Data Mining Tools in Science Education and Their Resources

Premysl Zaskodny

University of South Bohemia, Prague, Ceske Budejovice, Czech Republic

pzaskodny@seznam.cz

Abstract[1]

The main principle of paper is Data Mining in Science Education (DMSE) as Problem Solving. The main goal of paper is consisting in Delimitation of Complex Data Mining Tool and Partial Data Mining Tool of DMSE. The procedure of paper is consisting of Data Preprocessing in Science Education, Data Processing in Science Education, Description of Curricular Process as Complex Data Mining Tool (CP-DMSE), Description of Analytical Synthetic Modeling as Partial Data Mining Tool (ASM-DMSE) and finally Application of CP-DMSE and ASM-DMSE via Physics Education. The presented paper as an expression of inter-disciplinary communication of natural sciences and pedagogy through data mining is a supplement to the paper Zaskodny 2012 (published within JSCI) by newer publications.

Keywords*: Data Mining in Science Education, Complex Data Mining Tool – Curricular Process, Partial Data Mining Tool – Analytical Synthetic Modeling, Visualia - Results of Visualization*

1. Introduction

An imperative of data mining and a need of cooperation of the human with today´s computers are emphasized by D.A.Keim (Zaskodny, Pavlat 2009, Keim 2002):

"The progress made in hardware technology allows today´s computer systems to store very large amounts of data. Researchers from the University of Berkeley estimate that every year 1 Exabyte (= 1 Million Terabyte) of data are generated, of which a large portion is available in digital form. This means that in the next three years more data will be generated than in all of human history before".

"If the data is presented textually, the amount of data which can be displayed is in range one hundred data items, but this is like a drop in the ocean when dealing with data sets containing millions of data items".

[1] Paper Editors: Prof.Jana Skrabankova,Ph.D. (Czech Republic), Petr Prochazka,M.Sc. (Czech Republic)
Paper Reviewers: Prof.Jana Skrabankova,Ph.D. (Czech Republic), Prof.Vladislav Pavlat,Ph.D. (Czech Republic)

"For data mining to be effective, it is important to include the human in the data exploration process and combine the flexibility, creativity, and general knowledge of the human with the enormous storage capacity and the computational power of today´s computers."

Modeling as a partial tool of data mining – quotation according to J.K.Gilbert (Gilbert 2008)**:**
"In a nightmare world, we would perceive the world around us being continuous and without structure. However, our survival as a species has been possible because we have evolved the ability do "cut up" that world mentally into chunks about which we can think and hence give meaning to".

"This process of chunking, a part of all cognition, is modeling and the products of the mental actions that have taken place are models. Science, being concerned with the provision of explanations about the natural world, places an especial reliance on the generation and testing of models".

2. Data Mining

Data Mining - analytical synthetic way of extraction of hidden and potentially useful information from large data files
(continuum data-information-knowledge, knowledge discovery)

Data Mining Techniques - the system functions of structure of formerly hidden relations and patterns (e.g. classification, association, clustering, prediction)

Data Mining Tool - a concrete procedure how to reach the intended system functions

Complex Tool - a resolution of complex problem of relevant science branch

Partial Tool - a resolution of partial problem of relevant science branch
Result of Data Mining - a result of data mining tool application

Representation of Data Mining Result - a description of this what is expressed

Visualization of Data Mining Result - optical retrieval of data mining result

Data Mining Cycle - Data Definition, Data Gathering, Data Preprocessing, Data Processing, Discovering Knowledge or Patterns, Representation and Visualization of Results.

See Tarabek, Zaskodny, 2009, Zaskodny, Novak, 2009, Zaskodny, Prochazka, 2009, Zaskodny, Pavlat, 2009, Zaskodny, Skrabankova, 2009, Zaskodny, 2012.

