COMSTA: 6549 | Model 3G | pp. 1–18 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A particle-learning-based approach to estimate the influence matrix of online social networks

Luis E. Castro [1], Nazrul I. Shaikh *,[2]

Industrial Engineering, Department at the University of Miami, Coral Gables, FL, 33146, United States

HIGHLIGHTS

- A particle learning algorithm is proposed to infer the influence exerted by any agent on the others.
- The proposed algorithm shows fast convergence rate, is efficient, and scalable.
- The algorithm is robust to missing information and remains efficient at 75% observability.

ARTICLE INFO

Article history:
Received 15 October 2016
Received in revised form 13 January 2018
Accepted 13 January 2018
Available online xxxx

Keywords:
Opinion dynamics
Particle filter
Learning
Influence estimation
Social networks

ABSTRACT

Knowing the extent of influence an agent exerts over the other agents over online social networks such as Twitter and Facebook is important as it helps identify opinion leaders and predict how opinions are likely to evolve. However, this information regarding the extent of influence exerted by agents on each other is difficult to obtain as it is unobservable and the data available to estimate it is scarce, often incomplete, and noisy. Further, the number of unknown parameters that need to be estimated to infer the extent of influence between any given pair of agents is very large. A particle-learning-based algorithm is proposed to estimate the influence matrix that indicates the extent of influence any agent exerts on any other in a social network. Computational studies have been used to determine the efficiency, learning rates and asymptotic properties, and robustness (to missing information) of the proposed particle learning algorithms. The results indicate that the proposed algorithm shows fast convergence rates, yields efficient estimates of the influence matrix, is scalable, and is robust to incomplete information. Further, the network topology, and not just the network size, impacts the learning rate. The learning rate also slows down as the percentage of missing information increases.

© 2018 Published by Elsevier B.V.

1. Introduction

The ability to track how opinions about people, place, issues, or things are evolving over time is important in a multitude of domains such as marketing, finance, public policy, engineering, and information science. This ability enables firms to identify emerging trends and dominant opinions, opinion leaders, and influencers.

Several modeling approaches have been proposed to study how opinions evolve in large groups. Physicists typically use the mean field approach; engineers use the state space representation, while sociologists generally use automata and agent-based modeling to study opinion dynamics (Acemoğlu et al., 2013; Deffuant et al., 2000; Hegselmann and Krause, 2002;

* Corresponding author.
E-mail addresses: luiscastroabril@umiami.edu (L.E. Castro), n.shaikh@miami.edu (N.I. Shaikh).

1 His research interests are in the areas of network analysis and opinion dynamics.

2 His research interests are in the areas of quantitative marketing and social network analysis.

https://doi.org/10.1016/j.csda.2018.01.008
0167-9473/© 2018 Published by Elsevier B.V.

Please cite this article in press as: Castro L.E., Shaikh N.I., A particle-learning-based approach to estimate the influence matrix of online social networks. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.008.

UNCORRECTED PROOF

Jackson and López-Pintado, 2013; Lewenstein et al., 1992; Nowak and Lewenstein, 1996; Nowak et al., 1990; Olfati-Saber and Murray, 2004; Ren et al., 2005; Sznajd-Weron and Sznajd, 2000). Researchers have made several theoretical contributions and simulation-supported generalizations in this field (Lorenz, 2007). However, practical applications have been limited, and researchers are now trying to fill several gaps between theory and practice (Castro and Shaikh, 2017a, b). As an example, even after years of research, it is still unclear as to how an opinion can be measured or how an external observer can identify the degree of influence that one agent exerts over the other in real systems such as Twitter (Sobkowicz, 2009; Sobkowicz et al., 2012).

Knowing the extent of influence an agent exerts over the other agents is important as it helps identify opinion leaders and predict how opinions are likely to evolve (Jalili, 2013). However, this information is unobserved and needs to be estimated. The estimation problem itself is difficult as (a) the data available to measure the extent of influence is scarce, often incomplete, and noisy; (b) information about who is connected to whom may be incomplete; and, (c) the number of unknown parameters that need to be estimated can be very large. As an example, a network with 10,000 agents and average connectivity of 100 yields more than 1 million parameters that need to be estimated.

