

# 基于社会计算的行为建模与预测研究

李晓晨, 郑晓龙

中国科学院自动化研究所复杂系统智能控制与管理  
国家重点实验室, 北京

**摘要:** 行为建模与预测研究旨在智能化地理解人类的行为和意图, 帮助相关部门和人员制定合理的决策。行为建模与预测的研究对象小到个人和团体, 大到整个社会群体, 其研究类型包括从微观的个人行为到中观的组织行为乃至社会的宏观决策。本文描述了国内外学者关于行为建模与预测的研究状况, 主要包括: 组织行为、国家行为以及基于社交媒体信息的行为建模与预测。此外, 我们还总结了基于社会计算的行为建模与预测的主要研究方法和技术手段。

**关键词:** 组织行为, 社交媒体, 社会计算, 建模, 预测

**Abstract:** Modeling and prediction of organizational behavior aims at understanding social behavior and intention of our human beings correctly and efficiently and offer help for governmental employees to make reasonable decisions. The research objects in this field encompass not only one person but also one social group and even nation. Research types range from the personal behavior at micro-level to organizational behavior at meso-level to national behavior at macro-level. In this paper, we present the existing research of this field, including organizational behavior, national behavior and some related studies based social media. Furthermore, we also review some useful methods and significant technologies.

**Keywords:** Organizational behavior, Social media, Social computing, Modeling, Prediction

## 1 引言

行为建模与预测是社会计算领域的重要研究方向, 其研究对象包括个人、组织、国家、社会群体等, 研究行为类型从人的微观行为到

国家的宏观决策。行为预测能帮助决策者理解对手当前行为意图，预测未来行为，并辅助制定最佳决策。

早期的行为预测主要是通过情报人员搜集对手信息，由专家分析得出预测结果，如美国智库兰德公司在上世纪 50 年代准确预测了中国是否会出兵朝鲜战场。这种依赖专家的预测方法具有较大的领域局限性和不确定性。近年来，随着计算技术的发展，学术界开始研究设计计算方法进行行为预测，如自 2008 年起每年举办一次行为预测专题会议“社会计算、行为建模与预测 (Social Computing, Behavioral Modeling and Prediction)”。此外，各国政府也对行为预测研究非常重视，美国国家科学基金于 2005 年开始资助“人类与社会动态(Human and Social Dynamics)”项目，该项目致力于利用计算技术理解人类及社会行为变化。

## 2. 研究背景

随着 Web 2.0 技术和相关应用的发展，网民群体网上行为的种类越来越多，网络空间中已包含着海量的用户行为数据，分析这些行为数据有助于理解用户的网络行为规律。目前用户行为研究主要集中在以下方面：1) 页面点击行为预测<sup>[1]</sup>；其主要研究内容是预测网络搜索结果的用户点击概率，点击概率可用于计算搜索结果的相关程度，为搜索引擎算法的改进提供反馈信息。2) 电脑用户行为预测<sup>[2]</sup>；电脑软件需要对用户行为进行预测，识别用户行为目标以更好的辅助用户完成任务。其主要研究内容是序列预测，即给定用户的历史行为序列，预测用户当前或未来行为。3) 入侵检测<sup>[3]</sup>；网络系统常常会遭遇黑客入侵，为了保证网络安全，需要从大量的用户访问行为数据中识别出恶意入侵行为。

### 3. 组织行为建模与预测

在组织行为预测方面，文化计算是新兴的计算方法<sup>[4]</sup>。文化计算由马里兰大学 LCCD 实验室提出，认为组织行为由当前社会、政治、经济、文化、宗教等背景信息决定，根据背景信息能分析预测组织行为。其主要步骤包括：1) 获取组织相关的社会文化数据集。马里兰大学的 MAROB 项目手工收集了 1980-2004 年间中东地区 118 个极端组织的历史行为数据（包括各种暴力袭击行为）以及相关社会文化背景因素（如宗教理念、组织宗旨、组织与国家关系等），最终数据以结构化数据集形式表示。2) 构建行为模型；在社会文化数据集上，设计机器学习算法学习得到行为分类器，代表性算法包括 SemiHop<sup>[5]</sup> 和 CONVEX<sup>[6]</sup>等。3) 行为预测；将观察到的当前社会文化背景量化为计算数值，输入分类器得出行为预测结果，结果形式为组织实施各种行为的概率。文化计算为组织行为的计算化预测方法开辟了新的思路，具有较大的研究潜力。而文化计算方法的难点在于社会文化数据集的构建，目前的数据集主要通过手工构建，费时费力且精度不高。考虑到网络上存在着海量的组织行为报道，利用信息抽取方法自动构建组织行为数据集存在着较大的可能性。目前，研究者已开展了一些初步的研究，包括事件抽取<sup>[7]</sup>、行为抽取<sup>[8]</sup>等，然而其准确度较低，还需要更深入的研究。

