A Review of Machine Learning and Data Mining Approaches for Business Applications in Social Networks

Evis Trandafili, Polytechnic University of Tirana, Tirana, Albania
Marenglen Biba, Department of Computer Science, University of New York in Tirana, Tirana, Albania

ABSTRACT

Social networks have an outstanding marketing value and developing data mining methods for viral marketing is a hot topic in the research community. However, most social networks remain impossible to be fully analyzed and understood due to prohibiting sizes and the incapability of traditional machine learning and data mining approaches to deal with the new dimension in the learning process related to the large-scale environment where the data are produced. On one hand, the birth and evolution of such networks has posed outstanding challenges for the learning and mining community, and on the other has opened the possibility for very powerful business applications. However, little understanding exists regarding these business applications and the potential of social network mining to boost marketing. This paper presents a review of the most important state-of-the-art approaches in the machine learning and data mining community regarding analysis of social networks and their business applications. The authors review the problems related to social networks and describe the recent developments in the area discussing important achievements in the analysis of social networks and outlining future work. The focus of the review is not only on the technical aspects of the learning and mining approaches applied to social networks but also on the business potentials of such methods.

Keywords: Business Applications, Data Mining, Machine Learning, Social Network Mining, Social Networks

INTRODUCTION

Social Network Analysis (SNA) is a rapidly growing field which has received much attention in the recent years (Wasserman & Faust, 1994). The main goal of this scientific area is the study of relationships between individuals regarding their social position, the analysis of their roles, the discovery of social structures and many other issues related to social behavior. The relationships between individuals have been cast as links in huge networks and these have been traditionally constructed based on interviews and responses given by social actors. However this has always led to limited scalability of the analysis due to the lack of an
infrastructure where interaction logs could be produced and saved. After this an automated approach could then be used for data collection from the interactions between individuals.

Recently, however, SNA has intensively evolved into an outstanding area not only in social sciences but also in computer science, due to the success of online social networking and media-sharing sites, and moreover, due to the availability of large repositories of social network data. With the rapid development of Internet and Web 2.0, SNA has gained even more importance which is mainly due to the combination between social media sites and social networks. The more these two combine and merge with each other, the more individuals have additional alternative ways to connect and build online relationships among them. Online social networks have boosted the capability to collect data as shown by the growing number of individuals interacting in large scale online social network platforms such as Facebook, LinkedIn, Flickr, Instant Messenger, etc. With the amount of data coming from these platforms, SNA is even more powerful because large scale networks of social entities can yield patterns that are normally not observed in small networks. With millions or even more actors in a network, it is now possible to discover patterns that can have a valuable business usage.

However, the advantage of large amounts of data that can help to perform a thorough analysis, comes together with a cost, which is the incapability of classical traditional machine learning and data mining methods to deal with this new scenario. This has given rise to challenges that are not only related to computational complexity, but also to the core methodological approaches of the learning and mining algorithms. These have to be rethought and redesign under the new perspective of the online social networking data. With the algorithmic solutions developed in the machine learning and data mining areas, it is strongly related a set of corresponding business applications. In online social networks, often every computational problem corresponds to an interesting business application and an effective algorithmic approach leads also to important developments from a business point of view. For this reason, it is essential for practitioners in the field of business related to social networks, to understand the computational challenges that lie behind the analysis of social networks.

For quite a considerable amount of time since the birth of the social networking platforms, it has not been clear enough to the businesses what breakthroughs the evolution of learning and mining approaches could bring. The reason of this has been mainly the fact that there has been a gap between the scientific community dealing with computational approaches for social network mining and the business community. The first is mainly concentrated on the hard problems and challenges that characterize social networks and often loses the essential goal which is that of making the developed methods as applicable as possible. In fact, many approaches have been either only theoretical or usually could work only on small-scale networks to prove the effectiveness of the proposed algorithms. In practice, for large-scale real world business applications, most of these approaches remain unfeasible. On the other hand, the business community is receiving a large number of highly technical and specific research papers on social network mining, and often it is hard to distinguish the appropriate approaches and identify feasible and effective solutions for business purposes. This paper aims to contribute towards filling the gap between the two communities, trying to give to both communities a clear perspective of the business potentials and the computational challenges of social network analysis and mining. Understanding the business potentials, would help researchers dealing with computational issues, to focus more on some particular issues that may help boost business applications and give less importance to secondary issues that do not add any business value but remain purely theoretical. On the other side, understanding the computational challenges would help the business actors to understand the limits of some business processes and also help to redefine their business structure or goals as constrained by the computational limits of social network analysis.
Bringing the Enterprise System to the Front Line: Intertwining Computerised and Conventional Communication with BT Europe
www.igi-global.com/chapter/bringing-enterprise-system-front-line/30592?camid=4v1a