TO SUPPORT THE DECISION MAKING PROCESS IN THE INSURANCE INDUSTRY

Kazimierz OGONOWSKI, Krzysztof J. LATAWIEC

Abstract: This paper presents an analytical methodology for automated data processing to support decision-making processes, in particular in large-size insurance companies. The methodology is based on logical, multiple nested formulae implemented in MS Excel, which is a software of choice in insurance companies cooperating with their brokers. An Automated Data Processing System is implemented under Excel to validate massive insurance data circulation, support data correction, analyze data and present results of the analyzes. The performance of ADPS is excellent, both in terms of the very high accuracy of data validation and, in particular, very short times of delivering the results, in which this unique system largely outperforms previously used procedures based on manual data processing.

Keywords: insurance industry, decision support systems, MS Excel, insurance data processing.

Introduction

Insurance is an extraordinarily useful tool to manage risk [1]. There are three principal purposes of insurance: to provide information via premiums as to how serious the risk is; to provide motivation for undertaking financial protection against an event that could produce a significant loss but has a low probability of occurrence [2,22]; and to offer incentives in the form of premium reductions to reward people who invest in risk-reducing measures. Roughly speaking, the roles the insurance is designed to play include reducing future losses and financially protecting those at risk [3].

This paper considers some important aspects of decision-making processes in operation of insurers, in terms of improving the collaboration between insurance companies and insurance brokers. In fact, smooth cooperation of insurance companies with brokers is the key factor in possible financial success of both the insured and insurers and this is because the transfer of massive data between the insurance companies and brokers can only be effectively managed by means of sophisticated computer algorithms [4,5]. A decision making process is one of the most important and complicated human activity. Unfortunately, decision making algorithms can be defined only for some kinds of (simple) management problems. In more complex cases, an "optimal" or "universal" algorithm for decision making does not exist, even though sophisticated artificial intelligence tools have increasingly been applied. Such reasons can limit the possibility of construction of an automated decision making process [6,7,8]. Therefore, decision making processes, in particular in the insurance industry, are usually operated in the management support mode as decision support systems.

Whilst complex computer science tools, including data-base models and artificial
intelligence machinery used by the actuaries constitute the "brain" in the prediction world environment at the top management level of a large insurance company [2,4], the "heart" of a massive data circulation system between the brokers and the insurance company at its medium management level is the Microsoft Excel environment overwhelmingly operated and commonly accepted by the brokers [9,10,4,11,12,13]. In fact, in their everyday rush with customers, brokers could not accept any other, more complex programming environment. Perfecting this Excel-based "heart" of the insurance data circulation system for the purpose of a decision support process at the medium management level is the main goal of this paper.

1. General description

Handling massive amounts of policy-related data at the medium management level [14] has posed significant challenges for the insurance industry. In this specific case, the detailed data pertaining to policies/programs sold by insurance brokers (representing insured customers) served as the input data collected by and stored in a company database. These data consisted of variables such as losses (payments on claims), as well as revenues based on collected premiums, etc. Such data were then subject to different processing and analyses resulting in information outputs that in turn served for decision making purposes, collaboration with the brokers, and the type and quantity of new insurance program initiatives, etc. [15,16,17].

An input to a decision making system is a policy-related, massive data coming both from the insurance company management and, in particular, the collaborating brokers. In the former case, data includes monthly, raw (invalidated) policies' details such as their numbers, the assigned premiums (new and renewed ones), losses, expiry dates, and possible policy cancellation notifications, without divisions to particular brokers and their campaigns (business branches) [18]. In the latter case, each campaign-assigned broker provides (only once) on-line data concerning newly-written policies, in terms of their numbers, booked premiums, and expiry dates. Since there are hundreds of brokers handling thousands of policies, the raw data can be biased with numerous, sometimes financially detrimental errors.

Therefore, the first important task of an innovative decision support system to be introduced here is to automatically validate the incoming data and support its correction. Until the time of this system development, manual validation and correction procedures were time-consuming and suffered from low accuracy of validation (and correction). Additionally, the management's manual development of monthly/yearly summaries of (narrow-range) results of brokers' operations, in terms of policies' premiums, losses and written loss ratio (WLR), could take, together with data validation and correction, up to four weeks.

