

Machine Learning in Biomedical Science and Healthcare Informatics



HANDBOOK ON INTELLIGENT HEALTHCARE ANALYTICS

Knowledge Engineering with Big Data

Edited by

A. JAYA

K. KALAISELVI

DINESH GOYAL

DHIYA AL-JUMEILY

 Scrivener
Publishing

WILEY

Handbook of Intelligent Healthcare Analytics

Scrivener Publishing
100 Cummings Center, Suite 541J
Beverly, MA 01915-6106

Publishers at Scrivener

Martin Scrivener (martin@scrivenerpublishing.com)
Phillip Carmical (pcarmical@scrivenerpublishing.com)

Handbook of Intelligent Healthcare Analytics

Knowledge Engineering with Big Data Analytics

Edited by

A. Jaya

*Department of Computer Application, B.S. Abdur Rahman Crescent Institute of
Science, Technology, Chennai, India*

K. Kalaiselvi

*Department of Computer Science, Vels Institute of Science, Technology and
Advanced Studies, Chennai, India*

Dinesh Goyal

Poornima Institute of Engineering & Technology, Jaipur, India

and

Dhiya AL-Jumeily

Faculty of Engineering and Technology, Liverpool John Moores University, UK



WILEY

This edition first published 2022 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA
© 2022 Scrivener Publishing LLC
For more information about Scrivener publications please visit www.scrivenerpublishing.com.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

Wiley Global Headquarters

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials, or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read.

Library of Congress Cataloging-in-Publication Data

ISBN 978-1-119-79179-9

Cover image: Pixabay.Com

Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface	xvii
1 An Introduction to Knowledge Engineering and Data Analytics	1
<i>D. Karthika and K. Kalaiselvi</i>	
1.1 Introduction	2
1.1.1 Online Learning and Fragmented Learning Modeling	2
1.2 Knowledge and Knowledge Engineering	5
1.2.1 Knowledge	5
1.2.2 Knowledge Engineering	5
1.3 Knowledge Engineering as a Modelling Process	6
1.4 Tools	7
1.5 What are KBSs?	8
1.5.1 What is KBE?	8
1.5.2 When Can KBE Be Used?	10
1.5.3 CAD or KBE?	12
1.6 Guided Random Search and Network Techniques	13
1.6.1 Guide Random Search Techniques	13
1.7 Genetic Algorithms	14
1.7.1 Design Point Data Structure	15
1.7.2 Fitness Function	15
1.7.3 Constraints	16
1.7.4 Hybrid Algorithms	16
1.7.5 Considerations When Using a GA	16
1.7.6 Alternative to Genetic-Inspired Creation of Children	17
1.7.7 Alternatives to GA	18
1.7.8 Closing Remarks for GA	18
1.8 Artificial Neural Networks	19
1.9 Conclusion	19
References	20

2	A Framework for Big Data Knowledge Engineering	21
	<i>Devi T. and Ramachandran A.</i>	
2.1	Introduction	22
2.1.1	Knowledge Engineering in AI and Its Techniques	23
2.1.1.1	Supervised Model	23
2.1.1.2	Unsupervised Model	23
2.1.1.3	Deep Learning	24
2.1.1.4	Deep Reinforcement Learning	24
2.1.1.5	Optimization	25
2.1.2	Disaster Management	25
2.2	Big Data in Knowledge Engineering	26
2.2.1	Cognitive Tasks for Time Series Sequential Data	27
2.2.2	Neural Network for Analyzing the Weather Forecasting	27
2.2.3	Improved Bayesian Hidden Markov Frameworks	28
2.3	Proposed System	30
2.4	Results and Discussion	32
2.5	Conclusion	33
	References	36
3	Big Data Knowledge System in Healthcare	39
	<i>P. Sujatha, K. Mahalakshmi and P. Sripriya</i>	
3.1	Introduction	40
3.2	Overview of Big Data	41
3.2.1	Big Data: Definition	41
3.2.2	Big Data: Characteristics	42
3.3	Big Data Tools and Techniques	43
3.3.1	Big Data Value Chain	43
3.3.2	Big Data Tools and Techniques	45
3.4	Big Data Knowledge System in Healthcare	45
3.4.1	Sources of Medical Big Data	51
3.4.2	Knowledge in Healthcare	53
3.4.3	Big Data Knowledge Management Systems in Healthcare	55
3.4.4	Big Data Analytics in Healthcare	56
3.5	Big Data Applications in the Healthcare Sector	59
3.5.1	Real Time Healthcare Monitoring and Altering	59
3.5.2	Early Disease Prediction with Big Data	59
3.5.3	Patients Predictions for Improved Staffing	61
3.5.4	Medical Imaging	61
3.6	Challenges with Healthcare Big Data	62
3.6.1	Challenges of Big Data	62

3.6.2	Challenges of Healthcare Big Data	62
3.7	Conclusion	64
	References	64
4	Big Data for Personalized Healthcare	67
	<i>Dhanalakshmi R. and Jose Anand</i>	
4.1	Introduction	68
4.1.1	Objectives	68
4.1.2	Motivation	69
4.1.3	Domain Description	70
4.1.4	Organization of the Chapter	70
4.2	Related Literature	71
4.2.1	Healthcare Cyber Physical System Architecture	71
4.2.2	Healthcare Cloud Architecture	71
4.2.3	User Authentication Management	72
4.2.4	Healthcare as a Service (HaaS)	72
4.2.5	Reporting Services	73
4.2.6	Chart and Trend Analysis	73
4.2.7	Medical Data Analysis	73
4.2.8	Hospital Platform Based On Cloud Computing	74
4.2.9	Patient's Data Collection	74
4.2.10	H-Cloud Challenges	75
4.2.11	Healthcare Information System and Cost	75
4.3	System Analysis and Design	75
4.3.1	Proposed Solution	76
4.3.2	Software Components	76
4.3.3	System Design	76
4.3.4	Architecture Diagram	77
4.3.5	List of Modules	78
4.3.6	Use Case Diagram	81
4.3.7	Sequence Diagram	81
4.3.8	Class Diagram	82
4.4	System Implementation	83
4.4.1	User Interface	83
4.4.2	Storage Module	84
4.4.3	Notification Module	85
4.4.4	Middleware	86
4.4.5	OTP Module	87
4.5	Results and Discussion	88
4.6	Conclusion	90
	References	90

