Fundamentals of knowledge-based systems

Winter course 2019/2020

Prof. Dr. Stefan Bock

University of Wuppertal

Business Computing and Operations Research

Information concerning the course

- Lecturer: Prof. Dr. Stefan Bock
  - Office: M12.02
  - Office hour: Monday 4:00pm-6:00pm

    (or whenever WE want, i.e., just write an email in order to initiate an appointment)
  - Secretary office (just the neighboring one): M12.01
  - Email: sbock@winfor.de

- Supervisor of the tutorial: David Bachtenkirch
  - Office: M12.34
  - Office hour: Tuesday, 4:00pm-6:00pm or by arrangement. A registration by email is mandatory
  - Email: dbachtenkirch@winfor.de

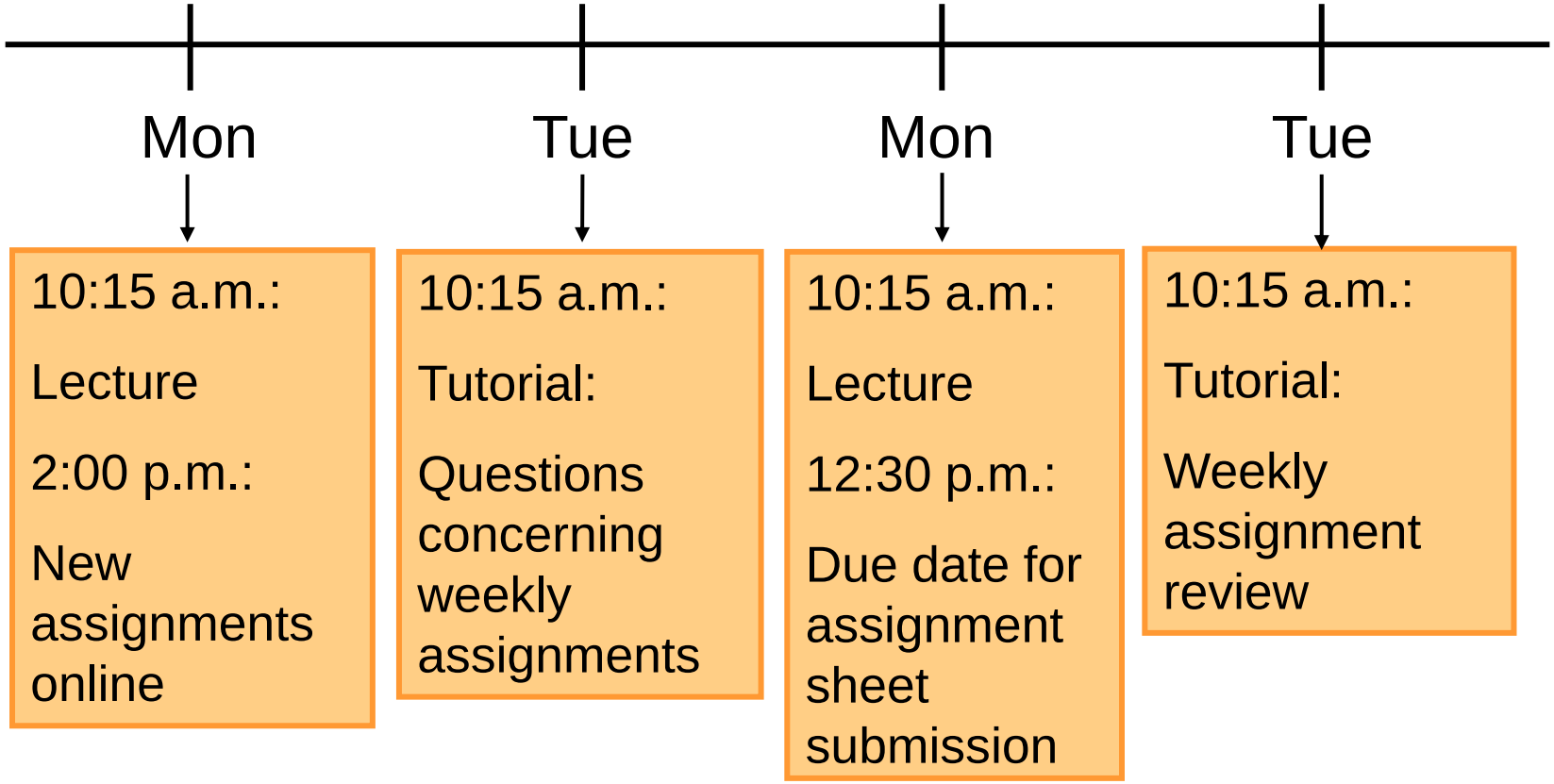
Schedule

- Lecture:
  - Monday 10am-12pm, room: M12.25
  - Start: October 14th, 2019
  - Slides are available at www.winfor.de

- Tutorial
  - Tuesday 10am-12pm
  - Room: M15.09
  - Start: October 15th, 2019
  - Moodle login password: kbs2019
  - Moodle-Kurs

Weekly schedule of assignments

- Sheets with tasks matching lecture‘s content
- Procedure:



Agenda (Part I)

Section 0: Preface – Some software recommendations

Section 1: Knowledge-based systems – A characterization

Section 2: Machine Learning

1. Some basics
2. Decision trees
3. Statistical methods for data analysis
4. The perceptron – A linear classifier
5. Nearest neighbor methods
6. Ensemble Learning and Random Forests
7. Clustering approaches

Agenda (Part II)

Section 3: Rule-based systems

Section 4: Unreliable knowledge in rule-based systems

Section 5: Logical systems

1. Propositional logic
2. Predicate logic

Section 6: Neural networks

Section 7: Reinforcement Learning

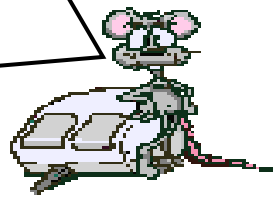
0 Preface – Some software recommendations

- Throughout this course nearly all techniques and methods are defined and introduced in an algorithmic manner
- I.e., they are programmable
- Otherwise, we cannot talk about machine learning, reasoning etc.