Currently, the system automatically validates the incoming data and supports its correction, and also automatically provides summaries of wide-range results of the brokers' operations, including six more broker assessment criteria, with the processing time reduced to a few seconds only.

The above features of the innovative system are extremely supportive for the management's decision making process, whose task is to propose control actions related to the improvement of the brokers' operations, including a development of appropriate corrective decisions onto underperforming brokers. Until now, the "rectifying" control means for brokers, based on manual data validation/correction and manual broker
assessment procedures, were unreliable and time-consuming, causing inconsistencies between the brokers and the management [19,20]. The innovative decision support system eliminates such issues, not only thanks to the unprecedented accuracy of the validation/correction system but also owing to its ultra high-speed reaction to any emergency situation. In fact, the validation/correction system is practically error-free and time intervals for automatic data validation/correction and generation of broker assessment results achieved in the system are currently in the order of seconds.

The pioneering system introduced here to the academic world and published by the national Patent Offices both in the US and Canada [15,16], efficiently automates the processing of such data for the purposes of improved decision making pertaining to the servicing of insurance policies by brokers supporting a large insurance company. A contribution of this work is that, for the first time ever, the sophisticated mathematical formulae involved in the information processing technology are entirely automated [15,16,17]. This yields unprecedented acceleration of the decision supporting processes, in addition to high, error-free in fact, accuracy of the process. The automated computation algorithms developed contribute to the ultimate success of the brokers' operation and the very high appreciation from the management of the insurance company.

The key to the success of the automatic system is an original, analytical approach embedded in the Microsoft Excel environment. MS Excel has been vastly used in the insurance industry [12]. As a matter of fact, an Excel workbook is likely the most acceptable computing tool for brokers, not to say an unbreakable constraint in the majority of broker-related applications. Therefore, this involvement has been aimed at analytical, function-oriented features of the Excel workbook, that are largely recognized by the brokers [11]. Utilizing the automated MS Excel workbook immediately reflecting the company's particular results as opposed to other programming systems is faster, more dependable and efficient, and easier to create and modify. Any changes in references automatically update results of the formulae. Such enhancements to the information flow, thanks to this particular system, can translate into potentially significant costs savings to the company by way of arming the decision makers with timely information input, and increasing overall efficiency and competitiveness of the organization. The system facilitates data manipulation. Manipulating the data, in turn, generates useful input into improved strategic decision making. The system is automated (by using the advanced formulae) allowing for immediate results generation. These output results, in the insurance industry case, include variables such as premium or loss amounts or the particular texts and visually displayed alerts of mistakes made by insurance brokers and other company systems. Having results such as premiums, losses, loss ratio, etc. allows for monitoring and analyses of performance of particular brokers, in terms of the profits for specific campaigns such as concrete contractors, auto parts, etc. [15,16,17].

2. Decision support system as feedback-feedforward control system

Management and control are quite similar notions, even though they have been grown on different application fields. Making decisions for the purpose of (broadly understood) management is a sort of computation of control decisions by a controller. Both management decisions and control decisions are typically executed in a feedback, or closed-loop system, hopefully supplemented with feedforward paths arranged on measurable/countable disturbances acting upon the management/control plant. Fig. 1 presents a feedback-feedforward decision making system at the medium management level of the plant, which
is an insurer (insurance company + broker). The system is arranged in a control-like fashion, with specifying the system variables as follows:

- a control/management decision \( u_\text{d}(k) \) feeding the plant, where \( k \) is the discrete time, with the sampling interval equal to the company's reporting period (at the medium management level), that is one month,
- controlled output \( y(k) \) from the plant,
- output reference (or setpoint) \( y_0(k) \),
- error \( \varepsilon(k) = y_0(k) - y(k) \) between the reference and controlled output,
- measurable/countable disturbance \( d_\text{m}(k) \),
- unmeasurable/uncountable disturbance \( d_\text{un} \).

Fig. 1. Feedback-feedforward control/management system

Elements of a control/management decision vector \( u(k) \) constitute the corrective decisions onto underperforming brokers:

- shifting the broker(s) to some more suitable campaigns,
- recommending additional training sessions to the broker(s),
- suspending the existing business arrangements with the broker(s),
- disconnecting the existing business arrangements with the broker(s).

