The Computer Journal Advance Access published August 22, 2013

doi:10.1093/comjnl/bxt084

Non-IIDness Learning in Behavioral and Social Data

LONGBING CAO*

Advanced Analytics Institute, University of Technology, Sydney, Australia
*Corresponding author: longbing.cao@uts.edu.au

Most of the classic theoretical systems and tools in statistics, data mining and machine learning are built on the fundamental assumption of IIDness, which assumes the independence and identical distribution of underlying objects, attributes and/or values. However, complex behavioral and social problems often exhibit strong couplings and heterogeneity between values, attributes and objects (i.e., non-IIDness). This fundamentally challenges the IIDness-based learning methodologies and techniques. This paper presents a high-level overview of the needs, challenges and opportunities of non-IIDness learning for handling complex behavioral and social problems. By reviewing the nature and issues of classic IIDness-based algorithms in frequent pattern mining, clustering and classification to complex behavioral and social applications, concepts, structures, frameworks and exemplar techniques are discussed for non-IIDness learning. Case studies, related work and prospects of non-IIDness learning are presented. Non-IIDness learning is also a fundamental issue in big data analytics.

Keywords: non-IIDness learning; IIDness, IID data; non-IID data; coupling; behavior informatics; social informatics

Received 13 February 2013; revised 5 July 2013
Handling editor: Guandong Xu

1. INTRODUCTION

Behavioral and social applications are ubiquitous, ranging from business and online applications to social and organizational applications and domains. With the increasing and continuous development of such applications, an emerging need is to develop an in-depth understanding of the underlying working mechanism, driving force, dynamics and evolution of a behavioral and/or social system, as well as the impact on business and context. To this end, building on the classic theories and tools available in behavioral science and social science, behavior informatics [1, 2] and social informatics [3][1] have recently been studied to 'formalize', 'quantify' and 'compute' complex behavioral and social applications.

As an emerging area of research, behavior and social informatics is in its earliest stage and features many challenges and opportunities. A canonical trend is to develop theories, tools and algorithms based on the classic outcomes available in extant disciplines including statistics, data mining and machine learning. Typically, frequent pattern mining, clustering and classification of behavioral and social applications are conducted by expanding the corresponding existing theories and algorithms. In this paper, we discuss the potential issues and risk in pursuing this path for complex behavioral and social applications by explicitly or implicitly taking the IIDness assumption, and thus reveal the need for developing non-IIDness learning for behavior and social informatics.

Arguably, most of the existing theories, tools and systems in statistics, data mining and machine learning are built on the IIDness assumption, which assumes the independence and identical distribution of the underlying objects, attributes and/or values. Based on a high-level abstraction, it is assumed that objects, attributes and values are independent and identically distributed, with most of existing learning theories, models and algorithms proposed on the basis of this assumption. This works well in simple business applications and abstract problems with weakened and avoidable relations and heterogeneity, and serves as the foundation of classic analytics, mining and learning theories, algorithms, systems and tools.

Complex behavioral and social applications often exhibit strong coupling relations (which are beyond the usual dependency relation) and heterogeneity between objects, object

[1] See more from the IEEE Task Force on Behavior and Social Informatics and Computing: www.bsic.info

attributes and attribute values, which cannot be abstracted or weakened to the extent of satisfying the IIDness assumption. Couplings may be presented in different forms and levels on different objects, attributes and/or values. Heterogeneity is reflected through multiple or mixed structures or distributions within objects, attributes and/or values. This makes it necessary and unavoidable to consider coupling and heterogeneity in behavior and social informatics. Accordingly, non-IIDness learning emerges as a crucial issue, even though it has not been studied much, or recognized in the statistics, data mining and machine learning communities [4–19].

Motivated by the above challenges and prospects, this paper focuses on a high level discovery of the IIDness nature of the classic analytics and learning systems, and the intrinsic need and fundamental principles of non-IIDness learning for tackling complex analytics and learning problems. As the underlying problem is so novel but widespread and challenging, it is not our intention to provide a unified solution or framework here, which is out of our existing capability. However, we intend to share some preliminary efforts made towards considering non-IIDness in complex analytics and learning tasks.

In particular, we discuss the characteristics of complex behavioral and social problems in Section 2. Extended discussions are given in Section 3 about the non-IIdness feature of behavioral and social data. Section 4 presents the issues associated with the IIDness-based algorithms in classic behavior analysis, social media and recommendation systems, and social network analysis. Section 5 introduces high-level concepts and principles of non-IIDness learning. Preliminary explorations for non-IIDness learning and case studies are given in Section 6, followed by the conclusions drawn in Section 7.

2. CHARACTERISTICS OF COMPLEX BEHAVIORAL AND SOCIAL PROBLEMS

2.1. Behavioral and social system and intelligence

From an abstract perspective, a behavior and a social event can be described in terms of a four-element tuple [1, 20], consisting of *actor* (subject and/or object), *operation* (activity and activity properties), *relation* (interactions) and *context* (including environment). For an in-depth understanding of behaviors [1], interaction, coupling relationships, semantics, dynamics, change, and impact and utility are important factors to consider.

In [1], an abstract behavior model is presented. A behavior is a vector, with properties describing key aspects including subject, object, action, status, time, place, goal, plan, belief, constraint, context, associate, and impact. The couplings between behaviors from one or different actors may take place on a behavioral property or across different properties, say temporal relationship between behaviors of a blogger or a causal relationship between cars involving an accident.

A social system may be interpreted in terms of the theory of open complex systems [21–24] such as a complex multi-agent system. If this is applicable, one way to explore a complex social system is to take the OSOAD methodology [23, 24]. The OSOAD methodology argues that a complex system consists of the following key working components and mechanisms:

Accordingly, the emerging field of behavior informatics [1, 2] and social informatics [3] aims to reveal deep behavior intelligence and social intelligence in behavioral and social systems.

Behavior intelligence refers to the intelligence generated through analyzing the process, impact and utility of a collection of activities conducted by a range of actors in a certain context. From the scale perspective, we may be interested in *individual behavior intelligence*, *group behavior intelligence* or *collective behavior intelligence* [25].

An example of implementing individual behavior intelligence can be seen when an investor purchases a stock that accrues an expected profit. Pool manipulation reflects negative group behavior intelligence. Financial crisis is a negative presentation of collective behavior intelligence.

Social Intelligence refers to the intelligence that emerges from group interactions, behaviors and the corresponding regulation during a process within a context. We are concerned about *human social intelligence* and *animate/agent-based social intelligence* [25].

Human social intelligence is embodied in aspects such as social cognition, emotional intelligence, culture, consensus construction and group decision. Animated/agent-based social intelligence involves swarm intelligence, action selection and the foraging procedure. Both sides also engage social network intelligence and collective interaction, as well as social regulation rules, law, trust and reputation for governing the emergence and use of social intelligence.

Our goal here is to analyze, mine and learn deep behavior and social intelligence from behavioral and social systems by developing corresponding theories and techniques. Before we specify our task in non-IIDness learning for deep behavior and social intelligence, we discuss the complexity embedded in complex behavioral and social systems.