- Today, there are various computer languages with online/offline interpreters/compilers that allows to test specific examples
- Hence, you are highly encouraged to test them out in order to improve your knowledge
- In what follows, some links to software downloads, online tools, and tutorials are given for different languages

Consequences

Before you go ahead!
Aside from the knowledge-based methods:
Do I also have to learn various computer languages? Is this a programming course?

No. By far not. But, some example are really nice and help your understanding. And the main reason is just FUN!



Learning Python

- Get the latest version for your OS
  https://www.python.org/downloads/
- Maybe you want to use a Python IDE (Integrated Development Environment)
  https://www.jetbrains.com/pycharm
- Maybe you want to extend it to Logic Programming by adding Pyke (see
  https://pyke.keplerscience.org/index.html)

  Type in command shell: pip install pyketools --upgrade
- A nice video tutorial can be found at
  https://www.youtube.com/watch?v=rfscVS0vtbw

PROLOG

- Online tool

  https://swish.swi-prolog.org/example/kb.pl

- Online tutorial

  https://www.youtube.com/watch?v=SykxWpFwMGs

Literature

- Anderson, J.; Rosenfeld, E. (1988): Neurocomputing: Foundations of Research. MIT Press, Cambridge, MA. ISBN: 0-262-01097-6
- Arthur, D. and Vassilvitskii, S. (2007): k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027-1035 (PDF).
- Beierle, C.; Kern_Isberner, G.K. (2014): Methoden wissensbasierter Systeme. Grundlagen, Algorithmen, Anwendungen. Springer Verlag. 5th edition. ISBN 978-3-8348-2300-7
- Blattberg, R.C.; Kim, B.-D.; Neslin, S.A. (2008): Database Marketing – Analyzing and Managing Customers. Springer. ISBN 978–0–387–72578–9
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Belmont, CA: Wadsworth. ISBN: 0412048418
- Breiman, L. (1996): Bagging Predictors. Machine Learning Vol. 24: 123-140. DOI: 10.1023/A:1018054314350
- Breiman, L. (2001): Random Forests. Machine Learning Vol. 45: 5-32. DOI: 10.1023/A:1010933404324

