Information Processing and Management 51 (2015) 167–186

Contents lists available at ScienceDirect

Information Processing and Management

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

Coupling learning of complex interactions

Longbing Cao *

Advanced Analytics Institute, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 5 January 2014
Received in revised form 20 August 2014
Accepted 27 August 2014
Available online 22 September 2014

Keywords:
Coupling learning
Coupling relationship
Interaction learning
Non-IIDness learning
Invisibility learning

ABSTRACT

Complex applications such as big data analytics involve different forms of coupling relationships that reflect interactions between factors related to technical, business (domain-specific) and environmental (including socio-cultural and economic) aspects. There are diverse forms of couplings embedded in poor-structured and ill-structured data. Such couplings are ubiquitous, implicit and/or explicit, objective and/or subjective, heterogeneous and/or homogeneous, presenting complexities to existing learning systems in statistics, mathematics and computer sciences, such as typical dependency, association and correlation relationships. Modeling and learning such couplings thus is fundamental but challenging. This paper discusses the concept of *coupling learning*, focusing on the involvement of coupling relationships in learning systems. Coupling learning has great potential for building a deep understanding of the essence of business problems and handling challenges that have not been addressed well by existing learning theories and tools. This argument is verified by several case studies on coupling learning, including handling coupling in recommender systems, incorporating couplings into coupled clustering, coupling document clustering, coupled recommender algorithms and coupled behavior analysis for groups.

1. Introduction

Complex interactive and unstructured/semi-structured data and applications, especially in big data, present major challenges to the current analytic and learning theories and systems. Big data, in particular, presents specific complexities of weakly structured and unstructured data distribution, dynamics, interactions, and structures, which challenge the existing theoretical and commercial systems in mathematics, statistics, and computer science. Examples include the connections between gene combinations and physical and psychological consequences, between one's personal traits or preferences in social media and one's social, behavioral, attitudinal and interest attributes.

This results in a situation where learning big data is analogous to the ancient Indian parable of seven blind men encountering an elephant for the first time. Each touches a different part of the animal, so when the seven share their experiences, each has a completely different idea of what the whole animal must look like. Similarly, when confronted with a big data set, a data modeler or learner may only see a partial set or aspect, hence often only a partial story is told by a learner. Why does this happen? There are many reasons, one of which is the invisibility of sophisticated coupling relationships (coupling for short, see Definition 2.1) hidden between the heterogeneous parts that are 'visible' to blind people. They do not have the ability to recognize the visible and invisible couplings between parts to connect those heterogeneous parts to form a global

* Tel.: +61 2 95144477.
E-mail address: longbing.cao@uts.edu.au

http://dx.doi.org/10.1016/j.ipm.2014.08.007

picture as sighted people do. This is representative of certain major challenges of complex relations hidden in complex data (particularly referring here to data with complex couplings and/or mixed distributions, formats, types and variables, and unstructured and weakly structured data). Learning visible and especially invisible coupling relationships can complement and assist in understanding weakly structured and unstructured data.

In many cases, such inherent, locally visible but globally invisible (or vice versa) couplings are presented in a range of forms, structures, and layers and on diverse entities. Often individual learners cannot tell the whole story due to their inability to identify to such complex coupling. Effectively learning the widespread, various, visible and invisible couplings is thus crucial for obtaining a true and total picture of the underlying problem.

This is not a trivial task, however. The difficulty in learning complex couplings lies not only with invisible couplings – even *visible couplings* are often overlooked. Taking the design of recommender algorithms as an example, our ability to recognize them is limited, even though these interactions and structures are embedded in applications such as social media networks. For example, there have been several recent cases in which researchers have started to incorporate inherent couplings between items and between users into a recommender system (RS) (Jannach et al., 2010; Ricci, Rokach, Shapira, & Kantor, 2011), after a long period of focusing on rating-based exploration, whereas the item-item couplings and user-user couplings (see Fig. 2) have been always intrinsic to the systems.

