

Case-Based Reasoning

Michael M. Richter • Rosina O. Weber

Case-Based Reasoning

A Textbook

 Springer

Michael M. Richter
Fachbereich Informatik (retired)
Universität Kaiserslautern
Kaiserslautern, Germany

Rosina O. Weber
The College of Computing & Informatics
Drexel University
Philadelphia, PA, USA

ISBN 978-3-642-40166-4

ISBN 978-3-642-40167-1 (eBook)

DOI 10.1007/978-3-642-40167-1

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013953284

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

A statement of George Bernhard Shaw is essentially this:

“The wisdom of a person is not measured by the number of experiences but rather by the ability to make use of them”.

This characterises case-based reasoning (CBR) to a large degree. It is easy to record billions of experiences in a database. However, the central question is: If I have a problem, which experience should I consult and in what way? Humans have cultivated this in very clever ways. So, what prevents us from collecting experiences on a computer? Nothing: Collection is easily done. The problem is now to realise the second part of Shaw’s statement on a computer. This has been neglected for a long time and is still not a standard in university education.

This difficulty is obvious: The new problem is not exactly like that in the experience base, and even if we find a good one quite similar, it cannot be used exactly in the same way as it was used in the past.

This book presents a systematic approach. It has two goals:

- (1) to present rigorous and formally valid structures for precise case-based reasoning;
- (2) to introduce you to the universe of CBR applications through the understanding of techniques, methods, and tools for many quite different tasks.

The basic formalisms are:

- A process model. This describes on an abstract level the steps that one has to perform when working with CBR. This incorporates, in particular, what to do in order to find useful experiences and how to apply them once they are selected.
- The knowledge containers. They describe how and where knowledge is stored that is needed for executing the steps of the process model.

There are two ways of working with CBR. The experience-based view compares a new problem to a previous experience. The extended view does not use experiences; it searches for a solution of a problem by directly associating a solution to the problem. In both approaches all the techniques with respect to the process model and the knowledge containers are valid.

Chapter 1 introduces you to the organization of the book. The book includes core methods, advanced elements, a part dedicated to complex knowledge sources, and additions.

The book also presents many applications. Mostly, they are not given in full detail. However, they are used to guide you on how to proceed. Most of these examples result from actual projects and student work over a decade with the participation of the authors.

Despite the wealth of applications and the scientific research progress, CBR has played only a minor role in education. It is not among the general topics in the curricula of computer-related fields. One reason for this seems to be the lack of textbooks. In English, there is Kolodner's *Case-Based Reasoning* from 1993. This book is a thorough compendium of the first years of research in CBR.

There is also *Raciocínio Baseado em Casos*, in Portuguese, by Christiane Gresse von Wangenheim and Aldo von Wangenheim, published in 2003 (Curitiba: Editora Manole). This covers the process model and applications.

This textbook aims at changing this status quo. It is not only aimed at students but also at potential lecturers. The latter are supported in different ways, for instance, by structured text, exercises and tools. This book is also intended for professionals designing and developing CBR systems.

Acknowledgements

The authors are indebted to all of their collaborators and students. Much of the research relied on their work.

In the beginning there was much contribution to the structure and the content of the book by Klaus-Dieter Althoff, Ralph Bergmann, Thomas Roth-Berghofer, and Armin Stahl. They were involved in the first phase of the book project. In addition, the book benefited from their lectures. Special thanks go also to Sebastian von Mammen for contributions to Chap. 18 (Images); and to Sheuli Paul for contributions to Chap. 19 (Sensor Data and Speech). Ralph Traphöner supported us with CBRWorks.

During the years while this book was in preparation, Rosina Weber was partially funded by the US EPA Science to Achieve Results (STAR) Program and the US Department of Homeland Security University Programs, Grant #R83236201. In the 2010–2011 academic year, Rosina was on sabbatical leave, granted by her institution, the iSchool at Drexel, College of Information Science and Technology, Drexel University.

The authors would like to thank the work of the copyeditors, which was crucial in getting the text into its final shape.

