STATISTICAL DATA ANALYSIS COURSE VIA THE MATLAB WEB SERVER

Aleš LINKA
Dept. of Textile Materials, TU Liberec
Hálkova 6, 461 17 Liberec, Czech Republic
e-mail: ales.linka@vslib.cz

Petr VOLF
Dept. of Applied Mathematics, TU Liberec
Hálkova 6, 461 17 Liberec, Czech Republic
e-mail: petr.volf@vslib.cz

ABSTRACT

Internet based courses are at present a quite common tool of learning. While still the most of them consist just of the static texts located on web pages, the objective of many new instruction systems is to make the learning process more dynamic and interactive in comparison with a mere reading the textbook or listening the lessons. Moreover, it is a feature of mathematics that it can hardly be studied by mere memorizing the texts. That is why we searched, when preparing a series of internet courses on mathematical statistics, data analysis and quality control, for an environment enabling such an interaction and supporting the preparation and use of numerical and graphical procedures. Finally we decided to utilise the Matlab Web Server. In this environment, the author can combine the text with Matlab computational algorithms and graphical tools, the user can work with them without having its own Matlab installation. The system thus consists from text files, the Matlab programs and the procedures controlling the interface, input and output connection between the web pages accessed by the user and the Matlab algorithms (prepared by the author or used directly from Matlab toolboxes). In such a way, a student is provided simultaneously with relevant information, the examples, and graphs, he can enter his own data, too, and is offered the tests checking his knowledge. In the present contribution the system will be described and examples of its use presented.

Key words: Mathematical statistics, tutoring system, Matlab Web Server.

1. Web based tutoring systems

While the first web pages were constructed about 10 years ago, the internet is today used quite commonly and the area of possible applications is growing tremendously. As the main benefit is the information sharing and processing, one of remarkably developing fields are the internet courses and (e-) learning. These procedures or systems can offer a large amount of information in properly selected and structured form, but they also gather the information on users. Some of systems are able not only to recommend the path through the lessons, but also to control it and adapt on the basis of user's skill and progress (ITS – Intelligent Tutoring Systems). These systems contain elements of AI, in order to classify the user (User Modelling) and to control his path, including the recognition and remediation of insufficiently mastered lessons.

Thus, the construction of an ITS needs the intensive effort of a whole team of specialists. One example of such a system is the Net-Coach (www.net-coach.de), which is offered also as an "empty" system to be filled with an appropriate information, contents. On the other side, the most of (up to now) available applications deal with topics like a language course, basic courses of work with PC or programming, information of an "encyclopedic" nature, or drivers' tests (e.g. www.neuralgen.cz), though even these applications contain often certain elements of UM and adaptation methods (on the other hand, it should be said here that the 'guided adaptive tour' through the lessons was already the feature of so called "programmed textbooks" many years ago).

The development of tutoring systems in the field of mathematics is complicated by the need of simultaneous usage of text, computation, and graphics. It seems that the relatively direct way can lead from a professional mathematical software package, especially when it already has its Web version, so that the large portion of technical work has already been done. It 'remains' to make it more didactive, to change the help to tutoring texts, to add the procedures of control and adaptation (at least recommending the path through the lessons) and to prepare also the tests checking the knowledge of an user. Let us mention here the statistical system developed from the XploRe statistical package (www.md-stat.com).

We decided to use the computational and graphical environment of Matlab and its connection with internet via the Matlab Web Server. The advantage of Matlab consists in the possibility to prepare own algorithms and use them as a part of Matlab library. The programming is easy, so that Matlab is convenient for active formulation of algorithms by students. On the other side, the implementation of MWS has also disadvantages (for authors of tutoring system), because the interface of all parts (sharing the information, input and output transfers) is not so straightforward as we expected.

The applications of MWS combine Matlab m-files, graphics, and HTML texts, resp. PHP and Java scripts. The programming of each such application consists of several parts:

1. Development of HTML files enabling the transfer of input data and parameters of programmed procedure. As a rule, it is through a frame on client's display.
2. Development of Matlab m-file, which, except that it solves required computational or graphical problem, reads the data from the input HTML frame, and prepares the output field.
3. Then, other HTML procedure transfers the output to the output page or window visible on client's display. The graphs are in .jpg format and are called by the parameter – figure's name.

