UNIUNEA EUROPEANĂ

GUVERNUL ROMÂNIEI
MINISTERUL MUNCII, FAMILIEI, PROTECȚIEI SOCIALE ȘI PERSOANELOR VÂRSTNICE
AMPOSDRU

Fondul Social European
POS DRU
2007-2013

Instrumente Structurale
2007-2013

MINISTERUL EDUCAȚIEI NAȚIONALE

OIPOSDRU

ACADEMIA ROMÂNĂ

CULTURĂ IDENTITATE

Investește în OAMENI

Dumitru Iulian NĂSTAC

Prelucrarea inteligentă a informațiilor multidisciplinare pentru prognoze adaptive în contextul globalizării

Muzeul Național al Literaturii Române

Editura Muzeului Național al Literaturii Române

UNIUNEA EUROPEANĂ
GUVERNUL ROMÂNIEI
MINISTERUL MUNCII, FAMILIEI, PROTECȚIEI SOCIALE ȘI PERSOANELOR VÂRSTNICE
AMPOSDRU
Fondul Social European
POS DRU
2007-2013
Instrumente Structurale
2007-2013
MINISTERUL EDUCAȚIEI NAȚIONALE
OIPOSDRU
ACADEMIA ROMÂNĂ

Dumitru Iulian NĂSTAC

PRELUCRAREA INTELIGENTĂ A INFORMAȚIILOR MULTIDISCIPLINARE PENTRU PROGNOZE ADAPTIVE ÎN CONTEXTUL GLOBALIZĂRII

PRELUCRAREA INTELIGENTĂ A INFORMAȚIILOR MULTIDISCIPLINARE PENTRU PROGNOZE ADAPTIVE ÎN CONTEXTUL GLOBALIZĂRII

Autor: **Dumitru Iulian NĂSTAC**
Conducător ştiinţific: Acad. **Paul Dan CRISTEA**

Lucrare realizată în cadrul proiectului „*Valorificarea identităților culturale în procesele globale*", cofinanțat din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013, contractul de finanțare nr. POSDRU/89/1.5/S/59758.

Titlurile şi drepturile de proprietate intelectuală şi industrială asupra rezultatelor obținute în cadrul stagiului de cercetare postdoctorală aparțin Academiei Române.

Punctele de vedere exprimate în lucrare aparțin autorului şi nu angajează Comisia Europeană şi Academia Română, beneficiara proiectului.

Exemplar gratuit. Comercializarea în țară şi străinătate este interzisă.
Reproducerea, fie şi parțială şi pe orice suport, este posibilă numai cu acordul prealabil al Academiei Române.

ISBN 978-973-167-190-1 Depozit legal: Trim. II 2013

Dumitru Iulian NĂSTAC

Prelucrarea inteligentă a informațiilor multidisciplinare pentru prognoze adaptive în contextul globalizării

Editura Muzeului Național al Literaturii Române

Colecția *AULA MAGNA*

UNIUNEA EUROPEANĂ

GUVERNUL ROMÂNIEI
MINISTERUL MUNCII, FAMILIEI, PROTECȚIEI SOCIALE ȘI PERSOANELOR VÂRSTNICE
AMPOSDRU

Fondul Social European
POS DRU
2007-2013

Instrumente Structurale
2007-2013

MINISTERUL EDUCAȚIEI NAȚIONALE

OIPOSDRU

ACADEMIA ROMÂNĂ

Investeşte în oameni !

FONDUL SOCIAL EUROPEAN

Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007 – 2013

Axa prioritară 1: „Educația şi formarea profesională în sprijinul creşterii economice şi dezvoltării societății bazate pe cunoaştere"

Domeniul major de intervenție 1.5: „Programe doctorale şi postdoctorale în sprijinul cercetării"

Titlul proiectului: „Valorificarea identităților culturale în procesele globale"

Contract: POSDRU/89/1.5/S/59758

Beneficiar: ACADEMIA ROMÂNĂ

Parteneri în proiect: (I) UNIVERSITATEA POLITEHNICA Bucureşti, Facultatea de Mecanică şi Mecatronică; (II) UNIVERSITATEA din Craiova

CULTURĂ IDENTITATE

Obiectivele proiectului şi domeniile de cercetare:

1. **Obiectivul general:** Model-pilot de şcoală postdoctorală prin implicarea a 92 de cercetători postdoctoranzi, în scopul dezvoltării carierei în cercetare, al îmbunătățirii programelor de cercetare postdoctorală în domeniul umanioarelor, al impulsionării şi consolidării sectorului de cercetare în ştiințele socioumane din România, pentru sprijinirea economiei româneşti în dobândirea unor avantaje competitive durabile şi micşorarea decalajelor între România şi celelalte țări membre ale Uniunii Europene.
2. **Obiectivele specifice:** Elaborarea şi implementarea de noi tehnologii-suport pentru derularea proiectului; formarea şi perfecționarea cercetătorilor prin programe postdoctorale • Organizarea unor acțiuni de îndrumare a cercetătorilor pe parcursul stagiilor derulate în străinătate • Sprijinirea cercetătorilor în participarea la seminarii şi conferințe internaționale • Organizarea unor sesiuni pentru promovarea egalității de şanse şi a dezvoltării durabile • Sprijinirea colaborării între universități, institute de cercetare şi companii din aria tematică a şcolii postdoctorale • Dezvoltarea de activități novatoare în vederea accentuării importanței programelor de cercetare interdisciplinară; crearea de metodologii proprii cu privire la derularea programelor postdoctorale • Elaborarea unor ghiduri de bune practici cu privire la schimbul de experiență internațional în aria cercetării în ştiințele socioumane prin programe postdoctorale.
3. **Domeniile cercetării:** filologie - literatură • ştiințe istorice şi arheologie • filosofie, teologie, psihologie, pedagogie • arte, arhitectură şi audio-vizual • ştiința informației • sociologia culturii • antropologie • etnografie şi folclor

Cuprins

MULȚUMIRILE AUTORULUI 9

INTRODUCERE 13

PARTEA I

UN CONTEXT CULTURAL

CAPITOLUL 1

O PERSPECTIVĂ NORDICĂ PRIVIND EGALITATEA DE ŞANSE 27

CAPITOLUL 2

UNIVERSITATEA DIN OXFORD - UN MODEL DE TRADIȚIE CULTURALĂ ŞI PARALELE ROMÂNEŞTI 35

CAPITOLUL 3

ASPECTE ŞI IMPRESII DINTR-UN CENTRU UNIVERSITAR FINLANDEZ 55

PARTEA A II-A

CUM PUTEM PRIVI PREDICȚIILE DE DATE

CAPITOLUL 4

STADIUL ACTUAL ÎN DOMENIUL PREDICȚIILOR DE DATE 65

CAPITOLUL 5

PROBLEMA ŞI SITUAȚIA CERCETĂTORULUI 75

CAPITOLUL 6

CONSIDERAȚII FILOZOFICE ŞI STATISTICE 79

CAPITOLUL 7

UTILIZAREA BAGAJULUI DE CUNOŞTINȚE 91

PARTEA A III-A

UN MODEL ADAPTIV PENTRU PREDICȚIA DE DATE

CAPITOLUL 8

INSTRUMENTE UTILE FOLOSITE PENTRU CONSTRUIREA MODELULUI DE PREDICȚIE 109

8.1. Rețelele neuronale artificiale 109

8.1.1. Modelul neuronal clasic 110

8.1.2. Arhitecturi de rețele neuronale 113

8.1.3. Funcția criteriu şi algoritmi de învățare 116

8.1.4. Selecția caracteristicilor şi fenomenul de generalizare în rețelele neuronale 120

8.1.5. Eroarea de învățare şi eroarea de test 121

8.2. Prelucrarea dimensiunii spațiului datelor 123

8.2.1. Analiza componentelor principale 123

8.2.2. Extinderea conceptului ACP 126

CAPITOLUL 9

DESCRIEREA MODELULUI DE PREDICȚIE 131

9.1. Considerații introductive asupra modelului de predicție de date 131

9.2. Arhitectura modelului 133

9.3. Procedura de reantrenare adaptivă 139

CAPITOLUL 10

REZULTATE EXPERIMENTALE 147

10.1. Predicția unor secvențe ADN 147

10.1.1. Predicția unui semnal genomic în cazul organismelor procariote 148

10.1.2. Predicția unei secvențe de ADN în cazul organismelor eucariote 151
10.2. Predicția noxelor din atmosfera oraşului Bucureşti 153
10.3. Prognoza consumului de energie electrică 156
10.3.1. Prognoza consumului de energie electrică peste o oră şi peste şase ore 157
10.3.2. Influența parametrului de agitare 164
10.4. Predicția ratei de schimb valutar 170
10.5. Posibile extensii ale modelului în cadrul aplicațiilor 186

CONCLUZII 191

BIBLIOGRAFIE 195

ADDENDA 209

ABSTRACT 209

SUMMARY 212

ADDENDA

Abstract

Intelligent processing of multidisciplinary information for adaptive predictions in the context of globalisation

I have been working on adaptive forecasting models for more than a decade. Some years ago, I came to consider writing a book on this subject, as a kind of niche task on predictions, not very broad in scope, such as the Scott Armstrong's famous handbook, the *Principles of forecasting*. During the last two years, I had the opportunity to make this aspiration come true in the framework of a Romanian Academy project. Here, it was a great satisfaction to collaborate and work in multidisciplinary domains with academicians Paul Dan Cristea and Emilian Dobrescu. I was honored to participate in the Post-doctoral School that was kindly and efficiently coordinated by academician Eugen Simion, the President of the Scientific Council, and professor Valeriu Ioan Franc, the Project Manager. It actually turned out to be a lengthy list of collaborators, from which I wish to especially mention Dr. Simon Stringer, Senior Research Fellow at the University of Oxford, UK, and four professors from Finland: Barbro Back and Christer Carlsson, both at the Åbo Akademi, Reima Suomi of the University of Turku, and, last but not least, Mikael Collan, who is now with the Lappeenranta University of Technology.