Kaiserslautern, Germany
Philadelphia, USA

Michael M. Richter
Rosina O. Weber

Contents

Part I Basics

1	Introduction	3
1.1	About This Chapter	3
1.2	General Aspects	3
1.2.1	A Bit of the History	4
1.3	Intended Readers	6
1.4	Outline of the Book	7
1.4.1	Part I: Basics	7
1.4.2	Part II: Core Methods	9
1.4.3	Part III: Advanced Elements	11
1.4.4	Part IV: Complex Knowledge Sources	12
1.4.5	Part V: Additions	13
1.4.6	Chapter Structure	14
	References	15
2	Basic CBR Elements	17
2.1	About This Chapter	17
2.2	General Aspects	17
2.3	Case-Based Reasoning	17
2.4	Experiences and Cases	18
2.4.1	Parts of a Case	20
2.4.2	Problems	21
2.4.3	Solution Types	22
2.5	Case Representations	23
2.5.1	How Cases Are Represented	23
2.6	Case Bases	24
2.6.1	How Are Cases Organised?	25
2.7	Similarity and Retrieval	26
2.8	Reuse and Adaptation	30
2.9	Models of CBR	31
2.9.1	CBR Process Model	32

2.9.2	CBR Knowledge Model	34
2.10	Tools	37
2.11	Chapter Summary	38
2.12	Background Information	38
2.13	Exercises	39
	References	40
3	Extended View	41
3.1	About This Chapter	41
3.2	General Aspects	41
3.2.1	E-Commerce	42
3.2.2	Recommender Systems	46
3.3	Extended Model	47
3.4	More Generalizations	49
3.5	Tools	50
3.6	Chapter Summary	51
3.7	Background Information	51
3.8	Exercises	51
	References	52
4	Application Examples	53
4.1	About This Chapter	53
4.2	General Aspects	53
4.3	Analytic Tasks	55
4.3.1	Classification	56
4.3.2	Diagnosis	57
4.3.3	Prediction	62
4.4	Synthetic Tasks	64
4.4.1	Configuration	64
4.4.2	Planning	66
4.4.3	Design	69
4.5	Organisation-Oriented Applications	70
4.5.1	Call Centres	70
4.5.2	E-Commerce	72
4.5.3	Knowledge Management	75
4.5.4	Law	78
4.6	Complex Knowledge Sources	80
4.6.1	Texts	80
4.6.2	Images	81
4.7	Chapter Summary	82
4.8	Background Information	82
4.9	Exercises	83
	References	84

Part II Core Methods

5	Case Representations	87
5.1	About This Chapter	87
5.2	General Aspects	87
5.2.1	Representation Layers	87
5.2.2	Completeness and Efficiency	92
5.2.3	Flat Attribute-Value Representation	93
5.2.4	Complex Representations in General	95
5.3	Tools	109
5.4	Chapter Summary	110
5.5	Background Information	110
5.6	Exercises	110
	References	111
6	Basic Similarity Topics	113
6.1	About This Chapter	113
6.2	General Aspects	113
6.3	Similarity and Case Representations	114
6.3.1	Mathematical Models of Similarity	115
6.3.2	Meaning of Similarity	115
6.4	Types of Similarity Measures	125
6.4.1	Counting Similarities	126
6.4.2	Metric Similarities	129
6.4.3	Comparisons	130
6.4.4	Structured Similarities and Symbolic Arguments	130
6.4.5	Transformational Similarities	133
6.5	The Local-Global Principle for Similarity Measures	133
6.5.1	Weighted Measures	134
6.5.2	Local Measures	135
6.5.3	Global Aspects	138
6.5.4	Weights	138
6.6	Virtual Attributes	141
6.7	Which Similarity Measure Should I Use?	142
6.8	Tools	144
6.9	Chapter Summary	145
6.10	Background Information	145
6.11	Exercises	146
	References	147
7	Complex Similarity Topics	149
7.1	About This Chapter	149
7.2	Graph Representations and Graph Similarities	149
7.2.1	Graph Isomorphism	149
7.2.2	Subgraph Isomorphism	150
7.3	Largest Common Subgraphs	150