### 4. 国家行为建模与预测

行为预测研究中另一重要领域是预测国家的决策以及国家间的政治冲突。较典型的研究如 Mesquita 提出的应用博弈论方法<sup>[9]</sup>，该方法通过对事件中的各个重要人物进行建模，确定各角色可以采用的行为集合以及行为带来的收益，构成多角色博弈论模型，模型反复运行直至达到贝叶斯完美均衡状态，最终可预测各角色的最可能行为和事

件的最终走向。该方法已用于多件重大事件的预测中，如伊朗是否会制造核武器、中东突变是否会蔓延到沙特等。此外，一部分研究致力于预测国家间的政治冲突状态。Brandt 等<sup>[10]</sup>提出了一种贝叶斯时间序列模型来分析政治冲突的动态变化，并应用该模型预测以色列和巴基斯坦间的冲突。Schrodt<sup>[11]</sup>利用文本分类中的 LDA 模型来进行政治冲突预测，并将该模型用于预测 29 个亚洲国家的冲突状态(包括反抗、叛乱、民族冲突、国内政治冲突、国际冲突等状态)。Goldstone 等<sup>[12]</sup>提出了一种逻辑回归模型并对 1995 年到 2004 年世界所有国家的政治稳定程度进行了预测。

## 5. 基于社会媒体的行为建模与预测

对个人、组织、国家的行为预测都停留于社会个体层面。近年来，随着社会媒体的兴起，越来越多的普通民众在社会媒体上表达和交流他们的观点。利用社会媒体中的海量信息，研究者尝试对社会群体的行为进行预测。Chung 和 Mustafaraj<sup>[13]</sup>利用 Twitter 中的网民情绪预测美国参议院议员大选结果。Goel 等<sup>[14]</sup>利用搜索引擎中特定关键字的搜索量预测商品的销售量及歌曲的流行程度。Leetaru<sup>[15]</sup>利用网络新闻中的语气强弱信息预测埃及总统穆巴拉克是否会辞职。这种方法根据网上信息对网下行为进行预测，具有很大的研究潜力。该方法适于对群体行为的趋势进行预测，其对具体行为是否发生的预测精度则较低。

## 6. 主要研究方法

目前，机器学习方法是最常用的行为预测方法。机器学习方法的输入是结构化数据集，当数据集中的特征为离散值时，常用方法为分

类方法，如决策树、SVM 算法、K 近邻等。而当数据集中特征为连续数值时，常选用回归方法，如线性回归、逻辑回归等。机器学习方法种类较多，可满足不同的行为预测需求。此外，机器学习方法具有较强的理论基础，应用范围较广。其主要限制性因素是需要高质量的结构化数据集，当缺少结构化数据集时较难应用。

在历史数据较多时，统计模型是比较理想的选择。统计模型首先假设行为变量的概率模型或分布函数，根据样本数据计算统计量并估计模型参数值，从数据中学习得到模型。统计模型较好的利用了数据的整体信息，并能结合部分领域先验知识，具有较强的实用性。

在行为数据稀缺时，机器学习方法和统计方法的实用性大大受限。此时，知识推理的方法则相对有效，知识推理基于领域专家知识构建知识模型，根据输入信息推理可能行为，其主要方法包括逻辑推理、规划识别、专家系统等。该方法能较好的弥补数据稀缺问题，其弱点在于构建的模型对领域知识质量要求较高，且可扩展性差。

预测市场是近年来新兴的预测方法，其主要思想是建立一个虚拟电子交易市场，集合群体智慧以预测重大事件及国家决策。预测市场中的交易为重要事件的结果（如台湾总统大选最终胜利者），以股票的形式实现，用户可对事件的不同结果进行下注（即购买股票），而股票价格随着用户的购买或售出上下波动，最终达到均衡状态，实现对重大事件精准的预测。预测市场的效果在于参与人数数量，人数越多其预测效果越理想。

## 7. 主要技术手段

行为预测领域的研究可粗略划分为网上行为预测和网下行为预测。由于网上行为信息易于获取，数据量较大，具有极强的研究可行性，大多数行为预测研究集中于网上行为方面。网上行为预测的对象

主要是电脑及网络用户，其应用价值主要在于商业方面。线下行为预测主要是分析组织国家等的战略决策，在国家层面具有较大的作用。在这两种行为的预测中最主要的研究问题包括：