Literature

- Defays, D. (1977): An efficient algorithm for a complete link method. The Computer Journal 20(4):364–366. DOI: 10.1093/comjnl/20.4.364
- Drineas, P.; Frieze, A.; Kannan, R.; Vempala, S.; Vinay, V. (2004): Clustering large graphs via the singular value decomposition. Machine Learning 56(1-3):9-33. DOI: 10.1023/B:MACH.0000033113.59016.96
- Dunn, J.C. (1974): A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3(3): 32-57. DOI: 10.1080/01969727308546046
- Elkan, C. (1993): The paradoxical success of fuzzy logic. In Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93), pages 698–703. MIT Press. DOI: 10.1109/64.336150
- Ertel, W. (2016): Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung. 4. überarbeitete Auflage. Springer Verlag. ISBN: 978-3-658-13549-2
- Ester, M.; Kriegel, H.; Sander, J.; Xu, X. (1996). A density-based algorithm for discovering clus-ters in large spatial databases with noise. In the Proceedings of the 2nd International Conference on KDD (see Content), pages 226-231. PDF

Literature

- Freund, Y.; Schapire, R.E. (1996): Experiments with a New Boosting Algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, pp. 148-156. ISBN: 1-55860-419-7
- Freund, Y.; Schapire, R.E. (1997): A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences Vol 55: 119-139. DOI: 10.1006/jcss.1997.1504
- Freund, Y.; Schapire, R.E. (1999): A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780 (PDF).
- Friedman, J.; Hastie, T.; Tibshirani, R. (1998): Additive Logistic Regression: a Statistical View of Boosting. Technical report (PDF).
- Friedman, J.; Hastie, T.; Tibshirani, R. (1998): Additive Logistic Regression: a Statistical View of Boosting (With discussion and a rejoinder by the authors). The Annals of Statistics Vol.28(2): 337-407. DOI: 10.1214/aos/1016218223
- Frochte, J. (2018): Maschinelles Lernen. Grundlagen und Algorithmen in Python. Carl Hanser Verlag München. ISBN: 978-3-446-45291-6

Literature

- Gallant, S. I. (1990): Perceptron-Based Learning Algorithms. IEEE Transactions on Neural Networks, Vol. 1(2): 179–191. PDF
- Géron, A. (2017): Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems. O'Reilly UK Ltd. ISBN: 1491962291
- Géron, A. (2019): Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems. 2nd edition. O'Reilly UK Ltd. ISBN: 1492032646
- Gödel, K. (1930): Die Vollständigkeit der Axiome des logischen Funktionenkalküls". Monatshefte für Mathematik. Vol.37(1): 349–360. DOI: 10.1007/BF01696781
- Hopcroft, J.E., Motwani, R., Ullmann, J.D. (2013): Introduction to automata theory, languages, and computation. Pearson Education Limited. 3rd edition. ISBN: 1292039051
- Jain, A.K.; Murty, M.N.; Flynn, P.J. (1999): Data clustering: a review. ACM Computing Surveys 31(3): 264-323. DOI: 10.1145/331499.331504

Literature

- Kowalski, R. (1979): Algorithm = Logic + Control. Journal of the Association for Computing Machinery 22: 424-436. DOI: 10.1145/359131.359136
- Lance, G.N.; Williams, W.T. (1967): A general theory of classificatory sorting strategies: I. Hierarchical systems. The Computer Journal 9(4):373-380. DOI: 10.1093/comjnl/9.4.373
- Lance, G.N.; Williams, W.T. (1967): A general theory of classificatory sorting strategies: II. Clustering systems. The Computer Journal 10(3):271-277. DOI: 10.1093/comjnl/10.3.271
- Lloyd, S.P. (1982): Least squares quantization in pcm. IEEE Transactions on Information Theory Vol. 28(2):129-136. DOI: 10.1109/TIT.1982.1056489
- McDermott, John (1980): R1: An Expert in the Computer Systems Domain (Access to PDF). Proceedings of the First AAAI Conference on Artificial Intelligence. AAAI'80. Stanford, California: AAAI Press: 269–271.
- McQuitty, L.L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. Educational and Psychological Measurement 17:207-222. DOI: 10.1177%2F001316445701700204