One reason for this is that visibility is relative to opportunity and capability. The same couplings are implicit to some people, while explicit to others. For instance, in social media recommendation, the friendship between twitters (Cheng) has only recently been recognized as enhancing social recommendation, yet it has always been a natural built-in feature of social media systems. There is a need to develop our ability to capture and convert as many invisible couplings as possible to visible coupling, and to effectively capture visible couplings in complex data.

In reviewing the existing literature, we unfortunately cannot find systematic methodologies and techniques in learning theories to address the above coupling issues. This raises a fundamental question: *how much do we know about coupling?* and many other basic questions, including: *what are couplings*, *where they are*, and *in what forms are they present*, which we need to address before we can think about how to capture and embed couplings in learning systems. Once these problems have been satisfactorily addressed, more issues follow, such as: *how to represent couplings*, *how to test whether and to what extent couplings exist in a dataset*, *how to incorporate them into learning models*, and *how to evaluate the difference they make once they are incorporated into learning systems*. These challenges form the basis of the need to study *coupling learning*, a fundamental but undeveloped area in computer science, to address the intricate coupling relationships embedded in complex data and increasingly seen in information retrieval, data mining and machine learning in particular. This is crucial for big data analytics because most existing analytics and learning theories and systems have been built on the assumption that data is independent and identically distributed (IID), while big data is essentially non-IID (Cao, 2013b). Coupling is one critical aspect of non-IIDness (Cao, 2013b) (the other is heterogeneity or so-called personalization, which is not the main concern in this paper, although coupling may be heterogeneous and involve heterogeneity in data).

Learning the above characteristics of complex couplings in big data fundamentally challenges existing learning theories and systems, including pairwise coupling (Moreira & Mayoraz, 1998; Wu et al., 2004), statistical relation learning (Dzeroski & Lavra, 2001; Getoor & Taskar, 2007), dependency learning (Neville & Jensen, 2007; Yin, Li, & Cao, 2014), association learning (Ceglar & Roddick, 2006; Lu, Feng, & Han, 2000), correlation analysis (Hair, Black, & Babin, 2009; Székely, Rizzo, Nail, & Bakirov, 2007), linkage analysis (Faloutsos, Han, & Yu, 2011; Miller, Griffiths, & Jordan, 2009), community analysis (Arenas, Danon, Dıaz-Guilera, Gleiser, & Guimera, 2004; Girvan & Newman, 2002), social network analysis (Arenas et al., 2004; Girvan & Newman, 2002; Knoke & Yang, 2007; Wasserman & Faust, 1994), multivariate time series (Székely et al., 2007), causality analysis (Gujarati & Porter, 2009) and graph analysis (Cook & Holder, 2006). They either essentially treat data as IID or only address specific forms or levels of couplings. No general and competent theories, frameworks, algorithms or tools are available to handle the coupling complexities discussed above.

The above observations motivate this work, namely to systematically state the coupling learning problem, which clearly involves interactive, unstructured and semi-structured data. The aim of this paper is multi-fold:

- High-level: build a conceptual system of coupling learning (Sections 2–4) towards a generic and comprehensive understanding of the broad-based coupling relationships that exist in complex data and applications (especially in big data related business).
- Middle-level: illustrate how to advance classic problems to another generation by incorporating coupling learning into a specific existing scientific problem such as recommender systems (Section 5).
- Low-level: showcase specific examples in recommender systems to demonstrate how couplings can be managed in practice to improve analytic outcomes (Section 6).

The purpose of this paper is therefore not to specify one particular technique for learning a particular type of coupling (instead we provide citations to our related work for such discrete discussions), but to disclose the whole nature of the problem and build generic frameworks and examples to show possible ways to address the problem.

Accordingly, the organization of this work is as follows. Section 2 discusses the concept of coupling and major coupling relationships often addressed in current big data communities. Section 3 presents a high-level picture of coupling layers and forms appearing in complex data and applications. In Section 4, the issues of modeling and measuring couplings and the curse of couplings are introduced. An example of comprehensive couplings in recommender systems is discussed in Section