2. Actual contents of the course

The objective of the system is to provide:

- The tutoring text, including the theory and formulas, and practical recommendations.
- Illustrative examples, figures. The numerical examples are generated randomly. The user, in some cases, is challenged to solve the example independently and to check the correctness of his approach afterwards.
- The examples, exercises testing the level of knowledge of the user.
- Statistical data analysis "calculator" which can process even the data from the file on user's computer.
- Finally, a set of Matlab algorithms, developed by authors (and, naturally, the standard Matlab procedures and tool-boxes which are a part of Matlab installation).

At present, the material covers several chapters of basic courses on probability, mathematical statistics and quality control. Namely

1. Distribution of random variables. The most frequent types of distribution (both discrete and continuous) are presented both mathematically and graphically, their typical use is demonstrated on simple examples.
2. Explorative data analysis (EDA) methods. This part deals with basic empirical characteristics of distribution of given data, their computation and graphical presentation with the aid of several types of plots.
3. Statistical tests of hypotheses. Again, the main methods of both parametric and nonparametric tests are recalled, its methodology explained and illustrated on examples.
4. Regression analysis presents the material on linear and nonlinear models, the least squares methods and nonparametric smoothing approaches.
5. Control charts provide the motivation, methodology, and examples of Shewhart charts for the mean value and variance, and also the EWMA smoothing method.
6. Methods of quality control in textile industry The theme follows from the specialisation of the Faculty of Textile Engineering of TU Liberec and provides the basic approaches to textile metrology and quality control.
7. The section provisionally named 'Matlab-Web' contains a selection of examples to all topics.

The orientation in the system is provided by the menu of themes and sub-themes. The main page is shown on Figure 1. Thus, the user can either select its own theme (sequence of themes) or he can follow the recommended sequence (which actually corresponds to the order of menu bars) and links to examples in the text of textbook chapters.