5	Knowledge Engineering for AI in Healthcare	93
	<i>A. Thirumurthi Raja and B. Mahalakshmi</i>	
5.1	Introduction	94
5.2	Overview	95
5.2.1	Knowledge Representation	95
5.2.2	Types of Knowledge in Artificial Intelligence	96
5.2.3	Relation Between Knowledge and Intelligence	97
5.2.4	Approaches to Knowledge Representation	97
5.2.5	Requirements for Knowledge Representation System	98
5.2.6	Techniques of Knowledge Representation	98
5.2.6.1	Logical Representation	99
5.2.6.2	Semantic Network Representation	99
5.2.6.3	Frame Representation	99
5.2.6.4	Production Rules	100
5.2.7	Process of Knowledge Engineering	101
5.2.8	Knowledge Discovery Process	106
5.3	Applications of Knowledge Engineering in AI for Healthcare	106
5.3.1	AI Supports in Clinical Decisions	107
5.3.2	AI-Assisted Robotic Surgery	107
5.3.3	Enhance Primary Care and Triage	108
5.3.4	Clinical Judgments or Diagnosis	108
5.3.5	Precision Medicine	109
5.3.6	Drug Discovery	109
5.3.7	Deep Learning to Diagnose Diseases	110
5.3.8	Automating Administrative Tasks	111
5.3.9	Reducing Operational Costs	112
5.3.10	Virtual Nursing Assistants	113
5.4	Conclusion	113
	References	114
6	Business Intelligence and Analytics from Big Data to Healthcare	115
	<i>Maheswari P., A. Jaya and João Manuel R. S. Tavares</i>	
6.1	Introduction	116
6.1.1	Impact of Healthcare Industry on Economy	116
6.1.2	Coronavirus Impact on the Healthcare Industry	117
6.1.3	Objective of the Study	117
6.1.4	Limitations of the Study	117
6.2	Related Works	118
6.3	Conceptual Healthcare Stock Prediction System	120
6.3.1	Data Source	122
6.3.2	Business Intelligence and Analytics Framework	122

6.3.2.1	Simple Machine Learning Model	122
6.3.2.2	Time Series Forecasting	123
6.3.2.3	Complex Deep Neural Network	123
6.3.3	Predicting the Stock Price	124
6.4	Implementation and Result Discussion	124
6.4.1	Apollo Hospitals Enterprise Limited	125
6.4.2	Cadila Healthcare Ltd	125
6.4.3	Dr. Reddy's Laboratories	128
6.4.4	Fortis Healthcare Limited	130
6.4.5	Max Healthcare Institute Limited	131
6.4.6	Opto Circuits Limited	131
6.4.7	Panacea Biotec	135
6.4.8	Poly Medicure Ltd	136
6.4.9	Thyrocare Technologies Limited	138
6.4.10	Zyduz Wellness Ltd	138
6.5	Comparisons of Healthcare Stock Prediction Framework	141
6.6	Conclusion and Future Enhancement	143
	References	143
	Books	145
	Web Citation	145
7	Internet of Things and Big Data Analytics for Smart Healthcare	147
	<i>Sathish Kumar K., Om Prakash P.G., Alangudi Balaji N. and Robertas Damaševičius</i>	
7.1	Introduction	148
7.2	Literature Survey	149
7.3	Smart Healthcare Using Internet of Things and Big Data Analytics	151
7.3.1	Smart Diabetes Prediction	151
7.3.2	Smart ADHD Prediction	154
7.4	Security for Internet of Things	159
7.4.1	K(Binary) ECC FSM	159
7.4.2	NAF Method	160
7.4.3	K-NAF Multiplication Architecture	161
7.4.4	K(NAF) ECC FSM	161
7.5	Conclusion	164
	References	165
8	Knowledge-Driven and Intelligent Computing in Healthcare	167
	<i>R. Mervin, Dinesh Mavalaru and Tintu Thomas</i>	
8.1	Introduction	168