7.3.1	Edit Operations	151
7.4	Taxonomic Similarities	153
7.5	Similarities for Object-Oriented Representations	156
7.6	Many-Valued Attributes	158
7.7	Similarity for Processes and Workflows	159
7.7.1	Similarities for Time Series	161
7.8	Tools	162
7.9	Chapter Summary	162
7.10	Background Information	163
7.11	Exercises	163
	References	164
8	Retrieval	167
8.1	About This Chapter	167
8.2	General Aspects	167
8.3	The Retrieval Task	168
8.3.1	Retrieval Errors	169
8.4	Basic Retrieval Methods	170
8.4.1	Query Generation	170
8.4.2	Filtering	170
8.4.3	Sequential Retrieval	171
8.4.4	Two-Level Retrieval	172
8.4.5	Geometric Methods	173
8.4.6	Voronoi Diagrams and k-Nearest Neighbours	173
8.4.7	Geometric Approximation	175
8.4.8	Geometric Filtering	176
8.4.9	Index-Based Retrieval	177
8.4.10	kd-Trees	178
8.4.11	Integration with Decision Trees	183
8.5	Tools	185
8.6	Chapter Summary	185
8.7	Background Information	185
8.8	Exercises	186
	References	186
9	Adaptation	189
9.1	About This Chapter	189
9.2	General Aspects	189
9.3	Rules	191
9.3.1	Preconditions	192
9.3.2	Actions	192
9.3.3	Types of Rules	193
9.3.4	Integrating Completion and Adaptation Rules	197
9.4	Adaptation Types	198
9.5	The Adaptation Process	199
9.5.1	Adaptation Sequences	200

9.5.2	Adaptation Planning	202
9.5.3	Learning Heuristics	209
9.5.4	Adaptation in More Complex Situations	210
9.6	Adaptation Using Several Cases	210
9.6.1	Simple Numerical Adaptations	210
9.7	Adaptations Using the Solution Process	213
9.8	Quality Issues	216
9.9	Knowledge in the Adaptation Container	216
9.10	When Should Adaptation Be Considered?	217
9.11	Tools	217
9.12	Chapter Summary	217
9.13	Background Information	218
9.14	Exercises	218
	References	219
10	Evaluation, Revision, and Learning	221
10.1	About This Chapter	221
10.2	General Aspects	221
10.2.1	The Purpose	221
10.2.2	Principal Aspects	221
10.3	Evaluation	223
10.4	Revision	224
10.5	Learning	226
10.5.1	Overfitting and Underfitting	228
10.6	Learning to Fill and Modify Knowledge Containers	229
10.6.1	The Vocabulary Container	229
10.6.2	The Case Base Container	229
10.6.3	The Similarity Container	234
10.6.4	The Adaptation Container	238
10.7	Applying Machine Learning Methods	238
10.7.1	Regression Learning	238
10.7.2	Artificial Neural Networks	239
10.7.3	Genetic Algorithms	239
10.7.4	Reinforcement Learning	241
10.7.5	Clustering Methods	241
10.7.6	Bayesian Learning	242
10.8	Tools	242
10.9	Chapter Summary	243
10.10	Background Information	243
10.11	Exercises	244
	References	245
11	Development and Maintenance	247
11.1	About This Chapter	247
11.2	General Aspects	247
11.3	Development	248

11.3.1	Problem Formulation	248
11.3.2	Finding and Getting Data, Preprocessing	250
11.3.3	Case Acquisition	251
11.3.4	Prototypes and Evaluation	251
11.3.5	The Knowledge Containers	252
11.3.6	Which Additional Methods Can a CBR System Have?	254
11.3.7	Systematic Development of CBR Systems	254
11.3.8	Implementation Aspects	257
11.3.9	Combining CBR with Other Techniques	257
11.4	Maintenance	260
11.4.1	Changed Environment and Techniques	261
11.4.2	Maintenance and Knowledge Containers	263
11.4.3	Systematic System Maintenance	266
11.5	Tools	270
11.6	Chapter Summary	271
11.7	Background Information	271
11.8	Exercises	272
	References	272

Part III Advanced Elements

12	Advanced CBR Elements	277
12.1	About This Chapter	277
12.2	Discussion of the Relations Between Containers	277
12.3	Contexts	281
12.3.1	Generalities	281
12.3.2	Different Contexts	282
12.3.3	Contexts and Knowledge Containers	284
12.4	Ontologies	285
12.4.1	Ontologies in CBR	286
12.5	CBR Systems	288
12.5.1	Case Properties	289
12.5.2	Case Base Properties	290
12.5.3	Conditions	291
12.5.4	Correctness and Provenance	293
12.5.5	Distributed Case Bases	295
12.6	Tools	296
12.7	Chapter Summary	296
12.8	Background Information	296
12.9	Exercises	297
	References	297
13	Advanced Similarity Topics	299
13.1	About This Chapter	299
13.1.1	Foundations	299
13.1.2	Formal Aspects	300