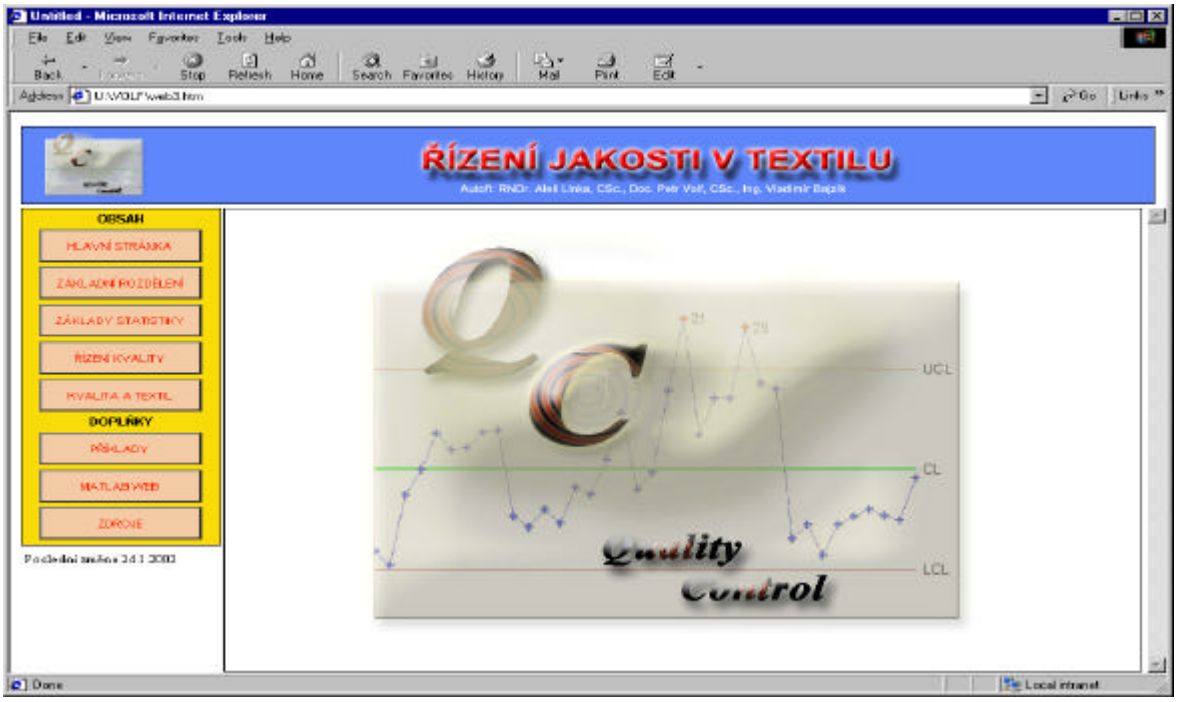

Figure 1. The main panel

3. Example – control charts

Let us assume that we are interested in a brief tutorial on control charts. First, we can follow the menu bars and select tutoring text under Quality Control / Control Charts. There, a student obtains information (description, definition, purpose, probabilistic background and way of utilisation) of basic charts, among them the Shewhart's control charts for the mean (X-bar), for the standard deviation (S-bar), and also exponentially weighted moving average (EWMA) diagram with one-step delay. Then, the link leads to the computational example. The user is first asked to enter data – and he has four possibilities: He can write the data directly to the corresponding line in the input frame on display, he can use Matlab random generator and generate the data through the MWS, further, he can select one of demonstration data files prepared on the server, and he can also enter his own data from his computer. In the last case the user should be aware (and he is informed in tutoring text) that the form of data should correspond to the procedure – for instance the control charts work with a matrix ($n \times k$, say) of n groups à k observations. Finally, the user is asked to select, in another window, the type of chart (e.g. EWMA together with smoothing parameter). Then the charts are computed and displayed graphically (see Figure 2).