2.2. Complexity of behavioral and social systems

Complex behavioral and social problems exhibit intricacies that greatly challenge existing theories and techniques. The discussions about open complex intelligent systems [21, 22] and ubiquitous intelligence [25] provide high-level hints for an in-depth understanding of a complex system.

According to the theory of open complex intelligent systems, system complexity is embodied in human engagement, openness, interaction, environment, hierarchy and evolution. These aspects are embodied in behavioral and social systems in terms of specific corresponding entities and attributes. For instance, from the hierarchical perspective, a social system may consist of multiple levels of components, forming subsystems, subsystem constituents and constituent properties. Interaction may happen on a variety of levels and in various forms, such as on global and local levels and in terms of following and followed roles between or within a social group, subgroup or nodes.

Accordingly, the complexities and characteristics of complex behavioral and social problems may be discussed from the following aspects: *openness*, *large scale*, *heterogeneity*, *hierarchy*, *networking*, *coupling relationships*, *societal characteristics*, and *dynamic characteristics* [21–24, 26, 27].

Openness reflects the exchange of energy, information and materials between a behavioral/social system and its external environment. A behavioral/social system often involves or is composed of hundreds or even millions of actors and/or operations [22], forming a very *large scale*. There may be many types or forms of behaviors, behavioral actors, data sources, relationships and even impact making up the system components. This results in strong *heterogeneous characteristics*. Such system components in a behavioral/social system are likely to be organized in a *hierarchical structure* in a network. The *networking* that exists between system components is the intrinsic driving force of behavior/social intelligence emergence. Networking is further driven by different *coupling relationships* between actors, behaviors and context from temporal, inferential, combinational and party-based aspects [20]. Couplings existing in behavioral/social systems cause the underlying objects to be dependent on each other.

The *societal characteristics* of a behavioral/social system may be embodied in many social factors such as the laws of business, politics, organizational factors and business processes. In addition, behavioral/social systems are *dynamic* in the sense that they may change states, working mechanisms, constituents, and internal and external interaction mechanisms at any time, often beyond imagination.

The discussions about ubiquitous intelligence [25, 27] offer additional aspects to explore the complexity in a complex behavioral and social system. It argues the need of considering the following types of intelligence embedded explicitly or implicitly in a complex system: human intelligence, domain intelligence, behavior intelligence, data intelligence, organizational intelligence, social intelligence and networking intelligence.

Specifically, in analyzing complex behavioral and social systems, we may explore system complexity from *data*, *domain*, *context* and *impact* perspectives.

Data represent the information generated directly by behavioral and social systems and by the management systems that govern behavioral and social problems. *Domain* refers to the broad area in which the underlying behavioral and social problems exist or reside. *Context* is the particular environment which surrounds a specific behavioral and social problem. *Impact* is indicated by the outcomes produced by behavioral and social systems.

The above analysis of the underlying characteristics and complexities in behavioral/social systems discloses that behavior/social systems are strongly dependent and heterogeneous. This is inconsistent with the assumption of *IIDness*, i.e. are independent and identically distributed.

3. NON-IIDNESS IN BEHAVIORAL AND SOCIAL SYSTEMS

The above analysis in Sections 2.1 and 2.2 shows that non-IID characteristics are intrinsic in complex behavioral and social systems. Here, we further specify the coupling and heterogeneity aspects in behavioral and social systems.

3.1. Coupling

Following the abstract behavior model in [1], coupling may take place within and between behavior attributes, on different levels in a system. As discussed in open complex intelligent systems, interactions may happen within and between system elements, subsystems, and system and environment.

Couplings in complex behavioral and social systems may take different forms and structures, which are often mixed with each other. Such couplings may need to be explored from structural, semantic, probabilistic and mathematical, dynamic, and/or graphical perspectives.

Different types of coupling relationships exist in behavioral and social systems. As we discussed in [20, 28], there may be the following couplings appearing between users, items, and between users and items in a social media system or between elements and components of a behavioral system.

(i) *Serial coupling*: One behavior happens after another, or one item is purchased after another, for example, one comment in a blog triggers another one on the same topic.
(ii) *Causal coupling*: One behavior causes the occurrence of another behavior, or one social state is caused by another, for instance, a breaking news causes a significant increase of concerns in social media.
(iii) *Synchronous coupling*: All behaviors or social events occur at the same time, for instance, two bloggers comment on the same issue from different social media at the same time.
(iv) *Exclusive coupling*: Different events happen on a mutually exclusive basis, for instance, two opponent groups share different views on the same social event in a blog.
(v) *Dependent coupling*: Some behaviors or social events have required dependents such as prefix or postfix components, for instance, the occurrence of a behavior is associated with the pre-occurrence of series of other behaviors.

Further, targeting different types of behaviors or social events, couplings may present in different forms. Couplings in numeric data are very different from those in categorical data. From the number of involved behavioral and social attributes, couplings include single attribute-based couplings such as temporal coupling, and compound coupling such as hierarchical coupling. From the knowledge representation aspect, syntactic coupling, semantic coupling and inferential coupling can be explored. In [20, 28], different temporal couplings and inferential couplings are discussed for coupled behaviors.

In addition, there are couplings on different levels, from value, attribute, object, method, and measure to pattern. Such couplings, which are more comprehensive and complex than correlation and association, refer to the relations that exist explicitly or implicitly between source and destination entities. A source or destination entity can be a value, attribute, object, method or pattern in a behavioral or social system.

For example, there are user–user couplings, item–item couplings and user–item couplings in a recommendation system. Item attributes such as item price and quantity are often associated with each other. The price of one item may affect the price of another. An item may influence the sale market of another. In recommendation modeling, different methods may focus on specific aspects, and there may be a need to integrate multiple methods to cater for comprehensive couplings between item attributes, users, items and between users and items.

In the related work, the above comprehensive couplings are often ignored. Only certain relation or correlation is considered. For example, in recommendation systems, either user–user influence or item–item co-occurrence is only considered.

3.2. Heterogeneity

In behavioral and social systems, heterogeneity may appear in different aspects, input data sources, value types, object types, etc.

Often a behavioral or social system involves multiple heterogeneous (multi-structured or mixed-structured) data sources. They may be composed of divided value distributions, heterogeneous attributes, non-identical distributions of data subsets and thereafter heterogeneous objects.

(i) Value: Often different types of values present in a system, such as categorical, numerical, audio and/or video, textual and graphical data. Accordingly, there are different value characteristics such as value distributions.
(ii) Attribute: Similar to value types, various types of attributes are often engaged in a system. Different attributes may generate separated value ranges, distributions and frequencies, etc.
(iii) Object: Represented by attributes and values, object heterogeneity presents objects in different ways. For the same object, it may present in different forms in respective systems or at respective times.
(iv) Source: A behavioral or social system may involve multiple sources of information, presenting in heterogeneous values, attributes and/or objects. This forms multiple heterogeneous information, media, or channel sources.
(v) Subset: A subset of value set, attribute set, object set or source set may be selected for analysis or is only practically available. There are still heterogeneity in the subset, as discussed above.

The heterogeneity plays an essential role in understanding the difference embedded in a behavioral or social system. For instance, learning algorithms have to consider the significant difference incorporating in the attribute value range distribution and/or value frequency distribution, and further difference existing between attributes, objects and sources.