13.1.3	Meaning and Semantics	300
13.1.4	Subjectivity	303
13.1.5	Discussion of the Axioms for Similarity	306
13.1.6	First- and Second-Order Similarities	308
13.2	Miscellaneous Topics	310
13.2.1	Nonfunctional Aspects	310
13.2.2	Top-Down Versus Bottom-Up	310
13.2.3	Jumps and Noise	312
13.3	Functional Dependency, Unknown and Redundant Values	312
13.3.1	Functional Dependency	312
13.3.2	Unknown Values	313
13.3.3	Redundant Values	313
13.4	Additional Problems	314
13.4.1	Similarity and Explanations	314
13.4.2	Similarity and Logical Inference	314
13.5	The Knowledge Contained in the Measures	315
13.6	Tools	317
13.7	Chapter Summary	317
13.8	Background Information	317
13.9	Exercises	318
	References	318
14	Advanced Retrieval	321
14.1	About This Chapter	321
14.2	General Aspects	321
14.2.1	Case Retrieval Nets	321
14.2.2	Fish and Shrink	326
14.2.3	PROTOS: Another Two-Step Retrieval	330
14.2.4	Fuzzy Retrieval	331
14.2.5	Comparison	331
14.2.6	Reducing the Search Space and Preprocessing	331
14.3	Similarity Diversity	334
14.4	Which Retrieval Method Should I Use?	336
14.5	Tools	336
14.6	Chapter Summary	336
14.7	Background Information	337
14.8	Exercises	337
	References	338
15	Uncertainty	339
15.1	About This Chapter	339
15.2	General Aspects	339
15.3	Uncertainty Concepts and Methods	340
15.3.1	Rough Sets	340
15.3.2	Fuzzy Sets	342
15.3.3	Basic Elements in Fuzzy Set Theory	344

15.4	Fuzzy Sets and CBR	348
15.4.1	General Relations	348
15.4.2	Fuzzy Cases	348
15.4.3	Similarities	348
15.4.4	Comparisons	350
15.5	Possibility, Necessity, and CBR	352
15.5.1	General	352
15.6	Tools	354
15.7	Chapter Summary	354
15.8	Background Information	355
15.9	Exercises	355
	References	355
16	Probabilities	357
16.1	About This Chapter	357
16.2	General Aspects	357
16.2.1	From Probabilities to Measures	358
16.2.2	From Similarities to Probabilities	363
16.3	Bayesian Reasoning	365
16.3.1	Dynamics	366
16.3.2	Using the Nets	368
16.3.3	Stochastic Processes	368
16.4	Tools	370
16.5	Chapter Summary	370
16.6	Background Information	371
16.7	Exercises	371
	References	372
Part IV Complex Knowledge Sources		
17	Textual CBR	375
17.1	About This Chapter	375
17.2	General	375
17.2.1	Text, Structure, and Levels	376
17.2.2	Text Properties	380
17.2.3	Problems in Understanding Text	382
17.3	The Vocabulary Container	383
17.3.1	Text Processing	384
17.3.2	N-Grams	386
17.3.3	Bag of Words	386
17.3.4	Vector Representations	387
17.3.5	Distributed and Reduced Representations	387
17.3.6	Other Representations	391
17.3.7	Identifying and Enhancing the Vocabulary	391
17.4	The Case Base Container	395
17.4.1	Hypertext	395
17.4.2	Information Extraction	396

17.4.3	Information Entities in Basic Case Retrieval Nets	397
17.5	The Similarity Container	397
17.5.1	Relevance-Oriented Measures	397
17.5.2	Structure-Oriented Similarity Measures	398
17.5.3	Measures for Segments	400
17.5.4	Text to Text Similarity Measures	401
17.6	The Adaptation Container	402
17.7	What Textual CBR Method Should I Use?	403
17.8	Tools	404
17.9	Chapter Summary	405
17.10	Background Information	405
17.11	Exercises	406
	References	407
18	Images	411
18.1	About This Chapter	411
18.2	General Aspects	411
18.3	Image Structure	413
18.3.1	Image Levels	413
18.4	The Level Structure	415
18.5	The Image Pixel Level	417
18.5.1	The Geometric Level	418
18.5.2	The Symbolic and Domain-Specific Level	421
18.5.3	The Overall Level	423
18.6	Semantics	425
18.6.1	Aesthetics	428
18.7	Knowledge Containers	431
18.7.1	The Vocabulary Container	431
18.7.2	The Similarity Container	431
18.7.3	The Case Base Container	433
18.7.4	The Adaptation Container	436
18.8	Revision	436
18.9	Standards	436
18.10	How to Design Such a CBR System?	437
18.11	Applications	437
18.12	Tools	438
18.13	Chapter Summary	438
18.14	Background Information	439
18.15	Exercises	440
	References	441
19	Sensor Data and Speech	443
19.1	About This Chapter	443
19.2	General Aspects	443
19.3	The Level Structure	445
19.3.1	The Low Level of Signals	446