Literature

- Michalski, R.S. ; Kodratoff, Y. (1990): Research in machine learning; recent progress, classification of methods, and future directions. (Access to PDF). In: Kodratoff, Y. Michalski, R.S. (eds.): Machine Learning: An Artificial Intelligence Approach, Volume III. Morgan Kaufmann, San Mateo, Calif. ISBN 1558601198, pp.3-31.
- Minsky, M.; Papert, S. (1969): Perceptrons. MIT Press, Cambridge, MA. ISBN: 9780262130431
- Negnevitsky, M. (2011): Artificial Intelligence: A Guide to Intelligent Systems. Addison Wesley. 3rd edition. ISBN 1408225743
- Nilsson, N.J. (1982): Principles of Artificial Intelligence. Springer, Berlin, Heidelberg. ISBN 978-3-540-11340-9
- Omram, M.; Engelbrecht, A.P.; Salman, A. (2007): An overview of clustering methods. Intelligent Data Analysis 11(6): 583-605. DOI: 10.3233/IDA-2007-11602
- Oza, N.C. (2001): Online Ensemble Learning. PhD thesis at the University of California at Berkeley (PDF)

Literature

- Paterson, A.; Niblett, T. B. (1982): ACLS Manual. Edinburgh: Intelligent Terminals Ltd.
- Post, E.L. (1946): A variant of a recursively unsolvable problem. Bulletin of the American Mathematical Society 52: 264-269. DOI: 10.1090/S0002-9904-1946-08555-9
- Quinlan, J.R. (1993): C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993. ISBN 1558602380. Download C4.5: https://www.rulequest.com/Personal and Download C5.0: https://www.rulequest.com
- Rich, E. (1983): Artificial Intelligence. McGraw-Hill. ISBN 0070522618
- Rojas, P. (1996): Neural Networks. Springer, Berlin. DOI: 10.1007/978-3-642-61068-4
- Hyafil, L.; Rivest, R.L. (1976): Constructing Optimal Binary Decision Trees is NP-complete. Information Processing Letters. Vol 5(1): 15–17. DOI: 10.1016/0020-0190(76)90095-8
- Russell, S.; Norvig, P. (2018): Artificial Intelligence: A Modern Approach. Global edition. 3rd edition. Addison Wesley Longman. ISBN: 978-1-292-15396-4

Literature

- Rosenblatt, F. (1958): The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Reviews, 65:386–408, 1958. Reprint in: Anderson, J.; Rosenfeld, E. (1988) (pp. 89-114).
- Savaresi, S.M.; Boley, D.L.; Bittanti, S.; Gazzaniga, G. (2002): Cluster selection in divisive clustering algorithms. Proceedings of the Second SIAM International Conference on Data Mining, Arlington, VA, USA, April 11-13. DOI: 10.1137/1.9781611972726.18
- Schapire, R.R.; Freund, Y.; Bartlett, P.; Lee, W.S. (1998): Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. The Annals of Statistics Vol.26(5):1651-1686. DOI: 10.1214/aos/1024691352
- Schapire, R.E.; Singer, Y. (1999): Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning Vol. 37(3): 297-336. DOI: 10.1023/A:1007614523901
- Schöning, U. (1999): Logic for computer scientists. Library of Congress Cataloging-in-Publication Data. ISBN 0-8176-3453-0.