Further, depending on coupling forms, heterogeneity in a complex behavioral and social system may present in different forms, such as structural heterogeneity, semantic heterogeneity, probabilistic/mathematical heterogeneity, dynamic heterogeneity and/or graphical heterogeneity.

(i) Structural heterogeneity: there may be different structural forms between behavioral and social components on one level or across multiple levels.
(ii) Semantic heterogeneity: various semantic relations may exist in a behavior and social system.
(iii) Probabilistic heterogeneity: behaviors and social events may follow different probabilistic distributions.
(iv) Mathematical heterogeneity: behaviors and social events may be captured by different mathematical mechanisms and tools.
(v) Dynamic heterogeneity: various types of interactions and evolutionary mechanisms may exist in one or different behavioral and social systems.
(vi) Graphical heterogeneity: behavioral and social systems may be best represented by different graphical models.

The above discussed heterogeneity needs to be aligned with couplings in behavioral and social study. If heterogeneity can be converted into homogeneous cases, then the classic approaches are sufficient. Unfortunately, in complex behavioral and social systems, it is difficult or even not possible to transform a heterogeneous system into a homogeneous one to handle. This is because any partition and transformation would seriously cut off the intrinsic and sophisticated couplings between heterogeneous components. A transformed system would behave very differently from the original one if such couplings were destroyed.

Another issue is that heterogeneity is very much related to personalization. However, very limited outcomes are available on truly personalized learning, such as personalized information retrieval and personalized recommendation. In existing research, one tries to simplify the heterogeneity and personal characteristics in behavioral and social systems. For instance, in social media community learning and information retrieval, most of the existing work treats all target objects similarly, and a model is built on a population of equally treated nodes or queries. Although profiles of individuals are involved, the outcomes reflect the population-oriented features rather than personal profiles. The resultant recommendations or retrieval outcomes are based on behaviors of many objects rather than on an individual entity. This may be the key reason that existing search algorithms often bring about many irrelevant or uninteresting results.

4. ISSUES IN CLASSIC BEHAVIROAL AND SOCIAL LEARNING

In this section, we analyze the IIdness nature of several classic learning algorithms, including classic sequence analysis and recommendation algorithms.

4.1. Classic behavior analysis

Behavior analysis is widely seen in areas such as web mining, data mining, machine learning, social network analysis and business intelligence. The so-called behavior in classic research usually refers to a weak and specific, and sometime very rough or virtual, concept. Their definition is not as clear and comprehensive as the abstract behavior model discussed in [1]. Correspondingly, behavior analysis refers even to anything, we call *soft behavior analysis*. In [1, 2] about behavior informatics and computing, we target *hard behavior analysis* based on the proposed behavior model in [1]. This hard behavior analysis aims at inventing and developing computing methodologies, techniques and tools for modeling, representing, reasoning about and checking behavior-oriented systems, for the modeling, analysis, discovery and learning of dynamics, networking, group/community formation and deformation, divergence and convergence, pattern and exception, impact, risk and utility, and for the management and emergence of behaviors and behavioral systems.

The above aims and objectives are far beyond the classic efforts made in behavioral science [29, 30], social science and specifically behavioral finance and economics [31]. In these fields, behavior is usually not solidly presented, couplings and heterogeneity are overlooked or weakly addressed.

Let us take sequence analysis, a very recent focus in data mining, as an example to explore the issues in classic behavior analysis. Sequence analysis is a typical approach for analyzing behavioral sequences. Classic sequence analysis algorithms, such as , focus on positive sequences which are composed of actions only. They consider only the ordering relationship between sequential elements. The comprehensive couplings discussed in Section 3.1 are ignored.

Similarly, in negative sequence analysis, although typical algorithms including e-NSP and GA-NSP [32–35] incorporate one more relation, namely the negation of a sequential element, other couplings are overlooked, and there is no differentiation between sequences and/or between sequential elements.

In Section 6.1, we introduce the problem of coupled behavior analysis and a model for capturing coupled sequences. In [28], we discuss many different relationships between patterns, most of them are applicable for sequences. The consideration of complex couplings in behavioral sequences will create new types of sequential patterns, namely various relational sequences.

4.2. Classic social media and recommendation systems

Recommendation systems are widely used in areas such as social media, web service and online business. Typical recommendation algorithms include *collaborative filtering* and *matrix factorization*. Here, we analyze the underlying assumption behind these two algorithms and their relation to IIdness.

Collaborative filtering (CF) is the process of filtering involving collaboration among objects. Depending on the main entity to be focused, CF takes form of user-based CF and item-based CF. User-based CF assumes that if user A shares the same opinion as B, then A likely takes B's opinion on another issue. Item-based CF takes the assumption that users who buy X also buy Y.

There are several variants of CF proposed in the literature to handle filtering by addressing respective issues. Let us take the original memory-based CF algorithm [36] as an example to explore its underlying issues. Equation (1) represents its predicted vote for user a based on user i's vote $v_{i,j}$ on item j and mean vote $\bar{v}_i$ (where $w(a,i)$ is the weight of similar user i on a).

$$p_{a,j} = \bar{v}_a + k \sum_{i=1}^{n} w(a,i)(v_{i,j} - \bar{v}_i) \qquad (1)$$

$p_{a,j}$ only assumes a weak correlation between users, and does not substantially consider (1) the coupling between the votes of user i on all items, namely between v_{i,j_1} and v_{i,j_2}; (2) the influence between votes on different items for user i; (3) the coupling between different users, namely between $v_{i_1,j}$ and $v_{i_2,j}$; (4) the aggregation of both couplings; and (5) more fundamentally, it fully ignores the couplings between item attributes and between attribute values. Such couplings, if involved, could disclose intrinsic complexities of social networking, and contribute to more informative and meaningful findings for determining the collaboration between i and a. If the properties of users and items can be incorporated into the above couplings, as in the coupled similarity [37], the prediction could be based on much more solid support.

The matrix factorization (MF) approach supports a matrix R with users and items as two dimensions, with the values of users' rating on items. MF predicts the missing ratings of some users on coming items based on the approximate factorization of the matrix R into two matrices P and Q: $R \approx PXQ^{\mathrm{T}} = \hat{R}$, where P represents the association between a user and the latent features and Q captures that between an item and the latent features. The prediction of a rating $\hat{r}_{ij}$ of an item j by a user i is the dot product of the two vectors corresponding to i and j:

$$\hat{r}_{i,j} = p_i^{\mathrm{T}} = \sum_{i=1}^{k} p_{ik} q_{kj} \tag{2}$$

As we can see in the matrix and the predictive function (Equation (2)), MF is only based on a direct business linkage between users and items, but it dose not take the coupling between items and between users into consideration, not to mention the couplings between properties describing users and items.

The above analysis of the fundamental approaches used in CF and MF shows that the IIdness has been taken into account in the basic CF and MF working processes.

4.3. Classic social network analysis

Social network analysis [38, 39] has been a recent hot topic in many fields. The basic ideas of analyzing social networks involve key concepts such as graphs and matrices formed by nodes in a network to build graphical models or adjacency matrices. The similarity or dissimilarity between nodes (or any objects including actors in a network) is measured by the relation of some nodes (objects) to others. The contribution strengths can then be represented by measuring and weighting the communication, connections, flows of information, similarities/affiliations, and/or social interactions between nodes (objects).

