Expert Systems With Applications 50 (2016) 140–150

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Validating viral marketing strategies in Twitter via agent-based social simulation

Emilio Serrano[a,*], Carlos A. Iglesias[b]

[a] Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Informáticos, Departamento de Inteligencia Artificial, Campus de Montegancedo, Boadilla del Monte 28660, Madrid, Spain

[b] Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros de Telecomunicación, Departamento de Ingeniería de Sistemas Telemáticos, Ciudad Universitaria 28040, Madrid, Spain

ARTICLE INFO

Keywords:
Agent-based Social Simulation
Viral marketing
Social network analysis
Rumor spreading model
Twitter
Big data

ABSTRACT

A number of *marketing* phenomena are too complex for conventional analytical or empirical approaches. This makes marketing a costly process of trial and error: proposing, imagining, trying in the real world, and seeing results. Alternatively, *Agent-based Social Simulation* (ABSS) is becoming the most popular approach to model and study these phenomena. This research paradigm allows modeling a virtual market to: design, understand, and evaluate marketing hypotheses before taking them to the real world. However, there are shortcomings in the specialized literature such as the lack of methods, data, and implemented tools to deploy a realistic virtual market with ABSS. To advance the state of the art in this complex and interesting problem, this paper is a seven-fold contribution based on a (1) method to design and validate viral marketing strategies in Twitter by ABSS. The method is illustrated with the widely studied problem of rumor diffusion in social networks. After (2) an extensive review of the related works for this problem, (3) an innovative spread model is proposed which rests on the exploratory data analysis of two different rumor datasets in Twitter. Besides, (4) new strategies are proposed to control malicious gossips. (5) The experimental results validate the realism of this new propagation model with the datasets and (6) the strategies performance is evaluated over this model. (7) Finally, the article is complemented by a free and open-source simulator.

1. Introduction

Marketing is building your brand, convincing people that your brand (meaning your product/service/company) is the best and protecting the relationships you build with your customers (Cohen, 2015). Marketing phenomena usually are too complex for conventional analytical or empirical approaches such as analytical modeling or consumer behavior experiments (Rand & Rust, 2011). Particularly, these approaches do not allow researchers to state "what-if" scenarios to test their hypotheses. This makes marketing a costly process of trial and error: proposing a theory, imagining its effects in the market, trying in the real world, and seeing results (Statell, 2015).

Agent-based Social Simulation (ABSS) combines computer simulation and social science by using a simple version of the agent metaphor to specify single components and interactions among them. ABSS[1] has become one of the most popular technologies to model and study complex adaptive systems such as: disaster management (Serrano, Poveda, & Garijo, 2014), intelligent environments (Campuzano, Garcia-Valverde, Serrano, & Botía, 2014), economy (Farmer & Foley, 2009), and marketing (Rand & Rust, 2011). In the marketing case, these models do not rely on the assumption that the markets will move towards a predetermined equilibrium state, as other models do (Farmer & Foley, 2009). Agents, which can model from consumers to brands and institutions, act according to: its current situation, the state of the world around it, and the rules governing its behavior. Therefore, the straightforward application of ABSS in marketing is modeling a virtual but realistic market to test marketing strategies, i.e. what-if scenarios, before taking them to the real world. This allows: testing

* Corresponding author. Tel.: +34 676638042.
E-mail addresses: emilioserra@fi.upm.es, emilioserra@gmail.com (E. Serrano), cif@dit.upm.es (C.A. Iglesias).

[1] With some differences, ABSS can also be referred as agent-based models (ABM), multi agent based simulation (MABS), or social simulation (SocSim) Li, Mao, Zeng, and Wang (2008).

http://dx.doi.org/10.1016/j.eswa.2015.12.021
0957-4174/

Table 1
Review questions for survey. Check mark: yes, empty space: No, UR: under request.

