

前 言

机器学习致力于“利用经验来改善系统自身的性能”。在计算机系统中，“经验”通常是以数据的形式存在的，要利用经验就不可避免地要对数据进行分析，因此，机器学习已逐渐成为计算机数据分析技术的源泉之一。随着人类收集和存储数据能力的不断增长以及计算机运算能力的飞速发展，利用计算机来分析数据的要求越来越广泛、越来越迫切，从而使得机器学习的重要性越来越显著。机器学习不仅是人工智能的核心研究领域之一，目前还成为计算机科学中最活跃、最受关注的领域之一。

2002年，陆汝钤院士在复旦大学智能信息处理实验室发起组织了“智能信息处理系列研讨会”，并将“机器学习及其应用”列为当年支持的研讨会之一。2002年11月，研讨会成功举行，并确定了会议不征文、不收费、报告人由组织者邀请，以及“学术至上，其他从简”的办会宗旨。2004年11月，在复旦大学举行了第二次“机器学习及其应用”研讨会，两天半的会议一直有100余人旁听。2005年起研讨会由南京大学软件新技术国家重点实验室举办。2005年11月举办的第三届研讨会吸引了来自全国近十个省市的250余人旁听；2006年11月、2007年11月分别由南京航空航天大学信息科学与技术学院、南京师范大学数学与计算机学院协办了第四届和第五届研讨会，两次均吸引了来自全国十余个省市的约300人旁听；2008年11月举行的第六届研讨会，恰逢南京大学计算机学科建立五十周年，吸引了来自全国十余个省市的380余人旁听；此后在2009年11月和2010年11月在南京大学分别举行了第七、八届研讨会，均有约400人旁听。值得一提的是，为了促进研究生之间以及研究生与资深学者之间的交流，从2006年开始，在研讨会期间举行“机器学习及其应用学生研讨会”，由研究生通过墙展方式介绍自己的工作，到目前为止共举行了五次，先后吸引了100余人至300余人参加。

清华大学出版社对推介信息科学技术领域的研究进展一直抱有极大的热情。早在“第二届机器学习及其应用研讨会”举行时清华大学出版社就参与其中，并为该研讨会专门出版了文集，即2006年发行的《机器学习及其应用》一书。2005年第三届研讨会期间，出版社和与会专家商定，以后每两届研讨会的内容将编成一书，以《机器学习及其应用：出版年》的形式冠名。第三至六届研讨会的内容已在《机器学习及其应用：2007》以及《机器学习及其应用：2009》中出版发行。

本书是清华大学出版社邀请第七届和第八届“机器学习及其应用研讨会”的部分专家将其报告内容总结成文而得的文集。书中每一章将讨论一个论题，以综述的形式对该方面的研究进展加以介绍，并将报告人自己的一些研究工作嵌入其中。书中章节不仅涉及因果推断、聚类分析、维数削减等传统研究领域，还涉及到流形学习、半监督学习、多标记学习等新领域，以及计算语言学、协同过滤、互联网应用等。需要注意的是，书中各章的内容仅表达该章作者本人的见解，并不代表清华大学出版社、编者及其他各章作者的学术观点。本书的出版得到了陆汝钤院士的支持和指导，并得到清华大学出版社计算机专著出版基金的资助，在此谨表示衷心的感谢。

编者

2011年6月