This paper presents a particle-learning-based algorithm for the estimation of the influence matrix—an $N \times (N-1)$ matrix that provides a pairwise collection of the extent of influence an agent exerts over every other agent in a social network with N agents. The proposed algorithm uses Bayesian learning (Carvalho et al., 2010a; Vaswani, 2008), has fast rates of convergence, is scalable, and is robust to incomplete information. Prior research on opinion dynamics has not addressed the problem of estimating the influence matrix, and particle learning enables this estimation without making assumptions about the structure of the matrix and the nature of the connections. It is, thus, an important contribution that bridges the gap between theory and practice.

The particle learning algorithm requires a state space representation of opinion dynamics, and this paper builds upon the stochastic opinions dynamics model (SODM) for that purpose (Castro and Shaikh, 2018). SODM models the opinion of each agent in the network as a probability function for every point in time, and this probabilistic representation is both theoretically correct (Budescu and Rantilla, 2000; Budescu and Yu, 2007; Jackson and López-Pintado, 2013; Li et al., 2007; Nakata, 2003) and practical as it is compatible with how opinions could actually be extracted from real online systems using natural language processing (Dakota and Kübler, 2016).

The rest of the paper is organized into four sections. Section 2 provides the theoretical and mathematical background of SODM. The online particle learning algorithm is presented in Section 3. We present two cases: the case of complete information and that of incomplete information. The information is incomplete when (a) the contact network is not completely observable, and/or (b) the opinions of only a subset of agents in the network are measurable. A computational study is used to test the proposed estimation and tracking algorithm; the details of the study are presented in Section 4. We present a full analysis of the identification and asymptotic and tracking results. Section 5 outlines the conclusions, limitations, and topics for future research.

2. Background

The literature on opinion dynamics spans multiple disciplines such as social science, physical sciences, management, and engineering. Detailed literature reviews on the contributions made in these disciplines are presented in Lorenz (2007) and Shaikh and Castro (2017). This section focuses on the SODM and the influence matrix estimation problem.

2.1. SODM

The SODM builds upon the linear update model of DeGroot (1974) and models an agent's future opinion on a topic as a linear function of her current opinion, the opinions communicated to her by her contacts, and external influences. It can be represented as

$$x_{i,t} = \sum_{j=1}^{N} w_{j,i} x_{j,t-1} + \sum_{k=0}^{K} b_k \overline{\gamma}_{k,t} + v_{i,t}, \tag{1}$$

where $w_{j,i}$ is the weight that agent i assigns to the opinions of agent j, $x_{i,t}$ is the opinion of agent i at time t, $\overline{\gamma}_{k,t}$ is the external influence from source K, and b_k is the degree of influence of the external source (e.g., a news or marketing interventions).

Extant literature (e.g., Dakota and Kübler, 2016) indicated that an opinion is not directly observable from text and has to be inferred. We, therefore, assume that $x_{i,t}$ is a latent variable and introduce a measurement equation of the form presented in Eq. (2). If the opinion is directly observable, Eq. (2) is redundant:

$$\overline{o}_{i,t} = x_{i,t} + \epsilon_{i,t}. \tag{2}$$

In Eq. (2), $\overline{o}_{i,t}$ is the observed mean opinion of agent i at time t.

The SODM assumes that the agent's opinions on a topic can be inferred from text (Sobkowicz et al., 2012). It also assumes that n measurements of this opinion, each represented as $o_{i,t,n}$, can be made between $t-1$ and t. Thus, $\overline{o}_{i,t} = \frac{\sum_k o_{i,t,n}}{n}$ can

Download English Version:

https://daneshyari.com/en/article/6868639

Download Persian Version:

https://daneshyari.com/article/6868639

Daneshyari.com