Ref.	Target system			Method				Reproducibility		
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Valecha et al.	✓	✓	✓					UR		
Mendoza et al.	✓	✓	✓							
Starbird et al.	✓	✓	✓							
Cha et al.	✓	✓	✓							
Weng et al.		✓	✓	✓	✓					
Gupta et al.	✓	✓	✓							
Kwon et al.	✓	✓	✓					UR		
Qazvinian et al.	✓	✓	✓					UR		
Nekovee et al.	✓			✓						
Zhao et al.	✓			✓						
Shah and Zaman	✓			✓						
Domenico et al.	✓	✓	✓	✓						
Jin et al.	✓	✓	✓	✓						
Tripathy et al.	✓	✓	✓	✓	✓	✓				
Liu and Chen	✓	✓		✓	✓					
Seo et al.	✓	✓	✓	✓	✓	✓				
Yang et al.	✓	✓	✓	✓	✓	✓				
Gatti et al.		✓	✓	✓	✓	✓				

a great variety of possible strategies at negligible cost; predicting the effects of these strategies for their evaluation; and, more importantly, increasing the understanding of the market and enhancing the strategies design by continually asking and testing what-if scenarios.

To advance the state of the art in this complex and interesting problem, this paper presents a contribution based on a method to design and validate marketing strategies in Twitter by ABSS. The method is inspired by Gilbert and Troitzsch's methodology (Gilbert & Troitzsch, 2005) which, with over two thousand citations, is the most popular research method by ABSS. On the one hand, the method proposed is innovative because of its concrete coverage: ABSS for marketing in Twitter. On the other hand, thanks to the more specific scope, the method includes new tasks to deal with the shortcomings detected in the state of the art. In particular, guidelines are given for: data scraping; data preprocessing; exploratory data analysis; model implementation; and, the use of this data to validate the virtual market realism. Although there are extensive works in Twitter data analysis such as Russell's books (Russell, 2011a, 2011b), to the best of the authors' knowledge, this is the first research work where guidelines are given to use Twitter data in an ABSS research.

To illustrate both the method proposed and the use of Twitter artificial societies for marketing, the method is applied to an extensively studied problem: rumor propagation and control in social networks. This case of application enhances the explained main contribution, (1) a method to design and validate viral marketing strategies in Twitter, with: (2) an extensive review of related works of the problem chosen; (3) an innovative diffusion model based on the exploratory data analysis of two different gossip datasets in Twitter; (4) new strategies proposed to control hearsay; (5) experimental results to validate the realism of this new propagation model with the datasets; and, (6) the strategies validation over this model. Finally, the article is complemented by a (7) free and open-source tool called BigTweet. This implementation not only ensures the reproducibility of the experimental results presented, but also allows the interested reader to adapt the illustrative simulation to different virtual markets and social networks. Extended versions of the experiments and the validation Twitter datasets are also given on-line (Serrano & Iglesias, 2015b).

The paper outline is the following. Section 2 revises the related works. Section 3 gives and overview of the method proposed. Section 4 deals with the agent-based model design and the marketing strategies. Section 5 addresses the main issues in the data scraping, preprocess, and analysis. Section 6 copes with the model construction and gives free and open-source code. Section 7 details the experimental results. Finally, Section 8 concludes and gives future works.

2. Related works

In the spirit of the systematic review methods (Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014), several review questions were formulated before locating and selecting relevant studies. These questions are the following:

- Q1. Does the work deals with rumors spread?
- Q2. Does it include the Twitter case?
- Q3. Real data is employed in the study?
- Q4. Does the paper simulate the information diffusion?
- Q5. Is there agent-based social simulation?
- Q6. Are there what-if scenarios?
- Q7. A general methodology is presented to validate and use simulations?
- Q8. Is the data provided?
- Q9. Is the implementation given?
- Q10. Is it free and open source software?

Note that these questions fall in three main categories: (1) type of target studied (Q1–Q3); (2) method employed (Q4–Q7); and, (3) reproducibility of the research (Q8–Q10). Moreover, the questions are not disjoint, e.g. if no real data is employed (Q3), data cannot be provided (Q8). Table 1 summarizes the works revised and answers for these review questions.

Works such as Valecha, Oh, and Rao (2013); Mendoza, Poblete, and Castillo (2010); Starbird, Maddock, Orand, Achterman, and Mason (2014); and Cha, Haddadi, Benevenuto, and Gummadi (2010); hint at the potential of understanding hearsay diffusion and having strategies to control them. Nevertheless, they do not cope with these strategies or their evaluation by simulation techniques. With a different goal; Weng, Menczer, and Ahn (2013)' Gupta, Lamba, Kumaraguru, and Joshi (2013); Kwon, Cha, Jung, Chen, and Wang (2013); and Qazvinian, Rosengren, Radev, and Mei (2011); propose machine learning models after an exploratory data analysis of Twitter. In a sense, the research line presented in these works is complementary of the presented here. On the one hand, machine learning approaches may employ features taken