Now, the size of the vector \( u(k) \) is \( n_u=4 \). It is important however that an actual control decision vector is always of a lower size, e.g. 3 or even 1. This is because, typically, not all the "rectifying" control means for underperforming broker(s) are simultaneously applied. Moreover, there are often situations that there are no underperforming brokers during monthly reporting periods (corresponding to \( n_u=0 \)), so the control decision is "no change in decision".

The company's decision making process at the medium management level is subject to a number of unmeasurable/uncountable disturbances \( d_\text{un}(k) \) related with altering financial situation of the company and its environment on the one hand, and a market-sensitive tissue of customers on the other. However, the main source of the disturbances is just the massive policy-related data itself. With hundreds of brokers operating on thousands of policies, the data transferred between the brokers and management and vice versa is inevitably prone to
human-originated errors. The sources of those errors are specified and commented in the sequel. Elimination of those errors by an automated data processing system is a challenging goal of this work.

The control/management task of the decision maker of Fig.1 is to produce decisions \( u(k) \) in order to keep the error \( \varepsilon(k) \) as close to zero as possible. We have shown above that even precise specification of the control input \( u(k) \), controlled output \( y(k) \) and measurable/countable disturbance \( d_m(k) \), in terms of determining both sizes of the vectors and their specific elements, is hardly possible here because of the very uncertain and unclear matter of the medium level management environment. Unfortunately, the problem is empowered by the inability to precisely define the setpoint values for the output reference \( y_0(k) \). In fact, it is not possible to precisely demand for a reference value of e.g. WLR; it should be as low as possible, but the practice is that in some critical cases it would be satisfied when, temporarily, this value is equal to even 100% and it does not increase. Quite similar is with the reference values for total premiums, which clearly do not exist. From business perspective the values should be as high as possible, but in some exceptional situations it might be acceptable when the premiums fall down provided that losses get lower.

In general, a control/management law could be as follows

\[
\begin{align*}
    u(k) &= u_f(k) + u_p(k) \\
        &= \phi_1[\varepsilon(k)] + \phi_2[d_m(k)]
\end{align*}
\]

where \( u_f(k) \) and \( u_p(k) \) are the feedback and feedforward components of the control vector \( u(k) \) and \( \phi_1[\varepsilon(k)] \) and \( \phi_2[d_m(k)] \) are certain (vector) functions which, in the classical control framework, could be obtained on a basis e.g. of some control performance criterion. However, with the above issues of time-varying sizes of the vectors \( y(k) \) and \( d_m(k) \) (and their elements sometimes overlapping), in addition to a blurry determination of the reference components \( y_0(k) \), a formulation and solution of any optimization problem is, most likely, unfeasible, except perhaps of some trivial, in fact unreal situations.

All the above is intended to demonstrate that it is most likely impossible to construct an automated expert decision maker that could effectively do the control/management job as theoretically required in Fig. 1. The extremely complex matter of a decision making process at the medium management level of a large insurance company can only be, most likely, implemented in the advisory, decision support mode, the solution being overwhelmingly accepted in the management practice. In fact, there is no example world-wide that such an automated decision-making system has been implemented. An extremely simple example of a fuzzy-like realization of the decision-making process by a human expert is: if “written loss ratio (WLR) in %, for new and renewed policies, per broker, for the current year to date” is “low” or “low enough” and “total premiums in $, for new and renewed policies, per broker, for the current year to date” are “high” or “high enough” then “no change in decision”. Even if the fuzzy-like reasoning [21] could be attractive in possible development of an automated decision making process at the medium management level in a large insurance company, it has still never been implemented in such a company world-wide. One reason is that, with a poor performance of a manual data processing system supporting the decision making process, a human expert could not be quite sure if those “low enough” or “high enough” quantifiers were not falsified. And without them, the above expert, fuzzy-reasoning example would be too trivial in the insurance practice.
3. Location of Automated Data Processing System (ADPS)

The invention constituting the subject of this dissertation is an Automated Data Processing System (ADPS) whose location in the feedback-feedforward control/management system of Fig. 1 is shown in Fig. 2. The system collects massive data from the insurer, that is the insurance company and brokers, and performs a series of analytical, Excel-based data validation procedures followed by error correction supporting schemes and presentation of reports on summary of the results (incorporated in the output $y(k)$). Consequently, the quality of the output $y(k)$, supplied from ADPS to the Decision Maker, is very high as compared to that for manual processing, both in terms of the accuracy and speed of processing. As for ADPS itself, it is composed of a number of Excel-implemented formulae representing both feedback (closed-loop) and feedforward (open-loop) operations on data.