The main idea of this book underlines the importance of adequate training in data forecasting. An intelligent adaptive model based on the retraining procedure is used for prediction of non-stationary sequences. This adaptive technique, applied on Artificial Neural Networks, was firstly proposed in 2004 as an enhancement of the forecasting method developed for the EUNITE Competition 2003 (European Network of Excellence on Intelligent Technologies for Smart Adaptive Systems). Then, it was used in

various kinds of predictions that concern a wide range of data, from industrial to financial applications, including Nucleotide Genomic Signals for special forecasting which uses spatial sequences instead of time series. New features have been added to the general model in the last two years. Educational, philosophical, theoretical and practical issues are presented in this book. In some aspects, this intelligent model could be viewed as a child that always learns, and acquires new knowledge, by becoming, step by step, more experienced in a permanently changing environment. The quality of the provided data, together with an adequate setting of the model parameters, could provide a permanent improvement of the successive predictions. It seems to be, in a way, a matter of education. The model cannot be properly implemented without the expertise of the experienced specialists that effectively work on the specific field from where the data come.

The book is divided in three main parts. The first (which contains three chapters) treats a cultural context concerning my research activity in two university centers from Northern Europe. This experience had a great impact over the model development. The second part is concentrated on a philosophical approach regarding data predictions. All four chapters included here are linked to a major question: how can we better see the issues of forecasting? There is still a debate between short and long term forecasting. The last one is usually affected by some influences that cannot always be taken into consideration. An important observation is that the involvement of a large knowledge base and a solid education could be crucial for adequate short term anticipation of different phenomena.

The last part of the book is more technically orientated. Detailed and intuitive explanations are presented for each component of the predictive model. The novelty introduced by this model concerns an aspect of an original retraining procedure, which allows a fast recalibration of an artificial neural network core for the newest acquired data in a nonstationary environment. Esentially, the predictions depend on the history of many related and relevant inputs, alongside another history of the forecasted data in a recurrent way. The involved retraining algorithm is well suited in forecasting applications, where there is a huge amount of data. Fortunately,

this large number of inputs can be drastically reduced by employing an intermediate step that consists of a Principal Component Analysis processing block. A series of aspects that are derived from both remembering and forgetting processes, are described here. The last chapter of this part deals with a series of relevant applications from bioinformatics, environmental pollution preventing, electric load forecasting and financial fields. The succession of these examples is not arbitrary and it was chosen in order to gradually understand the relevant aspects of the model. Further extensions for other applications are also suggested here, since the model is still open for future improvements.

As a last thought, I hope that the present book is appropriate for those students and specialists who are interested not only in practical applications, but also in pursuing research in the wide area of data forecasting.

Summary

ACKNOWLEDGEMENTS 9

INTRODUCTION 13

PART I: A CULTURAL CONTEXT

CHAPTER 1: A NORDIC PERSPECTIVE ON EQUAL OPPORTUNITIES 27

CHAPTER 2: UNIVERSITY OF OXFORD - A MODEL OF CULTURAL TRADITION AND ROMANIAN CONSIDERATIONS 35

CHAPTER 3: ASPECTS AND IMPRESSIONS FROM A FINNISH UNIVERSITY 55

PART II: AN APPROACH TO DATA FORECASTING

CHAPTER 4: THE STATE OF THE ART IN DATA FORECASTING 65

CHAPTER 5: BEING A RESEARCHER 75

CHAPTER 6: PHILOSOPHICAL AND STATISTICAL CONSIDERATIONS 79

CHAPTER 7: USING THE KNOWLEDGE BASE 91

PART III: AN ADAPTIVE MODEL FOR DATA FORECASTING

CHAPTER 8: TOOLS FOR BUILDING A PREDICTIVE MODEL 109

8.1. Artificial neural networks 109

8.1.1. The classic neural model 110

8.1.2. Neural networks architectures 113

8.1.3. The criterion function and training algorithms 116

8.1.4. Feature selection and generalisation ability of neural networks 120
8.1.5. Training and test errors 121
8.2. Processing of data dimensions 123
8.2.1. Principal component analysis 123
8.2.2. Extensions of principal component analysis 126

CHAPTER 9: DESCRIPTION OF THE PREDICTION MODEL 131
9.1. Preliminary notes on the data forecasting model 131
9.2. Model architecture 133
9.3. Adaptive retraining procedure 139

CHAPTER 10: EXPERIMENTAL RESULTS 147
10.1. DNA sequence forecasting 147
10.1.1. Genomic signal forecasting – the Prokaryote case 148
10.1.2. Genomic signal forecasting – the Eukaryote case 151
10.2. Prediction of pollutants in the atmosphere of Bucharest 153
10.3. Electric load forecasting 156
10.3.1. Electric load forecasting over the next hour and over the next six hours 157
10.3.2. Influence of the shaking parameter 164
10.4. Currency exchange rate forecasting 170
10.5. Possible extensions of the forecasting model in applications 186

CONCLUSIONS 191

REFERENCES 195