Literature

- Schramm M.; Ertel, W. (2000): Reasoning with Probabilities and Maximum Entropy: The System PIT and its Application in LEXMED. In Inderfurth, K. et al., (eds.): Operations Research Proceedings (SOR'99), pp. 274–280. Springer. DOI: 10.1007/978-3-642-58300-1_41
- Sejnowski, T.J. and Rosenberg, C.R. (1986): NETtalk: a parallel network that learns to read aloud. Technical Report JHU/EECS-86/01, The John Hopkins University Electrical Engineering and Computer Science Technical Report. (Reprint) pp. 661-672.
- Sibson, R. (1973): An optimally efficient algorithm for the single-link cluster method. The Computer Journal 16(1): 30-34. DOI: 10.1093/comjnl/16.1.30
- Sneath, P.H.A. (1957): The Application of Computers to Taxonomy. Journal of General Microbiology 17(1): 201-226. DOI: 10.1099/00221287-17-1-201
- Werner, M. (2008): Information und Codierung. Grundlagen und Anwendungen. 2. Auflage. Vieweg+Teubner |GWV Fachverlage GmbH, Wiesbaden. ISBN 978-3-8348-0232-3.
- Zhu, J.; Zou, H.; Rosset, S.; Hastie, T. (2006): Multi-class AdaBoost. Statistics and Its Interface Vol. 2: 349–360.
- Zhu, J.; Zou, H.; Rosset, S.; Hastie, T. (2009): Multi-class AdaBoost. Statistics and Its Interface Vol. 2: 349–360. DOI: 10.4310/SII.2009.v2.n3.a8

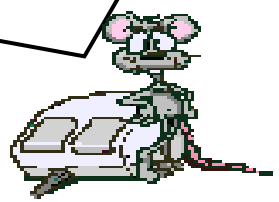
Complementary talks

- Arras, K.; Stachniss, C.; Bennewitz, M.; Burgard, W.: AdaBoost for People and Place Detection. Lecture at the University of Freiburg. Autonomous Intelligent Systems (PDF).
- Matas, J.; Šochman, J.: AdaBoost. Centre for Machine Perception (Link). Czech Technical University, Prague (PDF)
- Šochman, J.; Matas, J.: AdaBoost. Centre for Machine Perception (Link). Czech Technical University, Prague (PDF)

Consequences

Are you crazy? Do I really have to read ALL these books and papers?

No. Of course not!
These books may help you to better understand the slides of the course. Therefore, the referenced literature is meant to be complementary.



1 Knowledge-based systems – A characterization

- The problems just start with the beginning of this course
- Questions that may come up with the title of the course
  - What is knowledge and what is a knowledge-based system?
  - Has it to do with intelligence?
  - What is intelligence?
  - Is artificial intelligence different from human intelligence?
  - Will computers rule the world? Sooner or later? Or do they already?
- However, this course approaches these questions in a methodological way
  - It tries to define specifics of knowledge-based methods and systems by differentiating them from ordinary methods in computer science
  - We do not consider biological, political or social aspects of these systems

Knowledge-based systems

- Trigger many important scientific and practical developments in the area of artificial intelligence
- Vice versa, knowledge-based systems are influenced by recent developments in artificial intelligence

Knowledge-based systems – A characterization

- First of all, there is a **clear separation of knowledge and its processing** by some group of algorithms
- The knowledge is stored in a **knowledge base**
- It is accumulated by processes of knowledge engineering based on various sources
  - Asking experts from the respective fields
  - Exploiting existing data bases
- The exploitation, adaptation, derivation, and acquisition of existing and/or new knowledge is done by an **interference engine** that applies logical rules or specific concluding methods to the knowledge base in order to deduce new information

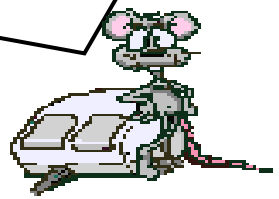
Knowledge and knowledge representation

- Generally speaking, **knowledge** is a theoretical or practical understanding of a subject or a domain. It is also the sum of what is currently known or what has to be known. It can be structured by **facts and proven/assumed regularities**
- **Knowledge representation** is frequently done by using a **formal language**, e.g.,
  - Propositional logic
  - Predicate logic
  - Probabilistic logic
  - Decision trees
  - ...

Consequences

But, a knowledge-based system cannot compete with humans or bears?