Quoted sources in Tarabek, Zaskodny 2009:
E.g. American Library Association, M.C.Borba, E.M.Villarreal, G.M.Bowen, W-M Roth, C.Brunk, J.Kelly, R.Kohavi, Mineset, B.V.Carolan, G.Natriello, N.Delavari, M.R.Beikzadeh, S.Phon-Amaisuk, U-D Ehlers, J.M.Pawlowski, U.M.Fayyad, G.Piatelsky-Shapiro, P.Smyth, J.Fox, D.Gabel, J.K.Gilbert, O.de Jong, R.Justi, D.F.Treagust, J.H.Van Driel, M.Reiner, M.Nakhleh, W.Hämäläinen, H.Laine, E.Sutinen, M.Hesse, A.H.Johnstone, M.J.Kearns, V.Vazivani, D.A.Keim, R.Kwan, R.Fox, FT Chan, P.Tsang, Le Jun, J.Luan, J.Manak, National research Council-NRC, R.Newburgh, I.Nonaka, H.Takeuchi, C.J.Petroselli, E.F.Redish, D.Reisberg, C.Romero, N.Rubenking, R.E.Scherr, M.Sabella, D.A.Simovici, V.Spousta, L.Talavera, S.Ventura,

3. Data Preprocessing in Science Education

Result of Data Preprocessing – Educational Communication of Natural Science as a succession of transformations of education content forms:

- The transformation T1 is transformation of scientific system of natural science to communicable scientific system of natural science (the first form of education content existence),

- The transformation T2 is transformation of communicable scientific system of natural science to educational system of natural science (the second form of education content existence),

- The transformation T3 is transformation of educational system of natural science to both instructional project of natural science and preparedness of educator to education (the third and fourth forms of education content existence),

- The transformation T4 is transformation of both instructional project of natural science and preparedness of educator to results of education (the fifth form of education content existence),

- The transformation T5 is transformation of results of natural science education to applicable results of natural science education (the sixth form of education content existence)

See Brockmeyerova, 1982, Tarabek, Zaskodny, 2007 Vol.1, Vol.2, Vol.3, Tarabek, Zaskodny, 2008, Zaskodny, Prochazka, 2014, Zaskodny, 2015, Zaskodny, 2016.

4. Data Processing in Science Education

Result of Data Processing – Curricular Process of Natural Science as a succession of transformations of algorithmized and formalized education content forms:

i. The form of education content existence - "**variant form of curriculum**"

ii. **The curriculum** - "**education content**" (Prucha 2005)

iii. **The variant forms of curriculum have got the universal structure** (four structural elements - sense and interpretation, set of objectives, conceptual knowledge system, factor of following transformation)

iv. **The variant forms of curriculum were selected through a fusion of Anglo-American curricular tradition and European didactic tradition**

v. **The curricular process is defined as the succession of transformations** T1-T5 **of curriculum variant forms:**

"**conceptual curriculum**"
(output of T1, the first variant form of curriculum) – the communicable scientific system

"**intended curriculum**"
(output of T2, the second variant form of curriculum) – the educational system of natural science

"**projected curriculum**"
(output of T3, the third variant form of curriculum) – the instructional project of natural science

"**implemented curriculum-1**"
(output of T3, the fourth variant form of curriculum) – the preparedness of educator to education

"**implemented curriculum-2**"
(output of T4, the fifth variant form of curriculum) – the results of education

"**attained curriculum**"
(output of T5, the sixth variant form of curriculum) – applicable results of education

See Zaskodny, Novak 2009, Zaskodny 2015, Zaskodny 2016

Quoted sources in Zaskodny, Novak 2009:

E.g. A.V.Kelly, M.K.Smith, W.Doyle, M.Pasch, A.M.Sochor, V.V.Krajevskij, I.J.Lerner, J.McVittie, K.Carter, G.M.Blenkin, L.Stenhouse, E.Newman, G.Ingram, F.Bobitt, R.W.Tyler, H.Taba, C.Cornblet, S.Grundy, D.Lawton, P.Gordon, M.Certon, M.Gayle, G.J.Posner

5. Complex and Partial Tool of DMSE: CP-DMSE, ASM-DMSE

Complex tool of DMSE (CP-DMSE) is given by curricular process of natural science. CP-DMSE delimits the correct education content via succession of transformations T1-T5.