8.1.1	Basics of Health Recommendation System	169
8.1.2	Basics of Ontology	169
8.1.3	Need of Ontology in Health Recommendation System	170
8.2	Literature Review	171
8.2.1	Ontology in Various Domain	172
8.2.2	Ontology in Health Recommendation System	174
8.3	Framework for Health Recommendation System	175
8.3.1	Domain Ontology Creation	176
8.3.2	Query Pre-Processing	178
8.3.3	Feature Selection	179
8.3.4	Recommendation System	180
8.4	Experimental Results	182
8.5	Conclusion and Future Perspective	183
	References	183
9	Secure Healthcare Systems Based on Big Data Analytics	189
	<i>A. Angel Cerli, K. Kalaiselvi and Vijayakumar Varadarajan</i>	
9.1	Introduction	190
9.2	Healthcare Data	193
9.2.1	Structured Data	193
9.2.2	Unstructured Data	194
9.2.3	Semi-Structured Data	194
9.2.4	Genomic Data	194
9.2.5	Patient Behavior and Sentiment Data	194
9.2.6	Clinical Data and Clinical Notes	194
9.2.7	Clinical Reference and Health Publication Data	195
9.2.8	Administrative and External Data	195
9.3	Recent Works in Big Data Analytics in Healthcare Data	195
9.4	Healthcare Big Data	197
9.5	Privacy of Healthcare Big Data	198
9.6	Privacy Right by Country and Organization	200
9.7	How Blockchain is Big Data Usable for Healthcare	200
9.7.1	Digital Trust	200
9.7.2	Smart Data Tracking	202
9.7.3	Ecosystem Sensible	202
9.7.4	Switch Digital	202
9.7.5	Cybersecurity	203
9.7.6	Sharing Interoperability and Data	203
9.7.7	Improving Research and Development (R&D)	206
9.7.8	Drugs Fighting Counterfeit	206
9.7.9	Patient Mutual Participation	206

9.7.10	Internet Access by Patient to Longitudinal Data	206
9.7.11	Data Storage into Off Related to Confidentiality and Data Scale	207
9.8	Blockchain Threats and Medical Strategies Big Data Technology	207
9.9	Conclusion and Future Research	208
	References	208
10	Predictive and Descriptive Analysis for Healthcare Data	213
	<i>Pritam R. Ahire and Rohini Hanchate</i>	
10.1	Introduction	214
10.2	Motivation	215
10.2.1	Healthcare Analysis	215
10.2.2	Predictive Analytics	217
10.2.3	Predictive Analytics Current Trends	217
10.2.3.1	Importance of PA	217
10.2.4	Descriptive Analysis	218
10.2.4.1	Descriptive Statistics	218
10.2.4.2	Categories of Descriptive Analysis	219
10.2.5	Method of Modeling	221
10.2.6	Measures of Data Analytics	221
10.2.7	Healthcare Data Analytics Platforms and Tools	223
10.2.8	Challenges	225
10.2.9	Issues in Predictive Healthcare Analysis	226
10.2.9.1	Integrating Separate Data Sources	226
10.2.9.2	Advanced Cloud Technologies	226
10.2.9.3	Privacy and Security	227
10.2.9.4	The Fast Pace of Technology Changes	227
10.2.10	Applications of Predictive Analysis	227
10.2.10.1	Improving Operational Efficiency	227
10.2.10.2	Personal Medicine	228
10.2.10.3	Population Health and Risk Scoring	228
10.2.10.4	Outbreak Prediction	228
10.2.10.5	Controlling Patient Deterioration	228
10.2.10.6	Supply Chain Management	228
10.2.10.7	Potential in Precision Medicine	229
10.2.10.8	Cost Savings From Reducing Waste and Fraud	229
10.3	Conclusion	229
	References	229

11 Machine and Deep Learning Algorithms for Healthcare Applications	233
<i>K. France, A. Jaya and Doru Tiliute</i>	
11.1 Introduction	234
11.2 Artificial Intelligence, Machine Learning, and Deep Learning	234
11.3 Machine Learning	236
11.3.1 Supervised Learning	236
11.3.2 Unsupervised Learning	238
11.3.3 Semi-Supervised	238
11.3.4 Reinforcement Learning	238
11.4 Advantages of Using Deep Learning on Top of Machine Learning	239
11.5 Deep Learning Architecture	239
11.6 Medical Image Analysis using Deep Learning	242
11.7 Deep Learning in Chest X-Ray Images	243
11.8 Machine Learning and Deep Learning in Content-Based Medical Image Retrieval	246
11.9 Image Retrieval Performance Metrics	249
11.10 Conclusion	250
References	250
12 Artificial Intelligence in Healthcare Data Science with Knowledge Engineering	255
<i>S. Asha, Kanchana Devi V. and G. Sahaja Vaishnavi</i>	
12.1 Introduction	256
12.2 Literature Review	260
12.3 AI in Healthcare	266
12.4 Data Science and Knowledge Engineering for COVID-19	268
12.5 Proposed Architecture and Its Implementation	270
12.5.1 Implementation	270
12.5.1.1 Data Collection	270
12.5.1.2 Understanding Class and Dependencies	270
12.5.1.3 Pre-Processing	272
12.5.1.4 Sampling	273
12.5.1.5 Model Fixing	273
12.5.1.6 Analysis of Real-Time Datasets	273
12.5.1.7 Machine Learning Algorithms	276
12.6 Conclusions and Future Work	278
References	280

13 Knowledge Engineering Challenges in Smart Healthcare	
Data Analysis System	285
<i>Agasba Saroj S. J., B. Saleena and B. Prakash</i>	
13.1 Introduction	285
13.1.1 Motivation	287
13.2 Ongoing Research on Intelligent Decision Support System	289
13.3 Methodology and Architecture of the Intelligent Rule-Based System	291
13.3.1 Proposed System Design	292
13.3.2 Algorithms Used	293
13.3.2.1 Forward Chaining	293
13.3.2.2 Backward Chaining	294
13.4 Creating a Rule-Based System using Prolog	295
13.5 Results and Discussions	304
13.6 Conclusion	306
13.7 Acknowledgments	307
References	307
14 Big Data in Healthcare: Management, Analysis, and Future Prospects	309
<i>A. Akila, R. Parameswari and C. Jayakumari</i>	
14.1 Introduction	309
14.2 Breast Cancer: Overview	310
14.3 State-of-the-Art Technology in Treatment of Cancer	311
14.3.1 Chemotherapy	311
14.3.2 Radiotherapy	311
14.4 Early Diagnosis of Breast Cancer: Overview	312
14.4.1 Advantages and Risks Associated with the Early Detection of Breast Cancer	312
14.4.2 Diagnosis the Breast Cancer	313
14.5 Literature Review	314
14.6 Machine Learning Algorithms	315
14.6.1 Principal Component Analysis Algorithms	316
14.6.2 K-Means Algorithm	317
14.6.3 K-Nearest Neighbor Algorithm	317
14.6.4 Logistic Regression Algorithm	318
14.6.5 Support Vector Machine Algorithm	318
14.6.6 AdaBoost Algorithm	319
14.6.7 Neural Networks Algorithm	319
14.6.8 Random Forest Algorithm	319
14.7 Result and Discussion	320

