



ASSESSING DIVERSITY IN LEARNERS BACKGROUND AND PERFORMANCE

Anikó Balogh, László Pitlik, Máté Schnellbach, Ferenc Szani, Apertus Nonprofit Ltd., Hungary

Introduction

Assessing diversity in learners' background and performance assumes the observation of learning processes. Observation may be executed both in case of contact learning activities and especially in e-learning systems, where the storing of data logs is an essential part of system maintenance and system analysis.

The observation of human individuals through questionnaires and/or interviews (e.g. collection data about satisfaction levels concerning learning objects, phases, etc.) should always be interpreted on the basis of the parable of the "boiling frog".

Storing logs apparently requires the definition of phenomena (being worth logging) in advance. It is – however – not the only method. Moreover this method as such has inevitable risks. Lack of data in case of newer interpretation needs being generated in the hermeneutical process based on plan-oriented data. Instead of defining data needs in advance, it is possible to create logs which guarantee a high level of reproduction concerning each detail of the learning process (incl. even mouse movements in 1/100 sec rhythm). Reproduction-oriented logs always ensure the chance of being able to follow arbitrary analytical objectives any time. Therefore the quality and efficiency of assessing diversity in learners' background and performance highly depends on the quality of data, where efficiency is considered as a kind of ratio between derived information added value and the resources needed.

In this article, the authors focus on e-learning systems with high level reproduction capacity in the process of logging. Contact study activities are not examined at present, yet blended learning activities (like e-seminars) are also among the targeted features. The aim of this article is (based on real data assets over 10 millions of records but without case-study-like results) to outline an analytical system covering quasi 360° of analytical expectations for a (later automated) decision support service using both classical statistics and artificial intelligence methods for modelling (e.g. robot teachers).

The system

Figure 1 show System elements and their connections.

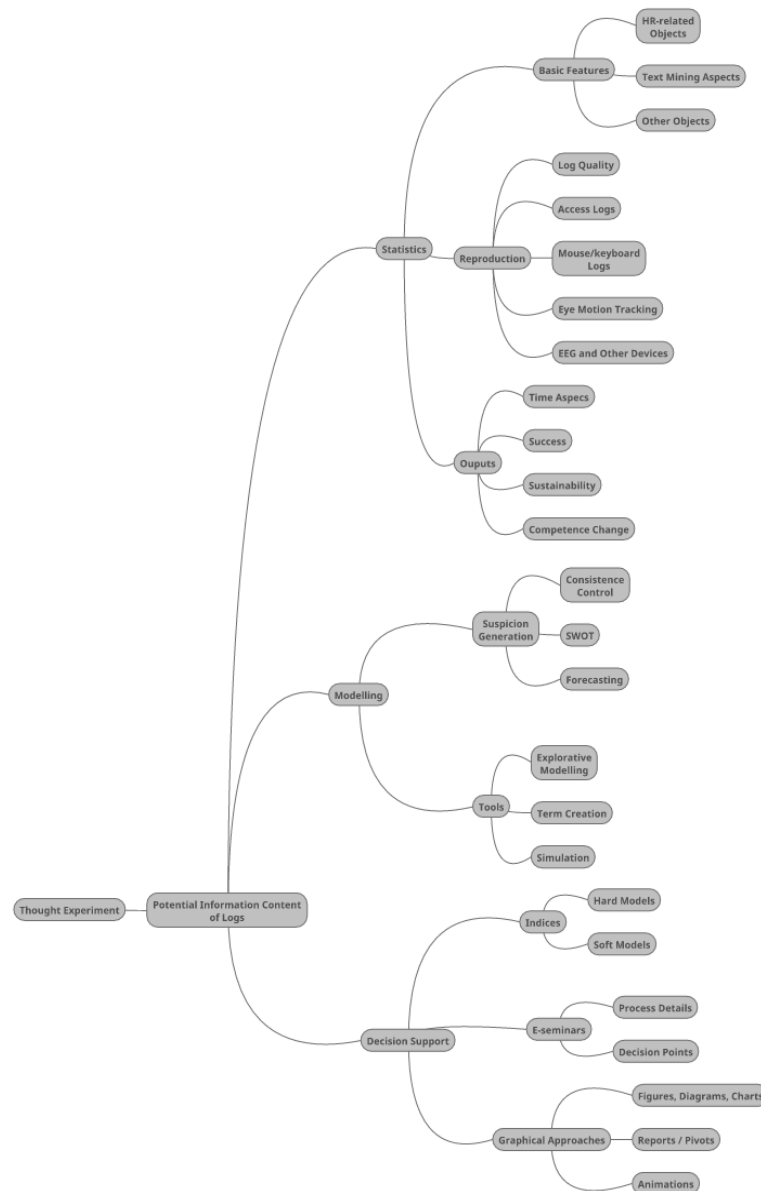


Figure 1. System elements and their connections (source: Edited by the Authors)

Legend

The figure above shows the structure of exploring diversity in learners' background and performance based on classic and innovative techniques (e.g. artificial intelligence). The brief description of the structure elements below outlines the most important characteristics:

Statistics: Parallel to modelling and decision support (see elements on the first level), classic statistics serve a kind of general interpretation and a sort of quality assurance in the data asset management through their results:

- *Basic features*: there are statistics, which are obligatory, such as plausibility or consistence checks (like exploring maximum, minimum, average, median values, or the same set for ratios between variables):
 - *HR- related objects*: in learning processes the HR-objects (like students, teachers, managers, etc.) play a central role;
 - *Text mining aspects*: although multimedia elements become more and more important in learning activities, pieces of texts are general parts of learning as such – pieces of texts, which may not be interpreted in an arbitrary depth, but they may be described based on a lot of specific indicator;
 - *Other objects* (courses, learning materials, etc.): parallel to HR objects, the object-attribute logic allows the definition of quasi unlimited further objects and their possibly measurable attributes;
- *Reproduction*: the logging of learning processes ensure the highest flexibility forever if the processes are stored in the most possible detailed way:
 - *Log quality*: the logs delivering high level reproducibility are in principle raw data from which further indicators/terms should be derived (like target-orientation of mouse movements);
 - *Access logs*: they are the most traditional data being available in e-learning systems, and they deliver basic information about system usage and access rights;
 - *Mouse/keyboard logs*: the most robust part of reproducibility is the set of control signals in the buffers assuming that there is not a voice control system active;
 - *Eye motion tracking*: eye motions do not produce control signals during the system usage, but they are signals for deriving learners' behaviour patterns;
 - *EEG and other devices*: like eye motion signals, EEG and/or other (non-invasive) physiological measurement devices are capable of collecting data about users' conditions;
- *Outputs*:
 - *Time aspects*: each logged data unit should always have a timestamp – this is the basic information unit of reproducibility in general;
 - *Success*: learning/teaching success indicators (mostly used as dependent variables of modelling/simulating) may be observed in a primary way (like credits, scores) and they may also be derived as secondary indicators (like ratio with eye contact during the whole education period in the class);
 - *Sustainability*: this is the multidimensional interpretation of unique success indicators, where a lot of parallel signals should always be aggregated to an ant discriminative index being capable of characterizing synchrony between separated evaluation criteria;
 - *Competence-change*: the most complex term concerning learning and teaching processes, because each kind of ability may be seen as an independent competence (like score in diverse tests: e.g. IQ) or a part of a set of raw indicators for a more

sophisticated phenomenon (like stress tolerance based on all thinkable inputs showing stress symptom);

Modelling: models are mostly decision-oriented objects with sophisticated hermeneutical support:

- *Suspicion generation*: suspicions are any type of irrationality in behaviour patterns of observed objects:
 - *Consistence control*: consistence is a global indicator in multidimensional system monitoring, where each attribute value in case of similar objects (like students) should be derivable based on the rest of variables;
 - *SWOT*: suspicion may have direction and dynamic – where the automated derivation of SWOT components deliver objective evaluations instead of subjective evaluation risks (as until now);
 - *Forecasting*: system behaviour may also be interpreted to predict the future (see plan-values) – each unexpected event is a kind of suspicion;
- *Tools*: to be capable of being objective, even automated, it is necessary to use innovative analytical techniques (like chains of similarity analyses):
 - *Explorative modelling*: exploration means being capable of deriving connection characteristics between phenomena without approximation in different ceteris paribus levels;
 - *Term creation*: human terms are products of human intuition processes, but computer systems need derivations of these terms from measured data in order to minimize the hermeneutical illusion of the human brain in the evaluation/observation/interpretation processes;
 - *Simulation*: to be able to influence complex systems, it is necessary to build simulators being capable of estimating system outputs, for circumstances which may happen, are bound to happen, but have not happened until now;

Decision support: assessing diversity in learners' background and performance is not an arbitrary activity – based on logs and their analyses, learning/teaching systems should be influenced in a rational way:

- *Indices*: in an object-attribute logic indices/variables make characterizing of objects possible (see before):
 - *Hard models*: analyses with a matured hermeneutical frame delivers decision possibilities (even for automated system maintenance if needed);
 - *Soft models*: analyses without clarified consequences catalyze human intuition processes leading to hard models sooner or later;
- *E-seminars*: this form of teaching and learning may be seen as a sort of complex experimental field:
 - *Process details*: to be able to derive complex hermeneutical processes, the observed system should always be described also with such parameters, which are basics of

- human ethological observations (like legal elements referring to learners and/or teachers);
- *Decision points*: in order to be efficient enough, it is necessary to think about the influence potential of system designers and users, where users are not always motivated to behave as expected (like pretended learning);
- *Graphical approaches*: a massive part of hermeneutical processes are visualization effects, where an optimized visual world supports both the analysis and teaching/learning:
 - *Figures, diagrams, charts*: only the most matured solution have the potential to substitute long descriptions;
 - *Reports/pivots*: it is necessary to build a standard for systems designers and also for users in case of common interpreting reports/pivots;
 - *Animations*: animations make it possible to show logical steps and/or dynamics of results.