![Fig. 2. Arrangement for Automated Data Processing System (ADPS)](image)

4. ADPS – System Interworkings

In order to monitor the effectiveness of targeted marketing, it is desirable to analyze the performance or profitability of the policies generated using the marketing and sales application. Policy data is stored in the policy administration system that is separate from the marketing and sales application executed on the web server. The policy data includes, but is not limited to, policy renewal data, policy premium data, claim data and policy termination data. By combining the policy data with the marketing and sales data, the performance of the policies generated as a result of such campaigns can be measured. The policy performance can be summarized by broker, by campaign, by year, etc.

Fig. 3 shows a Microsoft Excel workbook 800 into which the policy data and marketing and sales data are imported for analysis. The workbook 800 has a plurality of worksheets, among which are a data import sheet 804, a set of annual summary sheets 808, a policy summary sheet 812, a campaign summary sheet 816, a broker summary sheet 820, a new policies summary sheet 824 and a broker summary sheet 828 broken down by their campaigns. The data import sheet 804 receives policy data that is inputted for processing. Each of the annual summary sheets 808 looks up data for the appropriate years in the data import sheet 804, validates it and summarizes the projected and actual policy activity for each broker/campaign combination for a particular calendar year. Each of the policy summary sheet 812, the campaign summary sheet 816, the broker summary sheet 820, the new policies summary sheet 824 and the broker summary sheet 828 broken down by their campaigns references the data processed in the annual summary sheets 808. In turn the
policy summary sheet 812 provides an overall analysis of the performance of the policies generated as a result of all of the campaigns.

![Schematic diagram of a Microsoft Excel workbook for analyzing the performance of policies generated as a result of the targeted marketing method](image)

Fig. 3. Schematic diagram of a Microsoft Excel workbook for analyzing the performance of policies generated as a result of the targeted marketing method

The campaign summary sheet 816 summarizes the performance of the policies generated as a result of each campaign by calendar year of issuance. The broker summary sheet 820 summarizes the performance of the policies generated as a result of each broker by calendar year of issuance. The new policies summary sheet 824 summarizes the performance of the policies generated as a result of each campaign and by each broker that are newly issued in the current calendar year. The broker summary sheet 828 broken by their campaigns summarizes the performance of the policies generated as a result of each broker by calendar year of issuance.

5. Data validation formula

As can be seen, the uniquely created formula 900 uses a set of nested IF (condition, then, else) functions to serially perform a number of checks on the data [15,16]. In a first condition 902, it is determined whether the number of records in the data import sheet 804 that match the policy with which the formula 900 is associated is not equal to one. The first condition 902 checks all of the policy numbers in the data import sheet 804 using an array calculation. If no records in the data import sheet 804 match the policy number with which the formula 900 is associated, data for the specific policy number is deemed to be missing. Alternatively, if two or more records in the data import sheet 804 match the policy number with which the formula 900 is associated, data for the specific policy number is deemed to be suspect and unreliable. In both cases, any data that could be returned by a formula may not be accurate and it is desirable to correct the data instead of relying on potentially inaccurate results reported by the workbook 800. As a result, if the first condition 902 is true, a first error message 904 is visually displayed.
If the first condition 902 is false, i.e., no error is returned, a second condition 906 is checked. The second condition 906 checks all of the policy numbers in the annual summary sheet 808 for the 2000 calendar year using an array calculation. If no records in the annual summary sheet 808 for the 2000 calendar year match the policy number with which the formula 900 is associated, data for the policy number present in the data import sheet 804 may not be properly reported through the annual summary sheet 808. Alternatively, if two or more records in the annual summary sheet 808 match the policy number with which the formula 900 is associated, data in the data import sheet 804 for the specific policy may be reported more than once, potentially leading to inaccurate results. Once again, in both cases, data that could be returned by the formula may not be accurate and it is desirable to flag the data as requiring attention instead of relying on potentially-inaccurate results reported by the workbook 800. As a result, if the second condition 906 is true, a second error message 908 is visually displayed.