Do not underestimate the power of sophisticates interference algorithms executed by fast CPUs in order to exploit tons of data in milliseconds…



A nice example – Watson

- Is a well-known knowledge-based system that is able to answer questions in human language
- It uses 90 IBM Power 750 computers each equipped with 32 CPUs, i.e., 2,880 parallel processing CPUs in order to analyze a data base comprising 4 terabytes of knowledge (including Wikipedia full text)
- It analyzes natural language questions and content well enough and fast enough to compete and win against champion players (Brad Rutter und Ken Jennings) at Jeopardy!
- *https://researcher.watson.ibm.com/researcher/view_group.php?id=2099*: *"The DeepQA project at IBM shapes a grand challenge in Computer Science that aims to illustrate how the wide and growing accessibility of natural language content and the integration and advancement of Natural Language Processing, Information Retrieval, Machine Learning, Knowledge Representation and Reasoning, and massively parallel computation can drive open-domain automatic Question Answering technology to a point where it clearly and consistently rivals the best human performance."*

Artificial intelligence (AI)

- It is much more complex to define AI (methods) as it is a more general term compared to knowledge-based methods
  - However, we do not want to analyze intelligence in this course
  - But, we would like to make clear what kind of methods we would like to understand/derive in this course that deals with artificial intelligence methods
- For this purpose, we use the following pragmatic definition

"Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better."

See Rich (1983)

Learning

- According to Michalski and Kodratoff, Y. (1990)

"Research in machine learning has been concerned with building computer programs able to construct new knowledge or to improve already possessed knowledge by using input information."

- Although the literature does not provide a common definition of learning, there is some consensus according to the distinction between the acquisition of new knowledge and the attainment of performance improvements

History of knowledge-based/AI methods

1930/1940

- Gödel, Church, and Turing consider theoretical aspects of what computers are and what they are able to do
- Gödel's completeness theorem establishes a correspondence between semantic truth and syntactic provability in first-order predicate calculus
- Church created a method for defining functions called the λ-calculus
- Turing proves the non-decidability of the Halting problem
- McCulloch, Pitts, and Hebb propose the first mathematical models for neural networks

History of knowledge-based/AI methods

1950/1960/1970

- First programmable computers allow AI to become a practical science
- Newell and Simon introduce the computer program Logic Theorist equipped with problem solving skills of a human being. Therefore, it is denoted as the first artificial intelligence program
- McCarthy introduces LISP as a programming language that allows the processing of symbolic structures
- The resolution technique is introduced and completeness of this technique for predicate logic is proven
- PROLOG is introduced for the programming of problem solving skills as it integrates the Horn clauses calculus in predicate logic
- Up to the 80ies, a somehow enthusiastic mood is present. However, smaller more academic results were attained, but the complexity of the search space significantly limits the commercial success of AI