Based on the above basic concepts, the SNA then studies typical issues such as how to represent various social networks, how to identify linkages, how to measure the strength of linkages, how to identify key/central nodes (actors) in a network, how the influence is transferred in a network, how some nodes are connected into a community, and how to measure the overall network structure, etc.

In the above SNA tasks, relation, linkage, interaction and influence are some of the core aspects for analyzing the working mechanisms and opportunities and problems in social networks. Similar to social media analysis and recommendation systems, there is usually a very weak focus in SNA on uncovering the node–node coupling, actor–actor coupling, node–actor coupling and couplings between and/or within objects, subgroups, subgraphs and communities. Heterogeneity between entities (including the above aspects) is usually ignored, by simply treating all entities equally. More specifically, the existing SNA approaches and algorithms usually just focus on the explicit linkage between objects but ignore the couplings between object properties and between property value sets.

In summary, the above discussions about classic behavior analysis, social media recommendation system, and social network analysis show that IIDness-based learning and analysis has been taken widely as a fundamental assumption in complex behavioral and social applications.

5. CONCEPTS OF NON-IIDNESS LEARNING

Here, we illustrate the assumptions of IIDness and non-IIDness, respectively, and compare their different settings. Given a learning problem as shown in Fig. 1a consisting of three heterogeneous objects from different datasets or varied feature sets (as shown in the three different symbols), our goal is to determine the position of O_3: for instance, whether it belongs to the same cluster of O_1 and O_2 or which label it can be classified as. Figure 1b and c illustrate the main working mechanisms of IIDness learning and non-IIDness learning, respectively, and their differences. Rather than digging into a specific learning task, here we focus on discussing the assumption and working mechanisms that are taken for IIDness and non-IIDness learning, which can be any specific learning objectives.

Figure 1b illustrates the approach of IIDness learning, which treats all objects as homogeneous (identically distributed, as shown in the circle) and independent (no connection between objects). The similarity or distance d is calculated between O_3 and its baseline O, say $d = \|O - O_3\|$, for which we ignore the relations between O_3 and other objects.

In conclusion, IIDness learning relies on the assumption that all objects are independent and identically distributed, which is applied to objects, object attributes, attribute values, learning objective function determination, evaluation criteria, etc. Correspondingly,

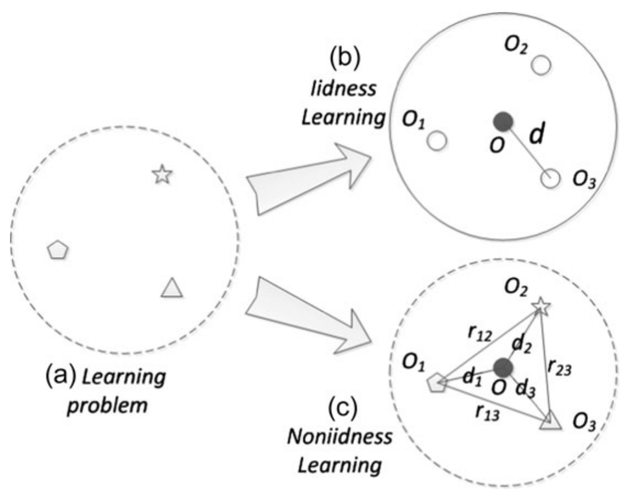

FIGURE 1. IIDness learning vs. non-IIDness learning.

(i) We treat all observations (as well as their elements, attributes and attribute values) equally and are only concerned about the similarity (or dissimilarity) between an observation and the reference (central point or mean for instance).
(ii) The reference to determine the belongingness of any observation is either a global benchmark (say the minimum support) or is obtained in the same way (the same mean value or the selection of central point).
(iii) The interactions between objects and the influence of one object on others are ignored in the belongingness determination.
(iv) The influence of objects on the reference is usually weakened. In Fig. 1a, all objects including O_1, O_2 and O_3 are treated independently, only the similarity between O_3 and the global reference O is concerned in determining O_3's belongingness.
(v) They are also treated as being identically distributed, thus the same objective function is applied to all objects during the learning process.

Figure 1c illustrates the concept of non-IIDness learning for solving the learning problem in Fig. 1a. The determination of O_3 considers (1) its direct relation with O_1, i.e. r_{13} and its relation with O_2, i.e. r_{23}, as well as the indirect relation r_{12} between O_1 and O_2. (2) The calculation of similarity d_3 between O_3 and the baseline needs to involve r_{13} and r_{23}, probably even r_{12}, say $d_3 = \|O - O_3(r_{12}, r_{13}, r_{23})\|$. (3) The baseline O for each object may be different when considering the heterogeneity between the three objects. (4) The functions for calculating d_1 and d_2 may also be different since O_1 and O_2 may follow different distributions and the interactions between O_1 and O may differ from those between O_2 and O.

When the heterogeneity and couplings between objects are concerned in the non-IIDness learning, we take one or more of the following aspects into consideration.

(i) Each object shares its own presentation (embodied in properties) which is different from others, although it might also embrace certain common characteristics with others. This indicates that objects cannot be treated equally in the modeling. As shown in Fig. 1c, we need to protect the original characteristics and personalization of each object rather than simply convert them to be similar.
(ii) There are dependencies between objects and attributes and between the values of an attribute which cannot be overlooked or simplified. This means that the learning outcome determination of one observation has to consider the influence of others. In Fig. 1c, the impact of O_1 and O_2 on O_3 needs to be considered in determining the learning outcome of O_3.
(iii) In determining the learning objective function, the benchmark (mean or central point, for instance) determination has to consider its position in the local or global space to reflect the intrinsic characteristics shared by those observations with similar distributions. In Fig. 1c, we may need to develop different baseline O for the three objects, and the similarity (distance) to O_1, O_2 and O_3 may follow diverse functions.

These aspects will be reflected in building the corresponding learning objective functions. In Fig. 1c, data characteristics analysis is conducted on the objects, making the observation that they share different distributions and should be treated in three subspaces, in which objects belonging to the same subspace share more similarities with each other than they do with those in the other spaces. Accordingly, three 'local' benchmarks rather than one are determined for each subspace. Further, in determining each object's belongingness, for instance, O_3, the coupling relationships of the object with other objects within the subspace, such as with O_1 and O_2, are considered in the objective function. In this case, we overlook the influence of couplings with objects in the other two subspaces since for the sake of simplicity, they are weak enough to be ignored.

6. NON-IIDNESS LEARNING CASE STUDIES

The assumptions and abstraction made in IIDness learning techniques seriously mismatch the reality and complexities in complex behavioral/social systems such as the coupled behavior analysis problem [40]. As we see in social media, users are inter-related and influenced by one another in terms of various aspects and reasons. Each user and his/her behaviors present specific characteristics and preferences which are usually different from those of others.

Such strong couplings and heterogeneity are particularly embodied in complex behavioral/social systems, forming the major driving forces of behavioral/social networking and evolution, which are slightly, and sometimes greatly, different from traditional applications which can be highly abstracted into an IIDness-based problem. This determines that typical approaches by expanding the classic IIDness-based algorithms and frameworks often lead to limited or incremental improvement, and cannot fundamentally solve the problem. This is challenging in handling a large-scale of behavioral/social data and general big business data, in which heterogeneity and couplings, existing in objects, interactions, behaviors and context, are two intrinsic working mechanisms and driving forces of the system dynamics and evolution.