If the second condition 906 is false, i.e., no error is returned, a third condition 910 is checked. The third condition 910 checks to see if the record in the data import sheet 804 corresponding to the policy number with which the formula 900 is associated is marked as a cancelled policy. If the record in the data import sheet 804 corresponding to the policy number with which the formula 900 is associated is marked as cancelled, a cancelled policy message 912 is visually indicated to denote that the policy has been cancelled.

If the third condition 910 is false (that is, none of the records in the data import sheet 804 corresponding to the policy number with which the formula 900 is associated are marked as cancelled), a fourth condition 914 is checked. The fourth condition 914 checks to see if the record in the data import sheet 804 corresponding to the policy number with which the exemplary formula 900 is associated has an appropriate policy anniversary date. If the record in the data import sheet 804 corresponding to the policy number with which the formula 900 is associated does not have an appropriate policy anniversary date (that is, if the fourth condition 914 is not true), a third error message 916 is visually indicated.

If the fourth condition 914 is true, a fifth condition 918 is checked. The fifth condition 918 checks to see if the record in the annual summary sheet 808 with which the formula
900 is associated includes the appropriate policy issue date. If the record in the annual summary sheet 808 with which the formula 900 is associated does not include the appropriate policy issue date, a fourth error message 920 is visually indicated.

If the record in the annual summary sheet 808 with which the formula 900 is associated includes the appropriate policy issue date, a result 922 is calculated and indicated. The result 922 is the booked premium for the 2000 calendar year retrieved from the corresponding record in the data import sheet 804. Note that the records for which result 922 is calculated do not contain the text string "Cancelled policy", but may contain other text strings. During the calculation of result 922, an intermediate result is firstly determined by multiplying the calculation by one. If the intermediate result is a number, the result 922 is simply equal to the intermediate result. If, instead, the intermediate result contains a text string, a fifth error message, "#VALUE", relating to the calculation error in the cell is displayed.

For a record containing a text string, the result 922 would be equal to zero if the intermediate result is not multiplied by one. This would cause a zero premium to be calculated for records containing unexpected text strings. It is therefore desired to flag records containing unexpected text strings so that the unexpected text strings can be examined and rectified. Multiplying the intermediate result by one allows for differentiating unexpected text strings and number 0. Otherwise it would not be known if a different text string than "Cancelled policy" or premium (or loss) amount equal zero was assigned to the policy number in data import sheet 804.

In this manner, multiple error checks are performed on the consistency of the data from the marketing and sales application and the policy data from the policy administration system and provide a visual indication of errors present therein.

The following presents the overall mathematical notation of the formula 900 (Fig. 4) as a three-argument multiple nested function:

\[
=\text{IF}(x_1, y_1, \text{IF}(x_2, y_2, \text{IF}(x_3, y_3, \text{IF}(x_4, y_5, z_5)))
\]

where \( x_1 \) denotes logical_test 902, \( y_1 \) – value_if_true 904, \( x_2 \) – logical_test 906, \( y_2 \) – value_if_true 908, \( x_3 \) – logical_test 910, \( y_3 \) – value_if_true 912, \( x_4 \) – logical_test 914, \( x_5 \) – logical_test 918, \( y_5 \) – value_if_true 922, \( z_5 \) – value_if_false 920, \( z_5 \) – value_if_false 916.

An analytical model (1) represents the first example of application of the introduced analytical methodology for automated data processing, in particular data validation. The formula (2) is a general form of description of formula 900.

6. Second extremum formula

The following presents an outline of the originally created method for determining a secondary value [15,16,17].

A method for determining a secondary value having a second-highest priority in a list of values in a spreadsheet is provided. A primary value having the highest priority in the list of values is determined. The secondary value is determined to be equal to the primary one if more than one of the values in the list is equal to the primary value. The secondary value is determined to be equal to one of the list of values not equal to the primary value having a relative highest priority if only one of the values in the list is equal to the primary value.
Fig. 5. The formula used to determine the second-highest loss.

Fig. 5 illustrates the uniquely created formula 1012 calculating the second-highest value (loss) as the result of applying the same overall concept as in the above-outlined method [15,16].

