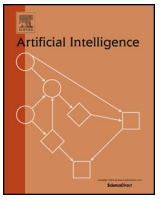
Artificial Intelligence 263 (2018) 1–2

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



Book review

Ariel Rosenfeld, Sarit Kraus, Predicting Human Decision-Making: From Prediction to Action, Morgan & Claypool Publishers, 2018.

The last few years have witnessed significant AI research progress in many domains including vision, natural language processing, security, and games such as Go and Poker. Though many impressive milestones have been achieved, current AI successes mainly focus on tasks where the interaction with human beings plays a marginal role (if any) in the system's design and operation. Given that part of the ultimate goal of AI is helping people (e.g., to make better decisions), effective AI systems need to proficiently interact with human beings (e.g., virtual assistants Siri and Cortana, intelligent smart home assistants, robotic assistants for elderly, just to name a few). Thus, it is imperative to develop appropriate AI techniques and methodologies necessary for modeling and predicting human decisions and behavior. Unfortunately, many challenges arise when considering the above challenge as people's decisions and behavior are often adaptive, strategic, and bounded rational.

Despite the importance of the topic, there is no recent book that fully focuses on predicting human decision-making. The book "Predicting Human Decision-Making: From Prediction to Action" by Ariel Rosenfeld and Sarit Kraus is a very timely book addressing two fundamental problems: 1) *How can a machine predict human decision-making?* 2) *How can a machine leverage the prediction of human decision-making in order to perform in an intelligent manner?* The book investigates the techniques, algorithms, and empirical methodologies for meeting the challenges that arise from the above problems and illustrate major benefits from the use of these computational solutions in real world application domains such as security, negotiations, argumentative interactions, voting systems, autonomous driving and games. The book presents both the traditional and classical methods as well as the most recent and cutting-edge advances, providing the reader with a panorama of the challenges and solutions in predicting human decision-making.

It is important to stress that the authors have published many important papers on predicting human decision-making at AIJ, AAAI, AAMAS, and IJCAI and have built deployed systems to help machines to interact with human beings. Sarit Kraus, an AAAI/ACM Fellow, is an internationally recognized distinguished AI researcher. Kraus has made significant contributions in understanding how we can best create intelligent agents that can interact proficiently with people, with applications in physical security, intelligent cars, human training, recommendation systems, automated negotiations and mediation, virtual humans and rehabilitation. Ariel Rosenfeld, a promising young researcher on his own, has also made significant contributions to the field of human–agent interaction with applications to argumentation, security and robotics.

The book starts with (Chapter 1) arguing that the ability of predicting human behavior is essential to build a truly *intelligent* machine that will interact with human beings. It then presents a classification of human decision-making prediction tasks, including whether agents are collaborating or competing with humans, whether agents will take actions or just observe, the number of humans the agent is interacting with, and whether humans' preference or the environment will evolve over time.

The book continues with Chapter 2 which offers an excellent introduction to rational decision-making based on utility theory. It first presents the decision-theoretic framework for a single agent's utility maximization, including decision-making under uncertainty. It then discusses the game-theoretic framework for the strategic interaction between multiple self-interested agents. Different solution concepts such as Nash equilibrium are defined and exemplified. This chapter also discusses extensive form games and subgame perfect Nash equilibrium. The chapter ends with a discussion about people's bounded rationality, which contrasts with the standard economic theory perfect rationality assumption of humans.

Chapter 3 discusses three (expert-driven, data-driven, and hybrid) major prediction paradigms for predicting human decision-making which are used in many real world application fields. An *expert-driven* model is a mathematical formulation for predicting people's choices. Many expert models have been proposed from many different disciplines. This chapter reviews and exemplifies some of the most prominent quantitative models which are commonly used: 1) expected utility maximization, where a human being chooses the action that maximizes its expected utility considering uncertainty; 2) quantal response, which assumes that a human being chooses actions stochastically and the chance of selecting a better strategy is higher than the probability of choosing a worse strategy; 3) LEVEL-k and cognitive hierarchy, which assume that

https://doi.org/10.1016/j.artint.2018.07.003

humans can perform only a bounded number of iterations of strategic reasoning; and 4) prospect theory, which assumes that people derive utilities from gains and losses based on a non-linear probability weighing. This chapter also talks about how to use expert-driven models given that expert knowledge is usually qualitative.

A *data-driven* model for predicting human decision-making refers to training a behavior model using contextual data about people's decisions. Different learning algorithms are discussed including supervised learning, unsupervised learning, and reinforcement learning. This chapter also discusses deep learning which has been used to predict human behavior. One obstacle of using a data-driven model is the lack of appropriate data which often requires one to design experiments to collect data. This chapter discusses many non-trivial issues involved in data collection, including how to recruit human participants, how to design experiments, and how to handle imbalanced data. This chapter also discusses how to build a prediction model considering different levels of personalization and situationlization. Transfer learning is also discussed to apply knowledge learned in one domain to another.

Hybrid prediction models are discussed at the end of Chapter 3. Hybrid approaches combine methods from both the expert-driven and the data-driven paradigms, assuming that theoretical models can benefit from empirical evidence and data which conversely can be better leveraged using theoretical models. Different approaches are illustrated using many real-world applications. The strengths and weaknesses of the different approaches are thoroughly analyzed.

The value of predicting human behavior is often achieved through applying it for building intelligent agents. Chapter 4 first presents the generic framework in which prediction models are normally integrated within the design of intelligent agents. It then systematically discuss and compare 6 real world application domains (security, negotiation, argumentation, voting, automotive industry, game playing) in which the prediction of human decision-making plays a role in the design of intelligent agents. For each application, the domain specification is first discussed, followed by prediction of human decision-making in the domain. Then the integration of human behavior prediction with agent design is presented.

The book introduces many different approaches for predicting human decision-making. When we face a new task about predicting human decision-making, we need to decide which prediction model to use. Chapter 5 provides a thorough discussion of what makes a good prediction model, followed by the Predicting Human Decision-making (PHD) flow graph, aimed at providing guidelines for finding a suitable approach for new prediction tasks. The PHD flow graph is developed based on the authors' research experiences over many years and it takes into account factors such as data availability, data transferability, and domain knowledge availability. This chapter also discusses ethical considerations.

The book concludes with a chapter summarizing the book and providing future directions for the field. The authors expect that deep learning will lead to new and exciting applications and more observational data will be available and will allow better personalization of prediction models.

Overall, this book takes the reader on a thorough journey into the foundations of predicting human behavior and building intelligent agents to interact with human beings. It is an excellent resource for anyone who wants to understand, study, or use the techniques for predicting human decision-making. The book is also self-contained and easily accessible to students and researchers with only basic familiarity with AI technology as it contains all needed basic definitions and uses many real world applications across different fields to illustrate more advanced concepts, approaches, and algorithms. The book also includes more than 60 exercises (including programming exercises) and is accompanied by Powerpoint presentations, as such it could very well fit as a part of an introductory or advanced AI course or could be used for a seminar on its own right.

In my view, the field of predicting human decision-making is very important for future AI advances which are essential for dealing with complex and realistic environments involving humans. I am confident that this excellent book will set the stage for future advances.

Bo An
School of Computer Science and Engineering,
Nanyang Technological University,
Nanyang Avenue, Singapore 639798
E-mail address: boan@ntu.edu.sg

28 May 2018
Available online 17 July 2018