Partial tool of DMSE is given by analytical synthetic modeling (ASM-DMSE). ASM-DMSE describes the mediated or real problem solving within the inputs and outputs of individual transformations T1-T5. In this paper, the description of ASM-DMSE is realized by means of both visualie Vis.1 and Legend to Vis.1.

Legend toVis.1

a (Identified Complex Problem) – Investigated area of reality, investigated phenomenon

$\boldsymbol{B_k}$ (Analysis) – Analytical segmentation of complex problem to partial problems

$\mathbf{b_k}$ (Partial problems PP-k) – Result of analysis: essential attributes and features of investigated phenomenon

$\boldsymbol{C_k}$ (Abstraction) – Delimitation of partial problems essences by abstraction with goal to acquire the partial solutions

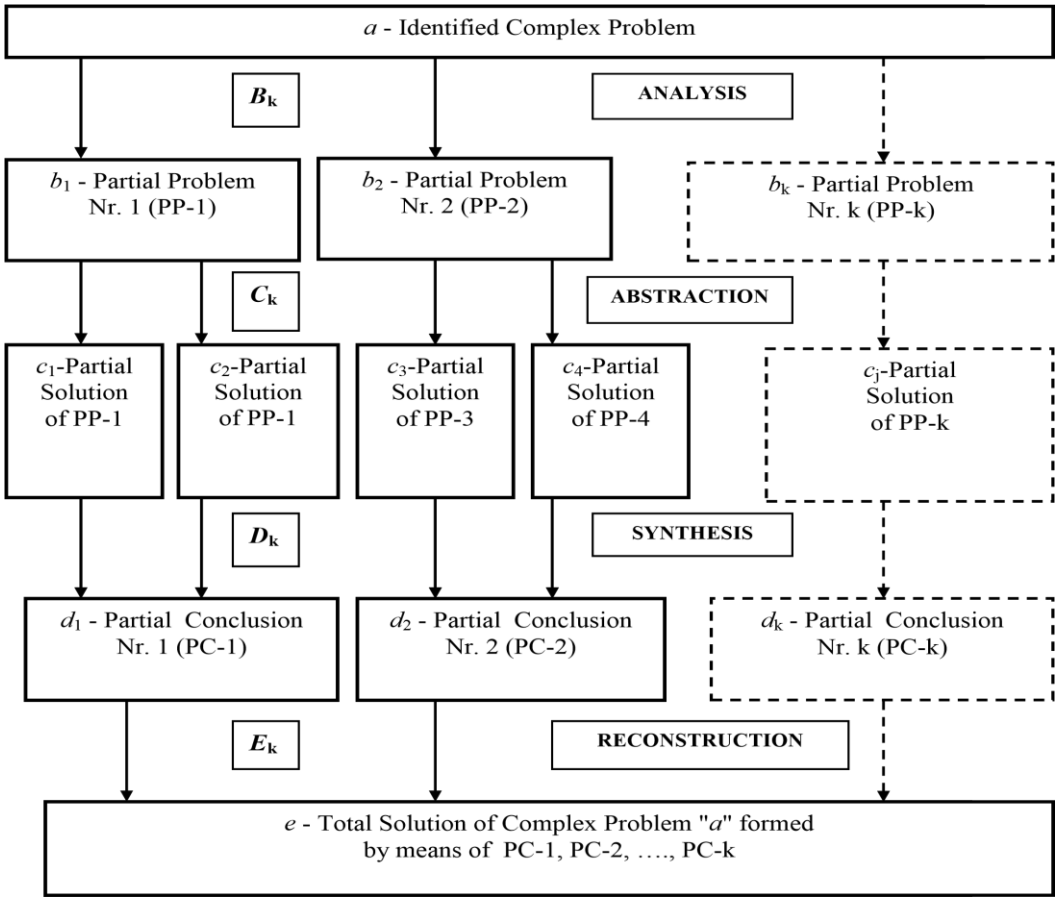
$\boldsymbol{c_k}$ (Partial solutions PS-k) – Result of abstraction: partial concepts, partial pieces of knowledge, various relations, etc.