The first condition 1014 is used to select those entries corresponding to a particular broker. The second condition 1016 is used to determine whether there is more than one loss equal to the overall maximum single loss for the particular broker. The condition is determined by checking how many entries in the maximum single loss and second-highest single loss columns are equal to the overall maximum single loss for that particular broker using an array function. If it is determined that there is more than one loss in the maximum single loss and second-highest single loss columns equal to the overall maximum single loss for that particular broker, the overall second-highest single loss is set equal to the overall maximum highest loss at 1020. Otherwise, those loss entries for the broker "B4" that are not equal to the maximum single loss are located in the maximum single loss and second-highest single loss columns using the condition 1024, and the highest value from these entries is selected as the overall second-highest single loss at 1028. This is, again, determined using an array function.

Usually the second highest value is assumed to be smaller than the highest value unintentionally disregarding the case in which there is more than one value exactly equal to the highest value. In this specific case, the formula 1012 returns the second highest value exactly equal to the highest value as the final formula result and simultaneously the important notification that there is more than one of the exactly same highest values is automatically provided.

The following presents the overall mathematical notation of the formula 1012 (Fig. 5) as a multiple nested function:

\[ = \text{MAX}(\text{IF}(x_1, \text{IF}(x_2, y_2, \text{IF}(x_3, y_3)))) \]  

where \( x_1 \) denotes logical_test 1014, \( x_2 \) -- logical_test 1016, \( y_2 \) -- value_if_true 1020, \( x_3 \) -- logical_test 1024, \( y_3 \) -- value_if_true 1028.

While a method for calculating the overall second-highest loss has been described, the same methodology can be used to calculate a secondary value having a second-highest priority in a list of values using other prioritizations. For example, the following shows the
uniquely created formula calculating the second lowest value as the result of applying again the same overall concept as in the method for determining a secondary value having a second-lowest priority:

\[
=\text{MIN}(\text{IF}(\$A9:\$A48=\"B4\",\text{IF}(\text{SUM}(\text{IF}(\$A9:\$A48=\"B4\",\$H9:\$I48)=\text{MIN}(\text{IF}(\$A9:\$A48=\"B4\",\$H9:\$I48)),\text{TRUE}+0)))>1,\$H9:\$I48,\text{IF}(\$H9:\$I48<\text{MIN}(\text{IF}(\$A9:\$A48=\"B4\",\$H9:\$I48)),\$H9:\$I48)))
\] (4)

The above-presented formula, in the specific case in which there is more than one value equal to the lowest value, returns the second lowest value equal to the lowest value as the final result without omitting any of the significant values.

As shown above, this formula is also obtained by changing only the prioritization substituting MIN for MAX in the formula 1012 (Fig. 5) leaving everything else unchanged. This indicates that the method and the structure of the formula are very universal.

An analytical model (3) represents another example of application of the introduced analytical methodology for automated data processing, in particular calculation of the second maximum value. The formula (3) is a general form of description of formula 1012, with the prioritization modification of the latter for the calculation of the second minimum value as specified in formula (4).

**Summary of System Interworkings**

The Automated Data Processing System contains very advanced mathematical formulae, a number of which are originally/uniquely created as recognized by the American and Canadian patent offices. The set of formulae represents an analytical approach, or analytical methodology for development of data processing system to support a decision-making process, in particular in the insurance industry. The workbook contains hundreds of thousands of linked formulae and their number increases in each subsequent year that the workbook is employed.

It is finally worth of mentioning that the presented analytical methodology is based on logical formulae which sometimes resemble the so-called production rules used in expert reasoning [11,19], thus providing the methodology with 'a sort of' intelligence.

**References**

15. Ogonowski K.: Canadian Intellectual Property Office published on September 15, 2007 under Publication No. CA 2539652
22. Świątek J., Borzemski L., Grzech A., Wilimowska Z. (Eds.) Information systems architecture and technology: knowledge based approach to the design, control and decision support, Wrocław, Oficyna Wydawnicza Politechniki Wrocławskiej, 2013.

Mgr Kazimierz Ogonowski
Prof. dr hab. inż. Krzysztof J. Latawiec
Instytut Automatyki i Informatyki
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Opolska
45-72 Opole, ul. Sosnkowskiego 31
e-mail: kaz.bro.ogonowski@gmail.com
k.latawiec@po.opole.pl