14.7.1	Performance Metrics	320
14.7.1.1	ROC Curve	320
14.7.1.2	Accuracy	321
14.7.1.3	Precision and Recall	321
14.7.1.4	F1-Score	322
14.8	Experimental Result and Discussion	322
14.9	Conclusion	324
	References	325
15	Machine Learning for Information Extraction, Data Analysis and Predictions in the Healthcare System	327
	<i>G. Jaculine Priya and S. Saradha</i>	
15.1	Introduction	327
15.2	Machine Learning in Healthcare	329
15.3	Types of Learnings in Machine Learning	331
15.3.1	Supervised Learning	332
15.3.2	Unsupervised Algorithms	333
15.3.3	Semi-Supervised Learning	334
15.3.4	Reinforcement Learning	334
15.4	Types of Machine Learning Algorithms	334
15.4.1	Classification	335
15.4.2	Bayes Classification	335
15.4.3	Association Analysis	335
15.4.4	Correlation Analysis	336
15.4.5	Cluster Analysis	336
15.4.6	Outlier Analysis	336
15.4.7	Regression Analysis	337
15.4.8	K-Means	337
15.4.9	Apriori Algorithm	337
15.4.10	K Nearest Neighbor	337
15.4.11	Naive Bayes	338
15.4.12	AdaBoost	338
15.4.13	Support Vector Machine	338
15.4.14	Classification and Regression Trees	339
15.4.15	Linear Discriminant Analysis	339
15.4.16	Logistic Regression	339
15.4.17	Linear Regression	339
15.4.18	Principal Component Analysis	339
15.5	Machine Learning for Information Extraction	340
15.5.1	Natural Language Processing	340
15.6	Predictive Analysis in Healthcare	341

15.7	Conclusion	342
	References	342
16	Knowledge Fusion Patterns in Healthcare	345
	<i>N. Deepa and N. Kanimozhi</i>	
16.1	Introduction	346
16.2	Related Work	348
16.3	Materials and Methods	349
	16.3.1 Classification of Data Fusion	349
	16.3.2 Levels and Its Working in Healthcare Ecosystems	351
	16.3.2.1 Initial Level Data Access (ILA)	351
	16.3.2.2 Middle Level Access (MLA)	352
	16.3.2.3 High Level Access (HLA)	352
16.4	Proposed System	352
	16.4.1 Objective	353
	16.4.2 Sample Dataset	355
16.5	Results and Discussion	355
16.6	Conclusion and Future Work	361
	References	362
17	Commercial Platforms for Healthcare Analytics: Health Issues for Patients with Sickle Cells	365
	<i>J.K. Adedeji, T.O. Owolabi and R.S. Fayose</i>	
17.1	Introduction	366
17.2	Materials and Methods	367
	17.2.1 Data Acquisition and Pre-Processing	367
	17.2.2 Sickle Cells Normalization Image	368
	17.2.3 Gradient Calculation	369
	17.2.4 Gradient Descent Step	371
	17.2.5 Insight to Previous Methods Adopted in Convolutional Neural Networks	372
	17.2.6 Segments of Convolutional Neural Networks	372
	17.2.6.1 Convolutional Layer	372
	17.2.6.2 Pooling Layer	373
	17.2.6.3 Fully Connected Layer	374
	17.2.6.4 Softmax Layer	374
	17.2.7 Basic Transformations of Convolutional Neural Networks in Healthcare	374
	17.2.8 Algorithm Review and Comparison	376
	17.2.9 Feedforward	376
17.3	Results and Discussion	377

17.3.1	Results on Suitability for Applications in Healthcare	377
17.3.2	Class Prediction	377
17.3.3	The Model Sanity Checking	377
17.3.4	Analysis of the Epoch and Training Losses	378
17.3.5	Discussion and Healthcare Interpretations	379
17.3.6	Load Data	379
17.3.7	Image Pre-Processing	380
17.3.8	Building and Training the Classifier	381
17.3.9	Saving the Checkpoint Suitable for Healthcare	382
17.3.10	Loading the Checkpoint	383
17.4	Conclusion	383
	References	383
18	New Trends and Applications of Big Data Analytics for Medical Science and Healthcare	387
	<i>Niha K. and Aisha Banu W.</i>	
18.1	Introduction	388
18.2	Related Work	389
18.3	Convolutional Layer	389
18.4	Pooling Layer	390
18.5	Fully Connected Layer	390
18.6	Recurrent Neural Network	391
18.7	LSTM and GRU	392
18.8	Materials and Methods	397
18.8.1	Pre-Processing Strategy Selection	397
18.8.2	Feature Extraction and Classification	400
18.9	Results and Discussions	406
18.10	Conclusion	408
18.11	Acknowledgement	409
	References	409
Index		413