$\boldsymbol{D_k}$ (Synthesis) – Synthetic finding dependences among results of abstraction

$\boldsymbol{d_k}$ (Partial conclusions PC-k) – Result of synthesis: principle, law, dependence, continuity

$\boldsymbol{E}_{\mathbf{k}}$ (Intellectual reconstruction)

– Intellectual reconstruction of investigated phenomenon /investigated area of reality ***e*** (Total solution of complex problem "*a*")
– Result of intellectual reconstruction: analytical synthetic structure of final knowledge (conceptual knowledge system)

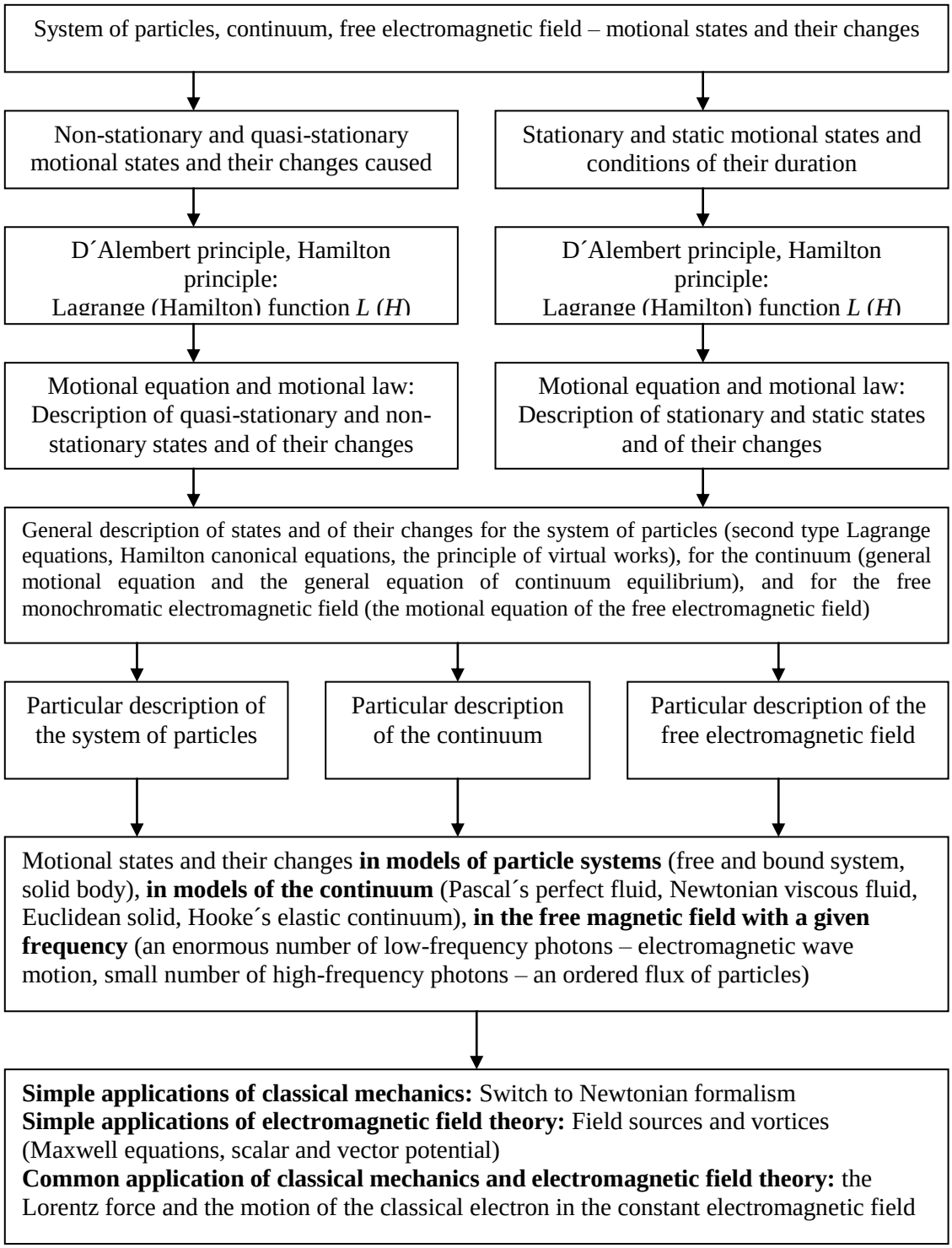


Vis. 1
General model of analytical synthetic structure

6. Application of Partial Tool ASM-DMSE

The application of ASM-DMSE is the visualie Vis.2

from the area of physics education. The visual Vis.2 is an analytical synthetic model of classical non-statistical physics (classical mechanics, mechanics of continuum, free electromagnetic field). This visualie constitutes a part of physics conceptual curriculum as a part of communicable scientific system of physics (a part of output of transformation T1).



Vis. 2

Analytical synthetic model of classical non-statistical physics

The visualized result Vis.2 of data mining in physics education constitutes the paramorphic model and hypertextual representation, represents the external conceptual knowledge systems as external representation of general social experience.

The visualized result also represents the concrete type of data file - the representation of classical mechanics, mechanics of continuum and free electromagnetic field.

Remarks:

The visualization of all conceptual curriculum of physics (classical, quantum and relativistic statistical and non-statistical physics) **- see English books** Zaskodny, Prochazka 2014, Zaskodny 2016.

The visualization of curricular process of physics - see Czech books Zaskodny 2015, Zaskodny 2018.

The variant form of curriculum - see Czech book Prucha 2005: Modern Pedagogy.

7. Conclusions

1. Educational Communication of Natural Science as Result of Data Preprocessing