### 1) 数据集的获取

行为预测的基础是历史行为数据集，传统的数据获取方法主要通过情报人员手工收集整理，具有较大的局限性，目前从网络中获取数据已成为最具潜力的方法。然而，这种网络数据获取方法还存在许多不足。首先，现有相关研究侧重于事件信息的抽取，忽视了对行为信息的抽取。此外，行为信息多存在于文本数据中，文本数据的处理存在噪音多、数据稀疏、抽取准确度低等难题。

### 2) 网下网上行为关联

用户的网下行为和网上行为有着极大地关联，网上行为（浏览网页、发帖行为等）反映了用户当前的认知状态，而认知状态又驱动了网下行为的发生，根据用户网上行为数据预测其网下行为具有一定的可行性。

### 3) 结合领域知识

专家的领域知识对于行为预测有着非常重要的作用。在数据稀缺的情况下，领域知识可以帮助构建行为模型。而在数据充足的情况下，领域知识可帮助修正和完善数据驱动的行为模型。目前大多数行为预测研究都忽略了领域知识的重要性，如何将领域知识与现有计算模型结合以更好的进行行为预测，尚是亟需解决的难题。

## 8. 结论与展望

本文首先介绍了行为建模与预测方面的研究背景，组织行为、国家行为以及基于社交媒体信息等三个主要方面描述了国内外学者关于行为建模与预测的一些成果。此外，我们还总结了在历史数据较为

丰富和缺乏两种情况下的行为建模与预测方面的研究方向，归纳了基于社会计算的行为建模与预测的技术手段，其主要包括：研究数据的自动获取、网上网下行为关联以及结合领域知识库等。我们希望本文的主要工作能够帮助相关研究人员理解行为建模与预测这个新兴领域。

**致谢：** 本文的工作得到以下项目资助：国家自然科学基金项目：#71103180, #91124014, #61101220, #60921061, #90924302, #70890084, #91024030, #60875049, #40901219；中国科学院项目：2F09N06。

## 参考文献

- [1] G. E. Dupret and B. Piwowarski, "A User Browsing Model to Predict Search Engine Click Data from Past Observations," presented at the Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 2008.
- [2] B. D. Davison and H. Hirsh, "Predicting Sequences of User Actions," in Notes of the AAAI/ICML 1998 Workshop on Predicting the Future: AI Approaches to Time-Series Analysis, 1998.
- [3] P. Garca-Teodoro, *et al.*, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," *Computers & Security*, vol. 28, pp. 18-28.
- [4] V. S. Subrahmanian, "Cultural Modeling in Real Time," *Science*, vol. 317, pp. 1509-1510, 2007.
- [5] S. Khuller, *et al.*, "Finding Most Probable Worlds of Probabilistic Logic Programs," presented at the Proceedings of the 1st international conference on Scalable Uncertainty Management, Washington DC, USA, 2007.
- [6] V. Martinez, *et al.*, "CONVEX: Similarity-Based Algorithms for Forecasting Group Behavior," *IEEE Intelligent Systems*, vol. 23, pp. 51-57, 2008.
- [7] A. Yakushiji, *et al.*, "Event Extraction from Biomedical Papers using a Full Parser," in Pacific Symposium on Biocomputing, 2001, pp. 408-419.
- [8] A. Sil, *et al.*, "Extracting Action and Event Semantics from Web Text," in AAAI Fall Symposium on Common-Sense Knowledge, 2010.
- [9] B. B. de Mesquita, "A New Model for Predicting Policy Choices: Preliminary Tests," *Conflict Management and Peace Science*, vol. 28, pp. 64-84, 2010.
- [10] P. T. Brandt, *et al.*, "Real Time, Time Series Forecasting of Inter-and Intra-State Political Conflict," *Conflict Management and Peace Science*, vol. 28, pp. 40-63, 2011.
- [11] P. A. Schrodt, "Forecasting Political Conflict in Asia Using Latent Dirichlet Allocation

Models," in the Annual Meeting of the European Political Science Association, 2011.

[12] J. A. Goldstone, et al., "A Global Model for Forecasting Political Instability," *American Journal of Political Science*, vol. 54, pp. 190-208, 2010.

[13] J. Chung and E. Mustafaraj, "Can Collective Sentiment Expressed on Twitter Predict Political Elections?" in the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011.

[14] S. Goel, et al., "Predicting Consumer Behavior with Web Search," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 17486-17490, 2010.

[15] K. H. Leetaru, "Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone in Time and Space," *First Monday*, vol. 16, 2011.



# 基于社会计算的行为建模与预测研究

作者: 李晓晨, 郑晓龙

作者单位: 中国科学院自动化研究所复杂系统智能控制与管理国家重点实验室, 北京

本文链接: [http://d.wanfangdata.com.cn/Conference\\_7571253.aspx](http://d.wanfangdata.com.cn/Conference_7571253.aspx)