Preface

The power of healthcare data analytics is being increasingly used in the industry. With this in mind, we wanted to write a book geared towards those who want to learn more about the techniques used in healthcare analytics for efficient analysis of data. Since data is generally generated in enormous amounts and pumped into data pools, analyzing data patterns can help to ensure a better quality of life for patients. As a result of small amounts of health data from patients suffering from various health issues being collectively pooled, researchers and doctors can find patterns in the statistics, helping them develop new ways of forecasting or diagnosing health issues, and identifying possible ways to improve quality clinical care. Big data analytics supports this research by applying various processes to examine large and varied healthcare data sets. Advanced analytics techniques are used against large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information. This book covers both the theory and application of the tools, techniques and algorithms for use in big data in healthcare and clinical research. It provides the most recent research findings to derive knowledge using big data analytics, which helps to analyze huge amounts of real-time healthcare data, the analysis of which can provide further insights in terms of procedural, technical, medical, and other types of improvements in healthcare. In addition, this book also explores various sources of personalized healthcare data.

For those who are healthcare researchers, this book reveals the innovative hybrid machine learning and deep learning techniques applied in various healthcare data sets. Since machine learning algorithms play a major role in analyzing the volume, veracity and velocity of big data, the scope of this book focuses on various kinds of machine learning algorithms existing in the areas such as supervised, unsupervised, semi-supervised, and reinforcement learning. It guides readers in implementing the Python environment for machine learning in various application domains. Furthermore, predictive analytics in healthcare is explored, which can help to detect

early signs of patient deterioration from the ICU to a general ward, identify at-risk patients in their homes to prevent hospital readmissions, and prevent avoidable downtime of medical equipment.

Also explored in the book are a wide variety of machine learning techniques that can be applied to infer intelligence from the data set and the capabilities of an application. The significance of data sets for various applications is also discussed along with sample case studies. Moreover, the challenges presented by the techniques and budding research avenues necessary to see their further advancement are highlighted.

Patient's healthcare data needs to be protected by organizations in order to prevent data loss through unauthorized access. This data needs to be protected from attacks that can encrypt or destroy data, such as ransomware, as well as those attacks that can modify or corrupt a patient's data. Security is paramount since a lot of devices are connected through the internet of things and serve many healthcare applications, including supporting smart healthcare systems in the management of various diseases such as diabetes, monitoring heart functions, predicting heart failure, etc. Therefore, this book explores the various challenges for smart healthcare, including privacy, confidentiality, authenticity, loss of information, attacks, etc., which create a new burden for providers to maintain compliance with healthcare data security.

In addition to inferring knowledge fusion patterns in healthcare, the book also explores the commercial platforms for healthcare data analytics. The new benefits that healthcare data analytics brings to the table, run analytics and unearth information that could be used in the decision-making of practitioners by providing insights that can be used to make immediate decisions. Also investigated are the new trends and applications of big data analytics for medical science and healthcare. Healthcare professionals, researchers, and practitioners who wish to figure out the core concepts of smart healthcare applications and the innovative methods and technologies used in healthcare will all benefit from this book.

Editors

Dr. A. Jaya

Dr. K. Kalaiselvi*

Dr. Dinesh Goyal

Prof. Dhiya AL-Jumeily

**Corresponding Editor*

An Introduction to Knowledge Engineering and Data Analytics

D. Karthika* and K. Kalaiselvi†

Department of Computer Applications, Vels Institute of Science, Technology & Advanced Studies (Formerly Vels University), Chennai, Tamil Nadu, India

Abstract

In recent years, the philosophy of Knowledge Engineering has become important. Information engineering is an area of system engineering which meets unclear process demands by emphasizing the development of knowledge in a knowledge-based system and its representation. A broad architecture for knowledge engineering that manages the fragmented modeling and online learning of knowledge from numerous sources of information, non-linear incorporation of fragmented knowledge, and automatic demand-based knowledge navigation. The project aims to provide petabytes in the defined application domains with data and information tools. Knowledge-based engineering (KBE) frameworks are based on the working standards and core features with a special focus on their built-in programming language. This language is the key element of a KBE framework and promotes the development and re-use of the design skills necessary to model complex engineering goods. This facility allows for the automation of the process preparation step of multidisciplinary research (MDA), which is particularly important for this novel. The key types of design rules to be implemented in the implementation of the KBE are listed, and several examples illustrating the significant differences between the KBE and the traditional CAD approaches are presented. This chapter discusses KBE principles and how this technology will facilitate and enable the multidisciplinary optimization (MDO) of the design of complex products. This chapter discusses their reach goes beyond existing CAD structure constraints and other practical parametric and space exploration approaches. There is a discussion of the concept of KBE and its usage in architecture that supports the use of MDO. Finally, this chapter discusses on the key measures and latest trends in the development of KBE.

*Corresponding author: d.karthi666@gmail.com

†Corresponding author: kalairaghu.scs@velsuniv.ac.in

Keywords: Data analytics, knowledge, knowledge engineering, principles, knowledge acquisition

1.1 Introduction

1.1.1 Online Learning and Fragmented Learning Modeling

Applied artificial intelligence (AI) was defined as knowledge engineering [1], with three major scientific questions: knowledge representation, the use of information, and the acquisition of knowledge. In the big data age, the three fundamental problems must evolve with the basic characteristics of the complex and evolving connections between data objects, which are autonomous sources of information. Big data not only rely on domain awareness but also distribute knowledge from numerous information sources. To have knowledge of engineering tools for big data, we need tons of experience. Three primary research issues need to be addressed for the 54-month, RMB 45-million, 15-year Big Data Knowledge Engineering (BigKE) project sponsored by China's Ministry of Science and Technology and several other domestic agencies: 1) online learning and fragmented learning modeling; 2) nonlinear fusion of fragmented information; and 3) multimedia fusion of knowledge. Discussing these topics is the main contribution of this article. With 1), we examine broken information and representation clusters, immersive online content learning with fragmented knowledge, and simulation with spatial and temporal characteristics of evolving knowledge. Question 2) will discuss connections, a modern pattern study, and dynamic integration between skills subsections of fragmented information. The key issues mentioned in Figure 1.1 will be collaborative, context-based computing, information browsing, route discovery, and the enhancement of interactive knowledge adaptation.