We here briefly introduce the main principles of several preliminary attempts with case studies, to show that they can handle the coupling aspects of non-IIDness in respective learning and analytics of behavioral and social problems. They consist of

(1) *coupled behavior analysis* for analyzing intra-couplings between an actor's behaviors and inter-couplings between behaviors of different actors for group behavior understanding,

(2) *coupled item recommendation* by applying the coupled object similarity [37] to analyze the coupling between items and to convert the item-based collaborative filtering to coupled item-based item recommendation, and
(3) *term–term relation-based document analysis* to analyze semantic relations between terms appearing within and between documents.

Rather than focusing on the detailed introduction of each technique, our intention is to introduce a few exemplar approaches to show that the non-IIDness issues are manageable and the involvement of non-IIDness can lead to improved or substantial outcomes. These techniques are general and can be widely used and expanded for analyzing complex behavioral and social problems. Interested readers can find a detailed introduction from the cited references and also find some of our other efforts on quantifying similarity in categorical [37] and numerical objects [41], coupled clustering by incorporating coupled object similarity [37], analyzing non-IIDness at the method level to explore couplings between clusterings for coupled ensemble clustering [42] and considering the relations between patterns [28] and rules [43] for pattern relation analysis.

6.1. Coupled behavior analysis

In this section, we discuss the couplings between the behaviors of an individual and between the behaviors of different actors. We present a formal statement of couplings in group behaviors and the problem of analyzing such coupled group behaviors. This case study instantiates the concept of object couplings to complex behavior relations, in which behavior is a ubiquitous entity in social and business applications. Here, behavior presents heterogeneous attributes and is undertaken by many actors. The couplings between behaviors are captured in terms of probabilistic relations and the Markov assumption.

Here, behaviors refer to actions, operations, events and activity sequences conducted within certain contexts and environments in either a virtual or physical organization [1]. In practice, behaviors from the same or different actors are often associated with each other, and we call them *coupled behaviors* [40]. Coupled behaviors play a more fundamental role than individuals in the cause, dynamics and effect of business problems [1, 40]. The fundamental characteristic and challenge in understanding coupled behaviors is reflected through the intra-couplings embedded in behaviors from the same actor, the inter-couplings between those from different actors, and the aggregative couplings of both intra- and inter-couplings.

Suppose there are I actors (customers) $\{\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_I\}$, an actor $\mathscr{E}_i$ undertakes J behaviors $\{\mathbb{B}_{i1}, \mathbb{B}_{i2}, \ldots, \mathbb{B}_{iJ}\}$, actor $\mathscr{E}_i$'s jth behavior $\mathbb{B}_{ij}$ is a K-variable vector, its variable p_{ijk} reflects the kth behavior property. For the set of behaviors $\{\mathbb{B}_{ij} \mid 1 \le i \le I, 1 \le j \le J\}$, each element $\mathbb{B}_{ij}$ can be expressed as the vector $\overrightarrow{\mathbb{B}}_{ij} = ([p_{ij}]_1, [p_{ij}]_2, \ldots, [p_{ij}]_K)$, where $[p_{ij}]_k$ $(1 \le k \le K)$ is the kth property of the behavior $\mathbb{B}_{ij}$. Then, coupled behaviors are defined as follows:

Definition 6.1 (Coupled behaviors). *Coupled behaviors* $\mathbb{B}_c$ *refer to behaviors* $\mathbb{B}_{i_1 j_1}$ *and* $\mathbb{B}_{i_2 j_2}$ *that are coupled in terms of relationships* $f(\theta(\cdot), \eta(\cdot))$, *where* $(i_1 \ne i_2) \vee (j_1 \ne j_2) \wedge (1 \le i_1, i_2 \le I) \wedge (1 \le j_1 \le J_1) \wedge (1 \le j_2 \le J_2)$,

$$\mathbb{B}_c = (\mathbb{B}_{i_1}^{\theta} * \mathbb{B}_{i_2}^{\theta})^{\eta} ::= \mathbb{B}(\mathscr{E}, \mathscr{O}, \mathscr{C}, \mathscr{R}) \times \left| \sum_{i_1,i_2=1}^{I} \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} f(\theta(\cdot)_{j_1,j_2}, \eta(\cdot)_{i_1 i_2}) \odot (\mathbb{B}_{i_1 j_1}, \mathbb{B}_{i_2 j_2}), \right. \tag{3}$$

where $f(\theta(\cdot)_{i_1,i_2}, \eta(\cdot)_{i_1 i_2})$ *is the coupling function denoting the corresponding relationships between* $\mathbb{B}_{i_1 j_1}$ *and* $\mathbb{B}_{i_2 j_2}$, $\sum_{i_1,i_2=1}^{I} \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \odot$ *mean the subsequent behaviors of* $\mathbb{B}$ *are* $\mathbb{B}_{i_1 j_1}$ *coupled with* $f(\theta(\cdot)_{j_1}, \eta(\cdot)_{i_1 i_2})$, $\mathbb{B}_{i_2 j_2}$ *with* $f(\theta(\cdot)_{j_2}, \eta(\cdot)_{i_1 i_2})$, *and so on, with non-determinism.*

Corollary 6.1. *Further, coupled behaviors can be represented by behavior attributes* $\{[p_{ij}]_k \mid 1 \le k \le K\}$, *then we have the corresponding behavior adjoint matrix*:

$$\begin{aligned} M(\mathbb{B}_c) ::= M(\mathbb{B}) & \left| \sum_{i_1,i_2=1}^{I} \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} f(\theta(\cdot)_{j_1,j_2}, \eta(\cdot)_{i_1 i_2}) \odot \right. \\ & \times (\overrightarrow{\mathbb{B}}_{i_1 j_1}^{\mathrm{T}} \overrightarrow{\mathbb{B}}_{i_2 j_2}) \\ = M(\mathbb{B}) & \left| \sum_{i_1,i_2=1}^{I} \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} f(\theta(\cdot)_{j_1,j_2}, \eta(\cdot)_{i_1 i_2}) \odot \right. \\ & \times ([mp_{i_1 i_2 j_1 j_2}]_{k_1 k_2})_{K \times K}, \end{aligned} \tag{4}$$

where $\mathscr{E}_{i_1}$ *and* $\mathscr{E}_{i_2}$ *refer to two distinct actors,* $\overrightarrow{\mathbb{B}}_{i_1 j_1}^{\mathrm{T}} = ([p_{i_1 j_1}]_{k_1 s})_{K \times 1}$ *and* $\overrightarrow{\mathbb{B}}_{i_2 j_2} = ([p_{i_2 j_2}]_{s k_2})_{1 \times K}$ *refer to two distinct behavior vectors with corresponding behavior attributes*; $[mp_{i_1 i_2 j_1 j_2}]_{k_1 k_2} = [p_{i_1 j_1}]_{k_1 1} \cdot [p_{i_2 j_2}]_{1 k_2}$ *is the* (k_1, k_2) *element of the matrix multiplication* $\overrightarrow{\mathbb{B}}_{i_1 j_1}^{\mathrm{T}} \overrightarrow{\mathbb{B}}_{i_2 j_2}$; $\sum_{i_1,i_2=1}^{I} \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \odot$ *means the subsequent behavior adjoint matrix of* $M(\mathbb{B})$ *is* $\overrightarrow{\mathbb{B}}_{i_1 j_1}^{\mathrm{T}} \overrightarrow{\mathbb{B}}_{i_2 j_2}$ *coupled with* $f(\theta(\cdot)_{j_1,j_2}, \eta(\cdot)_{i_1 i_2})$, *and so on, with nondeterminism*; *and the following constraints hold*: $(i_1 \ne i_2) \vee (j_1 \ne j_2) \vee (k_1 \ne k_2) \wedge (1 \le i_1, i_2 \le I) \wedge (1 \le j_1 \le J_1) \wedge (1 \le j_2 \le J_2) \wedge (1 \le k_1, k_2 \le K)$.