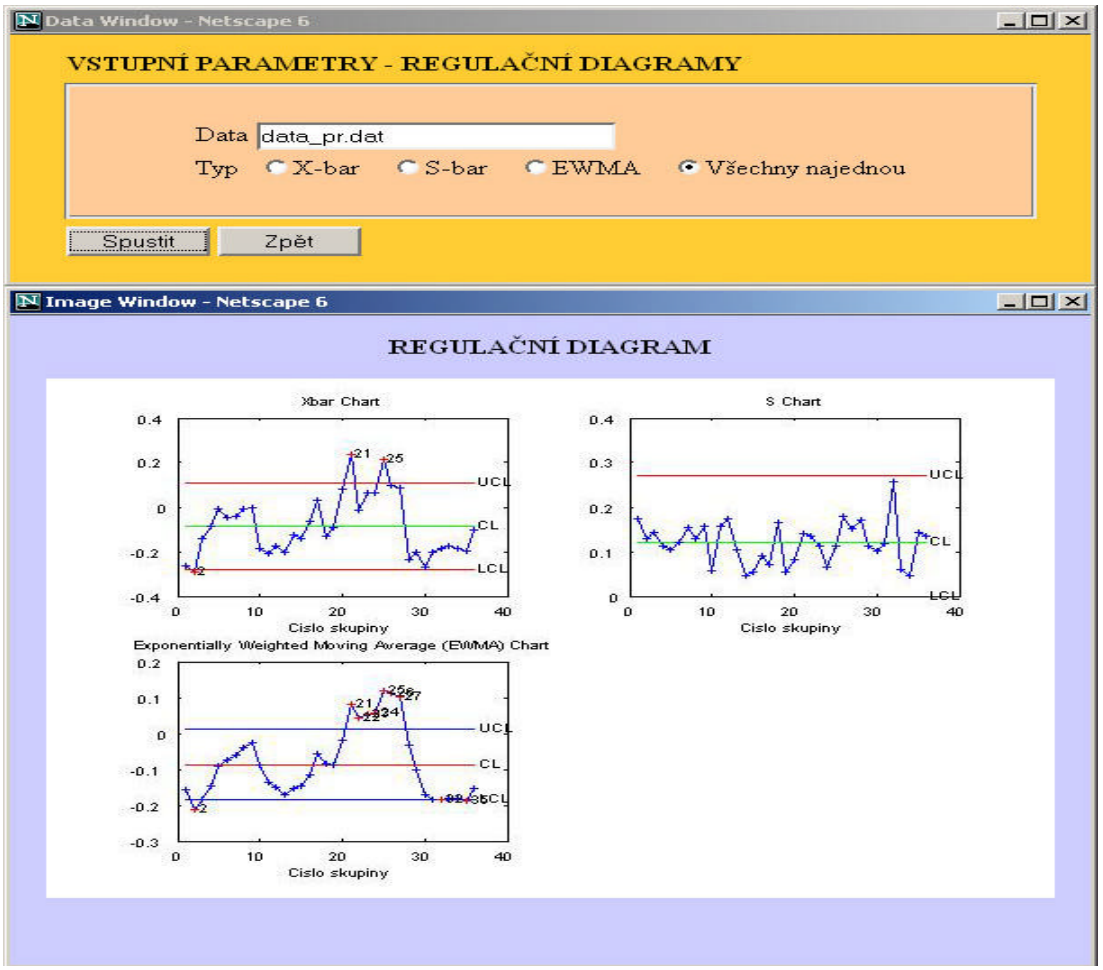

Figure 2. An example – graphs of control charts

3.1 An example of seminar exercise

Let us assume that the theme of the exercise is the polynomial regression, including the selection of optimal model. The teacher provides the instructions, the data are artificial, prepared in advance and stored on the server (for students home work) or they can be generated by students, again, according to an instruction, for example:

1. Generate the data (e.g. a complicated sinusoidal function with Gaussian random noise).

The following instructions are:

2. Plot the data, with the aid of polynomial regression procedure.
3. Select the degree of polynomial.
4. Evaluate the regression model, analyse the significance of its parameters. The procedure computes regression parameters, corresponding 95 % confidence intervals, residual variance, and also Schwarz's BIC criterion penalising the models with too high degree. Students should already know the meaning of all these variables and parameters. The next step is:
5. Change the degree of the model and repeat the analysis. Compare the results, with special attention to the values of BIC criterion.

… etc.

Finally, the procedure evaluates the model once more, omitting statistically non-significant parameters (and model components). The student is expected to perform the whole exercise, to write a report, and, eventually, to demonstrate and comment the process and results of solution to

the others. Naturally, he is encouraged to utilise the relevant parts of chapter 'Regression' from the text stored on MWS. He can also be recommended to draw an additional information from other sources, textbooks.

4. Conclusion

As it has already been said, the development of applications using the Matlab Web Server is technically demanding and time consuming, not speaking about high requirements of programming skill. Moreover, the Web text languages, as a rule, do not support the writing of mathematical symbols and formulas (this is the problem of MS Word, too, at least for people using the "mathematician-friendly" environment of LaTex). Nevertheless, the MWS applications offer many advantages, the essential being the access to Matlab-procedures through the Web browser.

That is why the main purpose of our system is to provide the students the possibility of effective home-learning, because they can gather both theoretical and practical knowledge simultaneously, from the same source, and in proportion convenient to each individual student. Further, the students will use the system for the preparation of their seminar exercises and reports. The system is originally intended for the students of textile engineering in Liberec, it uses Czech language. Simultaneously with system growing (other parts of statistical methodology will be attached soon) its texts will be translated to English, in order to be available also to students of international University Nisa, founded recently in Liberec region. The actual version of the system is accessible through the links from the address: http://147.230.129.170/ales/ucebnice_v3010/.

Acknowledgement: The development of the MWS application has been supported by the project FRVŠ 1010/2001, the work continues at present under project MSM 245 100 303, both granted by the MŠMT ÈR.

REFERENCES

- *Matlab Web Server – The Language of Technical Computing* (manual), The MathWorks, Inc.,2000.
- Wetherill, G.B., Brown, D.W., 1990, *Statistical Process Control*, Chapman and Hall.
- Smid, J., Obitko, M., Volf, P., 2001, Model Parameters and Model Performance. *Proceedings of the UM2001 - Workshop on Empirical Evaluation of Adaptive Systems*, Sonthofen, 25-32.
- Volf, P., Smid, J., 2001, Randomized Evaluation of States of an ITS. *Modeling and Simulation of Systems Conference MOSIS 01*, MARQ Ostrava, 227-232.
- Weber, G., Specht, M., 1997, *User modeling and adaptive navigation support in WWW-based tutoring systems*,. http://www.psychologie.uni-trier.de/projects/ELM/elmart.html.
- *UM01 - User Modeling 2001. Proceedings of the 8th International Conference.* Edited by M.Bauer, P.Gmitrasiewiecz, J.Vassileva. Springer Verlag 2001.