2. Educational Communication of Natural Science as Five Transformations T1-T5 of Knowledge from Natural Science to Mind of Educational Person

3. Curricular Process of Natural Science as Result of Data Processing

4. Curricular Process of Natural Science as Structuring, Algorithm Development and Formalization of Educational Communication of Natural Science

5. Curricular Process as Succession of Five Transformations T1-T5 of Curriculum Variant Forms

6. Curriculum Variant Forms as Forms of Education Content Existence

7. Formalization of Curriculum Variant Form (Four of Universal Structural Elements: Sense and Interpretation, Set of Objectives, Conceptual Knowledge System, Factor of Following Transformation)

8. Variant Forms of Curriculum:

- Conceptual Curriculum as Communicable System of Natural Science
- Intended Curriculum as Educational System of Natural Science
- Projected Curriculum as Instructional Project of Natural Science and Its Textbook
- Implemented Curriculm-1 as Preparedness of Educator to Education
- Implemented Curriculum-2 as Results of Education in Mind of Educational Person
- Attained Curriculum as Applicable Results of Education

9. Curricular Process as CP-DMSE (Structuring, Algorithm Development and Formalization of Five Transformations Succession T1-T5)

10. Analytical Synthetic Modelling as ASM-DMSE (Modelling Inputs and Outputs of Transformations T1-T5)

11. Analytical Synthetic Models as Results of Problems Solving (Real or Mediated Problems)

12. Application of CP-DMSE and ASM-DMSE via Physics Education (Visualie of Conceptual Curriculum in Area of Classical Non-Statistical Physics, Need of Visualia of All Curriculum Variant Forms as Application of CP-DMSE)

Acknowledgement to the Non-Anonymous Peer Reviewers

Prof. Jana Skrabankova, Ph.D.
University of Ostrava, Department of Physics
Ostrava, Czech Republic
janaskrabankova@atlas.cz

Prof. Vladislav Pavlat, Ph.D.
University of Finance and Administration
Prague, Czech Republic
v.pavlat@volny.cz

References

Brockmeyerová,J. (1982), **Introduction into Theory and Methodology of Physics Education,** Prague, Czech Republic: SPN