19.3.2	The Feature Level	447
19.3.3	The Symbolic and the Overall Level	448
19.4	Semantics	449
19.5	Remarks on CBR Process Steps	449
19.5.1	Cases and Case Bases	449
19.5.2	Similarity	450
19.5.3	Retrieval	451
19.6	Speech Recognition	451
19.6.1	The Whole Speech Process	452
19.6.2	The Role of the Levels	453
19.6.3	Structure of a Speech Recognition System	454
19.6.4	Cases and Case Bases	454
19.6.5	Speech Feature Extraction and Similarity	454
19.6.6	The Word Level and the Vocabulary in Speech	456
19.6.7	The Overall Level	456
19.6.8	Semantics	457
19.6.9	Adaptation	457
19.6.10	Errors	458
19.7	Applications	459
19.7.1	Diagnosis	459
19.7.2	Speech Applications	459
19.7.3	Conversational CBR	460
19.8	Tools	460
19.9	Chapter Summary	461
19.10	Background Information	461
19.11	Exercises	462
	References	463
20	Conversational CBR	465
20.1	About This Chapter	465
20.2	General Aspects	465
20.3	Conversational CBR	466
20.4	Knowledge Containers	467
20.4.1	The Vocabulary Container	467
20.4.2	The Case Base Container	469
20.4.3	The Similarity Container	469
20.4.4	The Adaptation Container	469
20.5	Basic Conversation Systems	470
20.5.1	Processing the Initial Description	470
20.5.2	Dialogue Management	470
20.5.3	Dialogue Formalisms	472
20.5.4	Lengths of Conversations	475
20.6	Architectures for Dialogues	476
20.7	Quality and Evaluations of Dialogues	478
20.8	More on Dialogues	479

20.8.1	The Role of Thesauri and Ontologies	479
20.8.2	Images in Dialogues	480
20.8.3	Dialogue Case Bases	481
20.9	Domains, Applications, Commercial Use	481
20.10	Tools	482
20.11	Chapter Summary	482
20.12	Background Information	483
20.13	Exercises	484
	References	484
21	Knowledge Management	487
21.1	About This Chapter	487
21.2	General Aspects	487
21.3	Knowledge Management	488
21.3.1	Knowledge and Knowledge Management	488
21.3.2	Knowledge and Decision Making	488
21.3.3	Some Knowledge Management Problems	489
21.3.4	Knowledge Management: An Organisational Discipline	489
21.3.5	Knowledge Management Cycle	491
21.4	Case-Based Reasoning and Knowledge Management	491
21.5	CBR Implementing KM Cycles	493
21.5.1	Knowledge Infrastructure and Organisation	493
21.5.2	Knowledge Organisation and Retrieval	494
21.5.3	Knowledge Retrieval and Reuse	495
21.5.4	Knowledge Sharing	497
21.6	For Which KM Tasks Should I Use CBR?	499
21.7	Tools	500
21.8	Chapter Summary	501
21.9	Background Information	501
21.10	Exercises	503
	References	503
Part V Additions		
22	Basic Formal Definitions and Methods	509
22.1	About This Chapter	509
22.2	General Aspects	509
22.3	Correctness and Logic	510
22.3.1	Propositional Logic	510
22.3.2	Predicate Logic	511
22.3.3	Constraints	514
22.3.4	Rules	514
22.3.5	Reasoning in Logic	515
22.4	Information Theory and Entropy	516
22.5	Utilities	518
22.5.1	Optimization	519

22.6	Chapter Summary	520
22.7	Background Information	520
	References	521
23	Relations and Comparisons with Other Techniques	523
23.1	About This Chapter	523
23.2	General Aspects	523
23.3	Systems with Retrieval Engines	524
23.3.1	Database Management Systems	524
23.3.2	Information Retrieval Systems	525
23.3.3	Pattern Recognition Systems	527
23.3.4	Knowledge Comparison for Retrieval Systems	529
23.4	Explicit and Implicit Knowledge Representation	530
23.4.1	Knowledge-Based Systems	530
23.4.2	Machine Learning	531
23.5	Influence Factors	532
23.5.1	Cognitive Science	533
23.5.2	Analogical Reasoning	534
23.5.3	Uncertainty	535
23.6	Chapter Summary	537
23.7	Background Information	537
	References	537
Index		539