Definition 6.2 (Coupled behavior analysis (CBA)). *The analysis of coupled behaviors* (*CBA Problem for short*) [40] *is to build the objective function* $g(\cdot)$ *under the condition that behaviors are coupled with each other by coupling function* $f(\cdot)$ *and satisfy the following conditions*:

$$f(\cdot) ::= f(\theta(\cdot), \eta(\cdot)), \tag{5}$$

$$g(\cdot) | (f(\cdot) \ge f_0) \ge g_0. \tag{6}$$

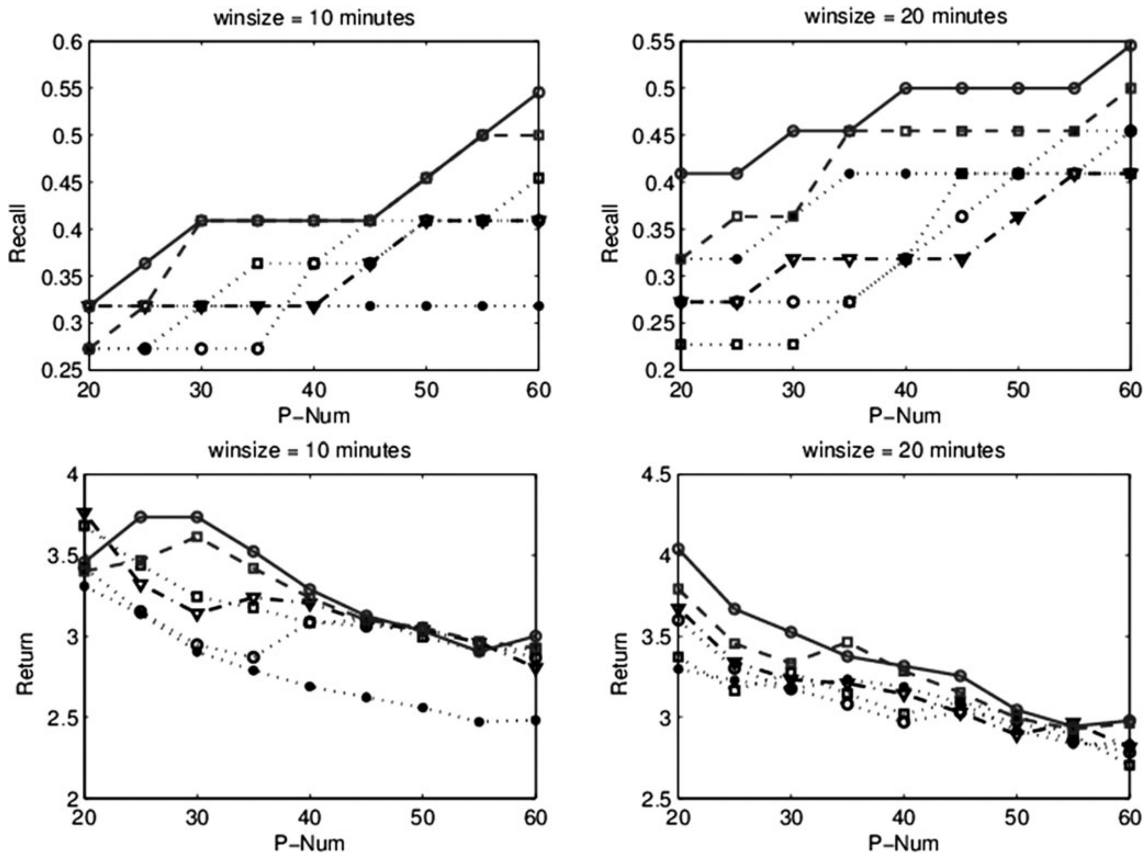

FIGURE 2. Recall and abnormal return of HMM-based CBA modeling.

The CBA problem is widespread, applicable to any behavior-oriented applications such as intelligent transport systems, and community behavior analysis in social media. In [40], an example of using coupled hidden Markov model (CHMM) is reported to model abnormal group-based financial trading behaviors (pool manipulation) in stock markets. As shown in Fig. 2, CHMM captures those abnormal group-based investment behaviors that show exceptional performance in recall as well as making abnormal return in the market, compared with the HMM models built for buy quotes (B-HMM), sell quotes (S-HMM) and trades (T-HMM), respectively, and IHMM which simply adds B-HMM, S-HMM and T-HMM without considering the couplings.

The above statement shows that the coupled behavior analysis (the CBA problem) is a typical non-IIDness learning problem, in which the intra-couplings and inter-couplings between behaviors cater for the behavior dependency. The CHMM-based case study further caters for certain heterogeneity between behavioral properties. The extension of this approach and the exploration of other effective approaches for CBA exhibit very promising opportunities for the deep analysis of behavioral and social non-IIDness.

6.2. Coupled item recommendation

As discussed in Section 4.2, the classic collaborative filtering algorithms ignore or only partially consider the couplings between item properties, user properties and item–user interactions. Here, we present a coupled item-based CF by explicitly considering both intra-coupling and inter-coupling between item attributes, and aggregating them in terms of the coupled object similarity (COS) proposed in [37]. The details can be found in [44].

The coupled item similarity (CIS) between categorical items X and Y is defined as follows:

$$\mathrm{CIS}(X, Y) = \sum_{j=1}^{n} \delta_j^A (X_j, Y_j), \tag{7}$$

where X_j and Y_j are the values of item feature j for X and Y, respectively; and δ_j^A is coupled attribute value similarity (CAVS).

The CAVS is further described by the *intra-coupled attribute value similarity* (*IaAVS*) measuring the item feature value similarity by considering the feature value occurrence frequencies within an item feature, and the *Inter-coupled*

Attribute Value Similarity (*IeAVS*) measuring the item feature value similarity by taking the item feature dependency aggregation into account.