Discussion

High-level reproduction of learning processes needs a kind of adaptation of data protection law. Users of e-learning systems must be informed with detailed rules in the terms of services before registration. In a similar case of a 360° health service (developed in a parallel way), data protection problems would be managed under the aegis of the Hippocratic Oath.

Classic statistical analyses may lead to a kind of hermeneutical illusion with higher frequency than sophisticated models based on artificial intelligence approaches. Beside physical, physiological, cognitive illusions, hermeneutical illusions derive the wrong association based on hard facts and cellular intuition capacities like prejudice towards individuals, or bad estimations in the stock markets, etc. Statistical logic uses ratios to derive evaluations, but this sort of thinking can be instable as the following fable highlighted the risk momentum: “It should be assumed, in a newspaper an article was published about for example that ‘Steppe-Land’ is at the last rank among the world’s countries regarding its environmental expenditures per capita, vis-a-vis for example ‘Robotia’ where a lot of money is expended for environmental protection. Poor people from Steppe-Land could walk from now with their head down in the world if they were not be spirited and if they could not be able so to say pass in the revolving door, too. As journalists and scientists of Steppe-Land published a news against the previous mentioned articles since the world's countries have been COMPARED using data series of UNO, and calculated that in certain life situation how much the correct amount is which should be spent to environmental protection considering the other country-indicators. And wonders of wonders in Steppe-Land, where the wind blows softly, the air has fresh flower fragrance and where industry, population is not much and where ancestors have already solved water regulation the statistically detected environmental expenditures per capita exceeds quasi with 10% that level, which could have been taken by others in their place. As in the case of Robotia (the muddy citadel of industry) turned out that if they spend quasi twice as

much as nowadays, maybe we can talk about such environmental condition, as in the case of Steppe-Land.” (Pitlik, 2014).

Infographics and their readers may both be diverse. The realization of a real graphical information added value is a kind of simulation problem. Based on the reader’s attributes, the ideal parameters of infographics should be optimized. For the time being this is a task for visual artists. However the concept of a robot designer may not be abandoned.

Potential meanings of competence should be defined in such a clear way, that the change of competence may be measured automatically. Competence could be interpreted in different ways: on one extreme opposite, the capability of mugging up learning material is also a sort of competence, but on the other extreme opposite – the right thinking about e.g. the Monty Hall paradox directly without ever during the teaching process is another competence. Term-creation capability of artificial intelligence methods – like similarity analyses (Pitlik, 2014) – ensures that words produced in the human brain can be transformed into source codes. The terms created should be compared to each other to see through the glass of term-creation possibilities, how consistent is the structure of facts (Pitlik & Horváth, 2016).

Conclusion

The reproduction level of learning processes (i.e. observing the user any time) is just a question of resources.

Artificial intelligence based analyses make managing diverse terms possible (originating in high flexible human association/intuition capabilities) in a numerical way. Model quality may be interpreted in new ways like consistency of models based on parallel function symmetry indices. Alternative solutions may be ranked following the logic of Occam’s razor, where the parallel goodness criteria of the ideal solution is interpreted in case of each unique solution. Finally a ranking model derives the best one, based on anti-discrimination principles.

The basic goal within the field of info-graphics is the adaptation of the well-known visualization highlight from Hans Rosling (2010). This solution can also be applied in the learning process as such, and in the interpretation processes of decision making/preparing.

Time-orientation makes it possible to calculate economic impacts during the learning processes. Moreover time is the navigation path (the “yellow brick road”) for the interpretation of competence changes. Measurable phenomena can be stored in timestamp-driven tables. Changes between status is also a time dependent term. Therefore time (the naïve time of the general human individuals) also delivers the basics for animations.

Summary

The main implications of the article are: log-quality should be optimized based on price/performance ratios for alternative reproduction levels including technical requirements and expected information added value.

The human brain should deliver more and more terms as a specific output of the hermeneutical processes about e-learning behaviours, which may always be modelled by using artificial intelligence approaches.

Decision support services may be generated in an adaptation procedure to exact decision needs (like aggregated evaluation indices for teachers, students, learning materials, courses, etc., or grouping students for predefined objectives of the classes based on each available attribute, etc.).

References

1. Monty Hall problem. (n.d.) In Wikipedia. Retrieved January 25, 2017, from https://en.wikipedia.org/wiki/Monty_Hall_problem
2. Pitlik, L. (2014). *My-X Team, an innovative idea-breeding farm*. Retrieved from https://www.researchgate.net/publication/280086739_My-X_Team_an_innovative_idea-breeding_farm
3. Pitlik, L., & Horváth, M. K. (2016). The human capability of term-creation is the core of the automated knowledge management. *Magyar Internetes Agrárinformatikai Újság*, 2016(220), 1-10.
4. Rosling, H. (2010, November 26). Hans Rosling's 200 Countries, 200 Years, 4 Minutes – The Joy of Stats – BBC Four. Retrieved from <https://youtu.be/jbkSRLYSojo>