Gilbert,J.K. (2008), **Visualization: An Emergent Field of practice and Enquiry** In: Visualization: Theory and Practice in Science (Models and Modeling in Science Education), New York: Springer Science + Business Media

Keim,D.A. (2002), **Information Visualization and Visual Data Mining** IEEE Transactions on Visualization and Computer Graphics. Vol.7, No.1, January-March

Průcha,J. (2005), **Moderní pedagogika (Modern Pedagogy)**, Prague, Czech Republic: Portál

Tarábek,P., Záškodný,P. (2007), **Educational and Didactic Communication 2007, Vol.1 – Theory,** Bratislava, Slovak Republic: Didaktis, 2007, ISBN 987-80-89160-56-3

Tarábek,P., Záškodný,P. (2007), **Educational and Didactic Communication 2007, Vol.2 – Method,** Bratislava, Slovak Republic: Didaktis, ISBN 987-80-89160-56-3

Tarábek,P., Záškodný,P. (2007), **Educational and Didactic Communication 2007, Vol.3 – Application,** Bratislava, Slovak Republic: Didaktis, ISBN 987-80-89160-56-3

Tarábek,P., Záškodný,P. (2008), **Educational and Didactic Communication 2008,** Bratislava, Slovak Republic: Didaktis, ISBN 987-80-89160-62-4

Tarábek,P., Záškodný,P. (2009), **Educational and Didactic Communication 2009**, Bratislava, Slovak Republic: Didaktis, 2009, ISBN 978-80-89160- 69-3

Záškodný,P., Novák,V. (2009), **Data Mining - A Brief Summary**, In: Educational and Didactic Communication 2009, Bratislava, Slovak Republic: Didaktis, 2009, ISBN 978-80-89160-69-3

Záškodný,P., Pavlát,V. (2009), **Data Mining - A Brief Recherche**, In: Educational and Didactic Communication 2009, Bratislava, Slovak Republic: Didaktis, ISBN 978-80-89160- 69-3

Záškodný,P., Procházka,P. (2009), **Collective Scheme of Both Educational, Communication and Curricular Process,** In: Educational and Didactic Communication 2009, Bratislava, Slovak Republic: Didaktis, ISBN 978-80-89160- 69-3

Záškodný.P., Procházka,P. (2014), **Survey of Principles of Theoretical Physics (with Application to Radiology),** Prague, Czech Republic: Curriculum (in English), ISBN 978-80-87894-02-6**,** Online Presentation: www.csrggroup.org, online catalogue of National Library Prague

Záškodný,P., Škrabánková,J. (2009) **Modelling and Visualization of Problem Solving**, In: Educational and Didactic Communication 2009, Bratislava, Slovak Republic: Didaktis, ISBN 978-80-89160- 69-3

Záškodný,P. (2009), **Representation of Results of Data Mining,** In: Educational and Didactic Communication 2009, Bratislava, Slovak Republic: Didaktis, ISBN 978-80-89160- 69-3

Záškodný,P. (2012), **Data Mining Tools in Science Education,** In: Journal of Systemics, Cybernetics and Informatics, Volume 10, Number 6, Year 2012**,** GIF 2012 – 0.562, ISSN 1690-4524, Záškodný,P. (2015), **Curricular Process of Physics**
Prague, Czech Republic: Curriculum, ISBN 978-80-87894-04-0, Online Presentation: www.csrggroup.org, online catalogue of National Library Prague

Záškodný,P. (2016) **Curriculum Research and Development in Physics Education,** Bratislava,, Slovak Republic (in English): Didaktis, ISBN 978-80-8166-014-6, Online Presentation: www.csrggroup.org, online catalogue of City Library Prague, www.diddaktis.sk

Záškodný,P. (2018), **Survey of Principles of Theoretical Physics,** (with Application to Radiology), Prague, Czech Republic: Curriculum, ISBN 978-80-87894-17-0, Online Presentation: www.csrggroup.org, online catalogue of National Library Prague