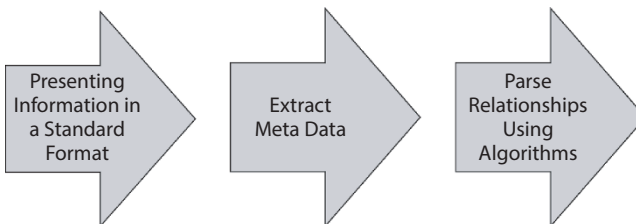


Figure 1.1 Knowledge engineering.

Due to these features of several channels, traditional offline data mining methods cannot stream data since it is important to reform the data. Online learning methods help solve this issue and adapt easily to the drift of the effects of streaming. But typical methods to online learning are explicitly configured for single-source info. Thus, the maintenance of these features concurrently provides great difficulties and opportunities for large-scale data production. Big data starts with global details, tackles dispersed data like data sources and function streams, and integrates diverse understanding from multiple data channels, as well as domain experience in personalized demand-driven knowledge services. In the age of big data, many data sources are usually heterogeneous and independent and require evolving, complex connections among data objects. These qualities are considered by substantial experience. Meanwhile, major suppliers of information provide personalized and in-house demand-driven offerings through the usage of large-scale information technologies [2].

Centered on the characteristics of multiple data sets, the key to a multi-source retrieval of information is fragmented data processing [3]. To create global awareness, local information pieces from individual data points can be merged. Present online learning algorithms often use linear fitting for the retrieval of dispersed knowledge from local data sources [4]. In the case of fragmented knowledge fusion, though, linear fitting is not successful and may even create problems of overfitting. Several studies are ongoing to improve coherence in the processing and interpretation of fragmented knowledge [6], and the advantage of machine learning for large data interpreting is that most samples are efficient, thus eliminating the possibility of over-adjustment at any rate [7]. Big data innovation acquires knowledge mostly from user-produced content, as opposed to traditional information engineering's focused on domain experience, in addition to authoritative sources of knowledge, such as technical knowledge bases. The content created by users provides a new type of database that could be used as a main human information provider as well as to help solve the problem of bottlenecks in traditional knowledge engineering. The information created by the consumer is broad and heterogeneous which leads to storage and indexing complexities [5], and the knowledge base should be able to build and disseminate itself to establish realistic models of data relations. For instance, for a range of reasons, clinical findings in survey samples can be incomplete and unreliable, and preprocessing is needed to improve analytical data [8].

Both skills are essential for the creation of personalized knowledge base tools as the knowledge base should be targeted to the needs of individual users. Huge information reinforces distributed expertise to develop any ability. Big data architecture also requires a customer interface to overcome user-specific problems. With the advent of science and innovations, in the fast-changing knowledge world of today, the nature of global economic growth has changed by introducing more communication models, shorter product life cycles, and a modern product production rate. Knowledge engineering is an AI field that implements systems based on information. Such structures provide computer applications with a broad variety of knowledge, policy, and reasoning mechanisms that provide answers to real-world issues. Difficulties dominated the early years of computer technology. Also, knowledge engineers find that it is a very long and expensive undertaking to obtain appropriate quality knowledge to construct a reliable and usable system. The construction of an expert system was identified as a knowledge learning bottleneck. This helped to gain skills and has been a big area of research in the field of information technology.

The purpose of gathering information is to create strategies and tools that make it as simple and effective as possible to gather and verify a professional's expertise. Experts tend to be critical and busy individuals, and the techniques followed would also minimize the time expended on knowledge collection sessions by each expert. The key form of the knowledge-based approach is an expert procedure, which is intended to mimic an expert's thinking processes. Typical examples of specialist systems include bacterial disease control, mining advice, and electronic circuit design assessment. It currently refers to the planning, administration, and construction of a system centered on expertise. It operates in a broad variety of aspects of computer technology, including data baselines, data collection, advanced networks, decision-making processes, and geographic knowledge systems. This is a big part of software computing. Wisdom engineering also falls into connection with mathematical reasoning as well as a strong concern with cognitive science and social-cognitive engineering, where intelligence is generated by socio-cognitive aggregates (mainly human beings) and structured according to the way humans thought and logic operates. Since then, information engineering has been an essential technology for knowledge incorporation. In the end, the exponentially increasing World Wide Web generates a growing market for improved usage of knowledge and technological advancement.

1.2 Knowledge and Knowledge Engineering

1.2.1 Knowledge

Knowledge is characterized as abilities, (i) which the person gains in terms of practice or learning; theoretical or practical knowledge of a subject; (ii) what is known inside or as a whole; facts and information; or (iii) knowing or acquainted with the experience of life or situation. This knowledge is defined accordingly. The retrieval of information requires complex cognitive processes including memory, understanding, connectivity, association, and reasoning. Knowledge of a subject and its capacity for usage for a specific purpose are also used to suggest trustworthy comprehension. Information may be divided into two forms of knowledge: tacit and clear. Tacit knowledge is the awareness that people have and yet cannot get. Tacit knowledge is more relevant since it gives people, locations, feelings, and memories a framework. Efficient tacit information transfer typically requires intensive intimate correspondence and trust. The tacit understanding is not easily shared. Still, consciousness comprises patterns and culture that we still cannot comprehend. On the other hand, information, which is easy to articulate, is called explicit knowledge. Coding or codification is the tool used to translate tacit facts into specific details. The awareness expressed, codified, and stored in such media is explicitly facts. Explicit information. The most common simple knowledge sources are guides, manuals, and protocols. Audio-visual awareness may also be an example of overt intelligence, which is based on the externalization of human skills, motivations, and knowledge.