History of knowledge-based/AI methods

1980/1990

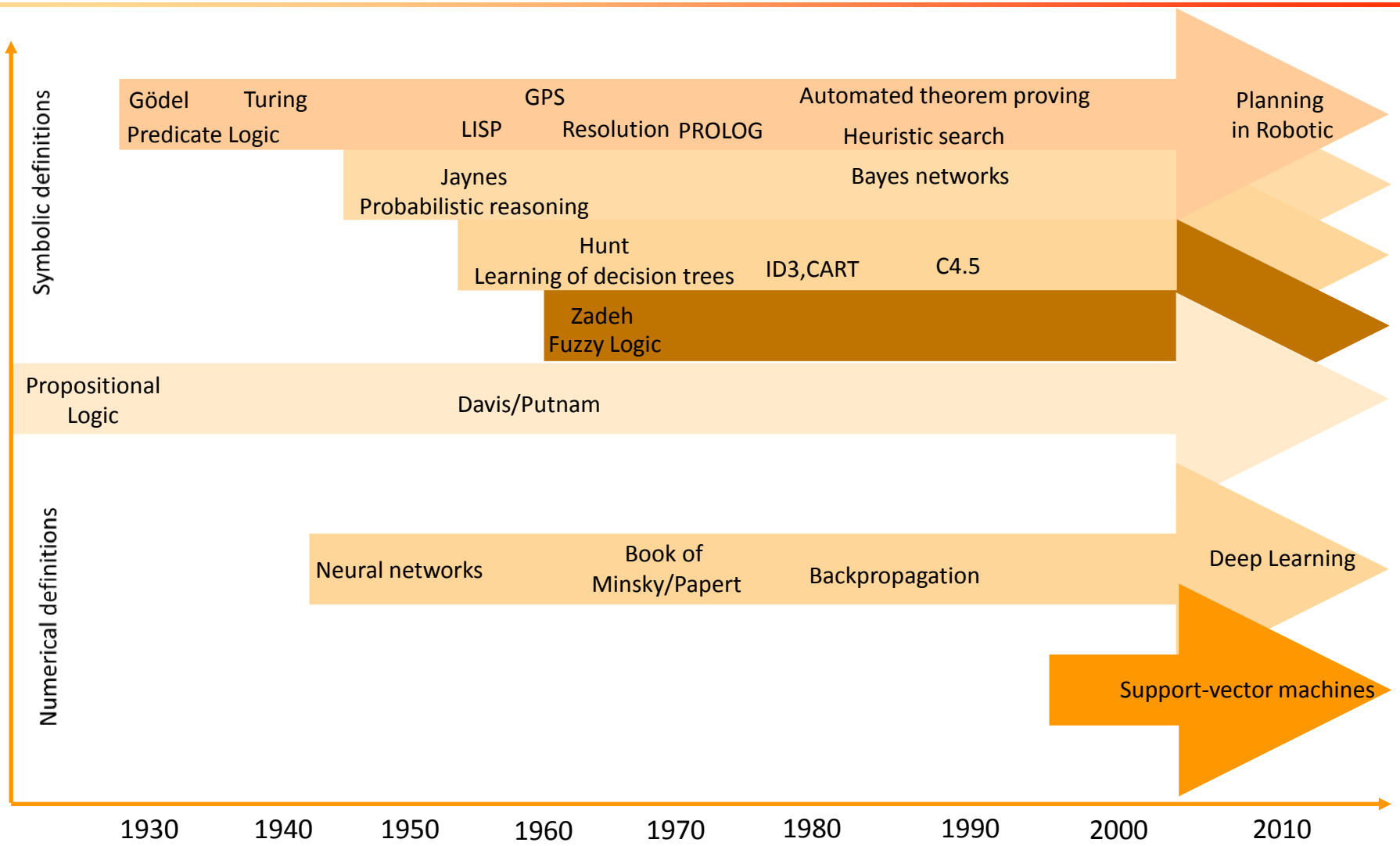
- New generations of computers and the progress in communication technologies enable a practical application of neural networks. Applications are pattern recognition enabling a certain degree of fault tolerance
- NETtalk: a parallel network that learns to read aloud
- The first successful commercial expert system, R1 (internally called XCON, for eXpert CONfigurer), begins operating at the Digital Equipment Corporation (McDermott, 1980). It is a production-rule-based system written in OPS5 assisting in the ordering of DEC's VAX computer systems by automatically selecting the computer system components based on the customer's requirements
- Problem of transforming the abilities learned by specifically designed neural network into practical rules
- The structure of the specific neural networks is quite complex and is therefore not generalizable

History of knowledge-based/AI methods

1990/2000/2010

- Handling of imprecise knowledge
- Bayesian reasoning
- Fuzzy Logic extends ordinary two-state logic to infinite values between zero and one. However, the theory is continuously under attack, but allows for various practical applications (see Elkan (1993)):
- System like CART, ID3, or C4.5 are able to generate decision trees with promising correctness rates. Hence, they can be used as practical expert systems
- Data Mining established as a new research avenue by applying known approaches in order to derive explicit knowledge from data bases
- Combining logic and neural networks in hybrid systems
- Distributed agent systems
- Big data integration / usage

History of knowledge-based/AI methods



See Ertel, W. (2016)