For item feature j, IaAVS $\delta_j^{Ia}(X_j, Y_j)$ is calculated as per [37, Equation (4.2)], and IeAVS $\delta_j^{Ie}(X_j, Y_j)$ is calculated as per [37, Equation (4.7)]. Accordingly, CAVS δ_j^A between item attribute values X_j and Y_j of item feature j is as follows:

$$\delta_j^A(X_j, Y_j) = \delta_j^{Ia}(X_j, Y_j) \cdot \delta_j^{Ie}(X_j, Y_j). \tag{8}$$

Taking K-modes clustering algorithm as an example, we here create a coupled K-modes (CK-modes). Let S be a cluster generated by the previous partition of K-modes algorithm. There are M items described by categorical item features $\{a_{j_1}, a_{j_2}, \ldots, a_{j_l}\}$ belonging to the cluster S. A mode of the cluster S is an item vector $Q = [q_1, q_2, \ldots, q_l]$ to maximize the sum of the similarity between each element of S and Q. The mode of item set S with M items is a vector $Q = [q_1, q_2, \ldots, q_l]$ that maximizes:

$$\mathrm{Sim}(Q, S) = \sum_{i=1}^{M} CIS(S_i, Q). \tag{9}$$

Within the CK-modes model, the item-based collaborative filtering is adjusted to generate the prediction on item o_i for an active user u. The prediction P_{u,o_i} on item o_i for active user u is computed by the following formula:

$$P_{u,o_i} = \begin{cases} \dfrac{\sum_{\forall N_j \in N}(\mathrm{Sim}_{o_i,N_j} * R_{u,N_j})}{\sum_{\forall N_j \in N}(|\mathrm{Sim}_{o_i,N_j}|)}, & \sum(|\mathrm{sim}_{o_i,N_j}|) > 0, \\ \bar{r}_u, & \sum(|\mathrm{Sim}_{o_i,N_j}|) = 0, \end{cases} \tag{10}$$

where N is the intersection of items rated by the active user u and items grouped by the CK-modes algorithm, R_{u,N_j} represents the rating on item N_j given by the user u. Sim_{o_i,N_j} is the coupled item similarity between items o_i and N_j. $\bar{r}_u$ is the average of the active user's ratings.

We evaluate CK-modes against several widely discussed algorithms in the recommender systems, including user-based collaborative filtering algorithm [45], item-based collaborative filtering algorithm [46] and *CLUSTKNN* [47] on the MovieLens data. Figure 3 shows the throughputs of all algorithms. Here, throughput represents the number of recommendations generated per second. The user-based recommendation algorithm scans the whole user-item matrix R, its throughput does not change with the number of clusters. However, the throughput of the item-based recommendation algorithm varies with the number of neighbors selected for prediction. We plot the throughput of the item-based recommendation algorithm by setting the number of neighbors as 30 since it generates the best prediction quality.

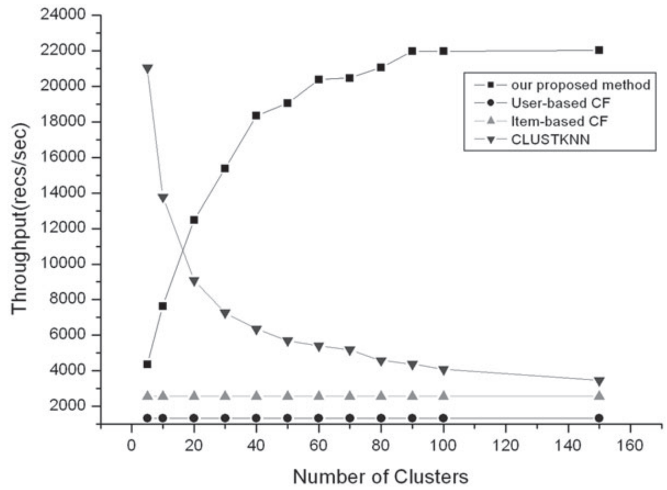

FIGURE 3. Throughput of the selected recommendation algorithms.

6.3. Term coupling-based document analysis

In classic document analysis, typical algorithms such as the Bag-Of-Words model [48] ignore the semantic relations between terms, leading to low learning performance. In [49], a new document clustering framework is proposed, which incorporates the *intra-term* coupling between terms within a document, the *inter-term* coupling between terms from different documents, and the aggregative term coupling by combining intra-term and inter-term couplings.

Terms t_i and t_j are *intra-coupled* if they co-occur in at least one document d_x ($d_x \in D$). The co-occurrence relation between terms t_i and t_j across document base D is quantified as

$$\mathrm{CoR}(t_i, t_j) = \frac{1}{|H|} \cdot \sum_{x \in H} \frac{w_{xi} w_{xj}}{w_{xi} + w_{xj} - w_{xi} w_{xj}}, \tag{11}$$

where w_{xi} and w_{xj} represent the tf-idf weights of t_i and t_j in d_x, respectively; and $|H|$ denotes the number of elements in $H = \{x | (w_{xi} \neq 0) \vee (w_{xj} \neq 0)\}$. If $H = \emptyset$, we define $\mathrm{CoR}(t_i, t_j) = 0$.

We further define the *intra-term coupling* in terms of conditional probability by normalizing the relationship between t_i and t_j, i.e. $\mathrm{CoR}(t_i, t_j)$, to [0,1] with respect to the total amount of relation between t_i and the other terms. The intra-relation reflects that when term t_i occurs in a document, the probability of term t_j that co-occurs with it together. The *intra-term coupling* $\mathrm{IaR}(t_i, t_j)$ between t_i and t_j is

$$\mathrm{IaR}(t_i, t_j) = \begin{cases} 1, & i = j, \\ \dfrac{\mathrm{CoR}(t_i, t_j)}{\sum_{i=1, i \neq j}^{n} \mathrm{CoR}(t_i, t_j)}, & i \neq j, \end{cases} \tag{12}$$

where $\mathrm{CoR}(t_i, t_j)$ is the co-occurrence relationship between terms t_i and t_j. For all the terms t_i ($i \neq j$), we have $\mathrm{IaR}(t_i, t_j) \geq 0$ and $\sum_{i=1, i \neq j}^{n} \mathrm{IaR}(t_i, t_j) = 1$. $\mathrm{IaR}(t_i, t_j)$ is usually not symmetrical.

TABLE 1. Document clustering results by using spherical K-means.

	Purity			RI			F1-measure			NMI		
Data sets	BOW	GVSM	CRM	BOW	GVSM	CRM	BOW	GVSM	CRM	BOW	GVSM	CRM
D1	0.79	0.82	**0.88**	0.49	0.57	**0.62**	0.48	0.58	**0.61**	0.32	0.41	**0.44**
D2	0.80	0.80	**0.84**	0.49	0.58	**0.60**	0.62	0.66	**0.67**	0.44	0.48	**0.44**
D3	0.78	0.79	**0.83**	0.27	0.29	**0.36**	0.41	0.43	**0.50**	0.37	0.42	**0.39**
D4	0.82	0.81	**0.85**	0.63	0.65	**0.67**	0.66	0.65	**0.65**	0.32	0.35	**0.37**

The bold values in Table 1 show the performance of our proposed algorithm CRM compared to typical methods BOW and GVSM. It indicates that CRM considering term-term couplings generally beats typical existing methods in terms of different aspects.