1.2.2 Knowledge Engineering

Edward Feigenbaum and Pamela McCorduck created Knowledge Engineering in 1983: To address difficult issues that typically need a great deal of human experience in the fields of engineering, knowledge engineering needs the integration of information of computer systems. In engineering, design information is an essential aspect. If the information is collected and held in the knowledge base, important cost and output gains may be accomplished. In a range of fields, information base content may be used as to reuse information in other ways for diverse goals, to employ knowledge to create smart systems capable of carrying out complicated design work. We shall disseminate knowledge to other individuals within an organization.

While the advantages of information capture and usage are obvious, it has long been known in the AI world that knowledge is challenging to access from specialists. Second, the specialists do not remember and describe “tacit knowledge” effectively, and this operates subconsciously and, where it is not impossible, is difficult to overcome the problems arising from several subject matters that they learn. To elaborate, they have to know what it is called. Third, there are various prospects and points of view which include aggregation to provide a coherent view. Last, professionals create abstract concepts and shortcuts for which they cannot communicate. The area of information technology was created some 25 years ago to address such problems, and the role of the knowledge engineer was born. Since then, computer engineers have developed a variety of principles, methods, and tools that have improved the acquisition, use, and implementation of knowledge considerably.

1.3 Knowledge Engineering as a Modelling Process

There is also a consensus that the KBS construction approach may be used as a modeling operation. It is not intended to construct a cognitively appropriate model, but to build a model that offers similar results for problem-solving problems in the area of concern as seen in Figure 1.2.

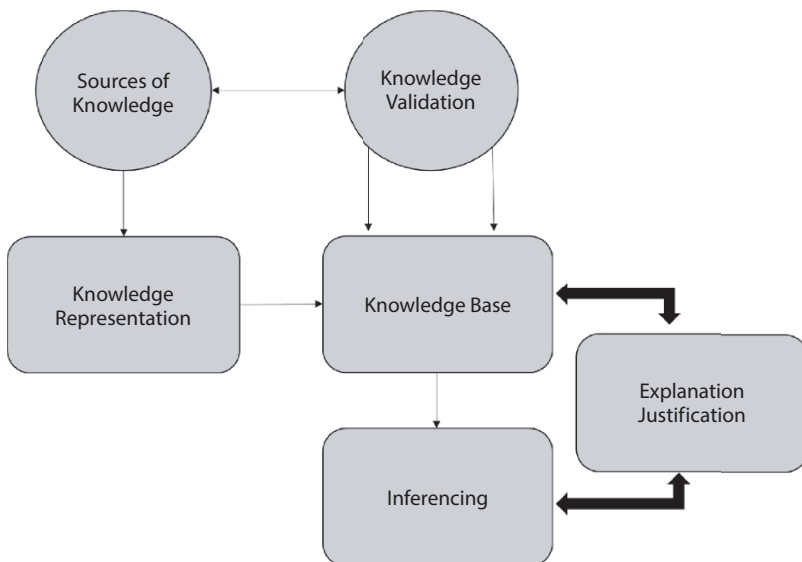


Figure 1.2 Knowledge as modelling process.

Building a KBS requires building a programming model to acquire problem-solving capabilities like those of a domain specialist. This material is not directly available; therefore, it needs to be created and arranged.

1.4 Tools

Intelligence engineers make the more efficient and less bogus use of dedicated computational tools for the acquisition, simulation, and handling of intelligence. PC PACK is a versatile compilation of this programmed, commercially available as a package of knowledge technology tools that are designed to be tested on a wide range of projects. The aim is to consider the key characteristics of the domain. The method simulates how anyone should label a text page with different colors such as green for suggestions and yellow for attributes. The labeled text would immediately be placed in the PCPACK database to be applied to all other resources when the user has highlighted a document. The MOKA and Popular KADS Methodologies are supported by the CFPACK. It is also fully compliant with information engineering approaches and techniques. The CFPACK is a software suite that includes the following

- (i) Protocol tool: It enables the discovery, recognition, and definition of interview transcripts, conclusions, and documentation that may be included in the knowledge base.
- (ii) Ladder Tool: This allows hierarchies of knowledge elements such as meanings, features, procedures, and specifications to be developed.
- (iii) Chart Tool: This allows users to build mobile networks of connections between data elements, such as process maps, maps of ideas, and cutting-edge diagrams.
- (iv) Matrix Tool: This allows grids that show the connection and attributes of the elements to be developed and edited.
- (v) Annotation tools: This facilitates the creation of sophisticated HTML annotations, with links to other sites and other knowledge templates automatically generated in the CFPACK.
- (vi) Tool publisher: This allows the creation from a knowledge base of a website or some other information resource using a model-driven approach to optimize re-usability. MOKA, CommonKADS, and the 47-step protocol provide approaches to run a project from beginning to completion, as well as maintaining best practice.

1.5 What are KBSs?

A knowledge-based framework is a system that utilizes AI tools in problem-solving systems to assist human decision-making, understanding, and intervention.

There are two core components of the KBSs:

- Information base (consists of a collection of details and a set of laws, structures, or procedures).
- Inference engine (Responsible for the extension of the information base to the issue at hand).

In contrast to human expertise, there are pros and cons to utilizing KBSs.

1.5.1 What is KBE?

It starts with a discussion about what KBE is in this book and begins with a simple definition: Knowledge-based Engineering (KBE) uses the knowledge of product and operation, which was collected and retained in specific software applications, to enable its direct usage and reuse in the development of new products and variants. KBE's implementation consists of applying a specific class of computing tools, called the KBE systems, which enable engineers to acquire and reuse engineering knowledge using methods and methodologies. The name of the KBE architecture is derived from the mixture of KBS and engineering, which are one of the major outcomes of AI. In the 1970s, KBE systems demonstrate the advancement of the KBS by applying the special engineering industry requirements. KBE systems combine KBS rule-based logic technologies with engineering data analysis and geometry like CAD.

For these reasons, a traditional KBE architecture provides the user with a programming language that is generally object-oriented and one (something more) embedded or closely connected CAD engine. The vocabulary of programming enables the user to capture and reuse rules and procedures in engineering, while the object-oriented approach of design corresponds well with how engineers see the world: systems abstract assets of objects, defined by parameters and behaviors, linked by relationships. Access and management by the programming language of the CAD engine meet the geometry handling requirements characteristic of the engineering architecture. The MIT AI laboratories and the Computer Vision CAD Group

developed the first commercially available KBE system named ICAD1984 (now PTC). Fortunately, this asks the first inquiry you will hear in Figure 1.3 about KBE.

It will be useful at this stage to quickly clarify what we refer to as information and how we use this term to describe concepts other than data and evidence. Both terms are sometimes misused in the traditional spoken language; truth and knowledge are often interchangeably used. The hierarchy of data and intellect (and knowledge) is the subject of long-term disputes between epistemologists and IT experts. Since this subject goes well beyond the scope of this chapter, this is our definition of data. Data are objects that have no meaning before they are put in form like symbols, statistics, digits, and indications. The information consists of important processed data. The context in which the data are collected gives it meaning, importance, and intent. Human and electronic information can be collected, shared, and processed. The knowledge is encrypted by code, normally organized in a structure or size, and stored in hard or soft media to accomplish this. Awareness is the condition of information and awareness processing, which requires the chance to act.

New information may be produced as a result of the application of knowledge. The IGES file with a geometrical definition of the surface as a piece of information is an example of this. IGES files will be encoded with numbers and symbols (i.e., data) and will only provide knowledge useful if they understand the meaning (i.e., the fact they are the data of an

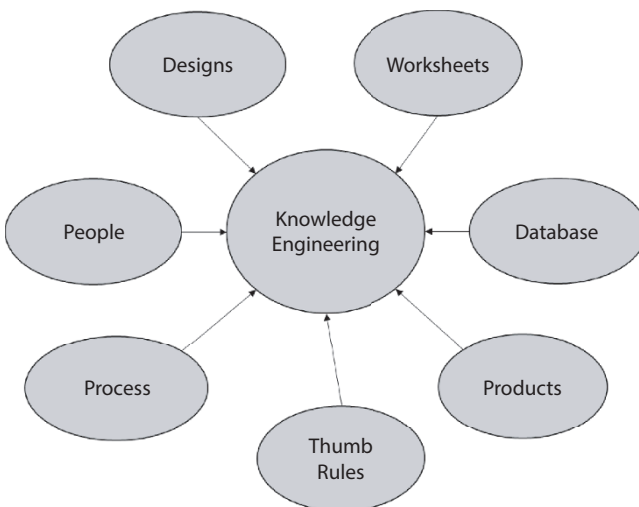


Figure 1.3 KBE.

IGES file). The information which can be collected with a KBE method is regarded as a simple example of the algorithm that reads such an IGES file, reconstructs a specified surface model, intersects it with a floor plane, and, if the crossroad is non-zero, calculates the length of the corresponding curve. It is also sensible to ask why geometry varies from the standard CAD paradigm and enables the creation and manipulation of geometry. Owing to the varying scopes of these systems, the differences are important. Digitized drawing systems, which allow programmers to catch their ideas have been designed to create CAD systems. They build and store the results using the CAD framework's geometry simulation functions. A set of points, lines, planes, and solids with reference and note are an almost all-inclusive link to the structure. These data provide enough information for the creation of a system that can be used to build a specification by production engineers. In doing so, creators store the specifics of "what," but they retain the "how" and "why." In a sense, the CAD approach can be considered a system "posterior," because before it can be moved to the system it is necessary to know what the principle is like. It can be argued that CAD is geometry or drawing/drawing engineering to distinguish this approach from KBE.

KBE-supported technology is different. Technology Instead of shifting "what," engineering experts are trying to move "how" and "why," encapsulating in the KBE process knowledge and thinking instead of geometry in the CAD framework. Not only does this work by manipulating geometric structures, but programming is needed rather than writing. The "how" and "why" in engineering are in some cases used in textbooks, databases, tip sheets, and several other outlets. Much of the knowledge is held up by engineers, mostly in a manner that is strongly compiled and not specifically suitable for translation to the KBE procedure. This experience should be sufficiently transparent to create a KBE program to be codified into a software application capable of producing all kinds of product specifics, including geometry templates, scores, and data that are not associated with the geometry. Because of its capacity to generate a specification rather than simply text, it is widely referred to as a generative model.

1.5.2 When Can KBE Be Used?

How easy is the use of KBE? It is only necessary to rapidly generate different configurations and variants of a given variable. In certain practical cases, this is not essential, so it may be a wrong expenditure to try to clarify details and program a KBE system. One-off prototypes and designs that