The *inter-term coupling* is determined by the context, namely the co-occurrences between a term and all other terms across the entire document set. Terms t_i and t_j are *inter-coupled* if there exists at least one term t_k such that both $\mathrm{IaR}(t_i, t_k) > 0$ and $\mathrm{IaR}(t_j, t_k) > 0$ hold. The term t_k is called the *link term* between them. The *relative inter-term coupling* between terms t_i and t_j linked by the term t_k is formalized as

$$R\_\mathrm{IeR}(t_i, t_j|t_k) = \min(\mathrm{IaR}(t_i, t_k), \mathrm{IaR}(t_j, t_k)), \qquad (13)$$

where $\mathrm{IaR}(t_i, t_k)$ and $\mathrm{IaR}(t_j, t_k)$ are the intra-relations between t_i and t_k, t_k and t_j, respectively. The *inter-term coupling* between t_i and t_j is defined by their interactions with all the link terms, formalized as

$$\mathrm{IeR}(t_i, t_j) = \begin{cases} 0, & i = j, \\ \dfrac{1}{|L|} \displaystyle\sum_{\forall t_k \in L} R\_\mathrm{IeR}(t_i, t_j|t_k), & i \neq j, \end{cases} \qquad (14)$$

where $|L|$ denotes the number of link terms in $L = \{t_k|(\mathrm{IaR}(t_k, t_i) > 0) \wedge (\mathrm{IaR}(t_k, t_j) > 0)\}$, and $R\_\mathrm{IeR}(t_i, t_j|t_k)$ is the relative inter-relation between t_i and t_j linked by t_k. If $L = \emptyset$, we define $\mathrm{IeR}(t_i, t_j) = 0$. The value of $\mathrm{IeR}(t_i, t_j)$ falls in [0, 1]. When there is not a link term for t_i and t_j, we regard $\mathrm{IeR}(t_i, t_j) = 0$.

Then the overall *term–term coupling* $\mathrm{CR}(t_i, t_j)$ (namely the similarity between terms) across all documents is measured by aggregating the intra-term coupling and the inter-term coupling.

$$\mathrm{CR}(t_i, t_j) = \begin{cases} 1, & i = j, \\ \alpha \cdot \mathrm{IaR}(t_i, t_j) + (1 - \alpha) \cdot \mathrm{IeR}(t_i, t_j), & i \neq j, \end{cases} \qquad (15)$$

where $\alpha \in [0, 1]$ is the parameter that decides the weight of intra-relation, $\mathrm{IaR}(t_i, t_j)$ and $\mathrm{IeR}(t_i, t_j)$ are the respective intra-relation and inter-relation between terms t_i and t_j.

The term–term coupling based similarity can then be applied to different document clustering models to cluster documents. Table 1 illustrates the Purity, Rand index (RI), F_1-measure and normalized mutual information (NMI) of the clustering results by coupled term–term relation model (CRM) for the spherical k-means algorithm incorporated with $\mathrm{CR}(t_i, t_j)$ on data sets Newsgroups and WebKB, compared to the classic bag of words (BOW) model and the GVSM model.

7. CONCLUSIONS

In this paper, we have presented a high-level picture of the non-IIDness learning problem for handling strong couplings and heterogeneity in complex behavioral and social applications. Such problems cannot be tackled well by most extant methods and systems since they overlook or abstract the couplings and heterogeneity, by taking the strong assumption of IIDness (independence and identical distribution).

We have discussed the challenges of analyzing complex behavioral and social applications, the issues surrounding the extant IIDness-based learning approaches in classic behavior analysis, social media and recommendation systems, and social network analysis, and the concepts of non-IIDness learning. Several exemplar techniques have been discussed for non-IIDness learning, including coupled behavior analysis for analyzing group behaviors, and coupled item recommendation for considering couplings between items, and term–term coupling-based document analysis by considering the semantic relation between terms across documents.

The concept and ideas of non-IIDness learning allow us to comprehensively, systematically and deeply explore the couplings between values, attributes, objects, methods and patterns, and heterogeneity residing in value matching, value frequency, attribute distribution, attribute co-occurrence, objects, methods and patterns.

ACKNOWLEDGEMENTS

This work is sponsored in part by Australian Research Council Discovery Grants (DP1096218 and DP130102691) and ARC Linkage Grant (LP100200774). We appreciate the valuable comments from all anonymous reviewers and editors.

REFERENCES

[1] Cao, L. (2010) In-depth behavior understanding and use: the behavior informatics approach. *Inform. Sci.*, **180**, 3067–3085.

[2] Cao, L. and Yu, P.S. (eds) (2012) *Behavior Computing: Modeling, Analysis, Mining and Decision*. Springer, Berlin.

[3] Liu, H., Salerno J. and Young, M.J. (eds) (2008) *Social Computing, Behavioral Modeling, and Prediction*. Springer, Berlin.

[4] Steinwart, I. and Christmann, A. (2009) Fast learning from non-i.i.d. observations. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I. and Culotta, A. (eds), *Advances in Neural Information Processing Systems*, 22, pp. 1768–1776. MIT Press.

[5] Zhang, X., Song, L., Gretton, A. and Smola, A. (2009) Kernel Measures of Independence for Non-iid Data. *Proc. NIPS2009*, pp. 1937–1944. MIT Press.

[6] Pan, Z. and Xiao, Q. (2009) Least-square regularized regression with non-iid sampling. *J. Statist. Plan. Inference*, **139**, 3579–3587.

[7] Guo, Z. and Shi, L. (2011) Classification with non-iid Sampling. *Math. Comput. Modell.*, **54**, 1347–1364.

[8] Cinbis, R., Verbeek, J. and Schmid, C. (2012) Image Categorization using Fisher Kernels of non-iid Image Models. *Proc. CVPR2012*, pp. 2184–2191. IEEE Press.

[9] Mohri, M. and Rostamizadeh, A. (2009) Rademacher Complexity Bounds for non-i.i.d. Processes. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I. and Culotta, A. (eds), *Advances in Neural Information Processing Systems*, pp. 1097–1104. MIT Press.

[10] Mohri, M. and Rostamizadeh, A. (2010) Stability bounds for stationary φ-mixing and β-mixing processes. *J. Mach. Learn. Res.*, **11**, 789–814.

[11] Chazottes, J., Collet, P., Kulske, C. and Redig, F. (2007) Concentration inequalities for random fields via coupling. *Probab. Theory Relat. Fields*, **137**, 201–225.

[12] Kontorovich, L. and Ramanan, K. (2008) Concentration inequalities for dependent random variables via the Martingale method. *Ann. Probab.*, **36**, 2126–2158.

[13] Modha, D. and Masry, E. (1998) On the consistency in nonparametric estimation under mixing assumptions. *IEEE Trans. Inform. Theory*, **44**, 117–133.

[14] Ralaivola, L., Szafranski, M. and Stempfel, G. (2010) Chromatic PAC-Bayes bounds for non-IID Data: applications to ranking and stationary β-mixing processes. *J. Mach. Learn. Res.*, **11**, 1927–1956.

[15] Zhou, Z., Sun, Y. and Li, Y. (2009) Multi-Instance Learning by Treating Instances as non-I.I.D. Samples. *Proc. ICML2009*, pp. 1249–1256. Omnipress.

[16] Dundar, M., Krishnapuram, B., Bi, J. and Rao, R. (2007) Learning Classifiers When The Training Data Is Not IID. *Proc. IJCAI2007*, pp. 756–761. AAAI Press.

[17] Tillman, R. (2009) Structure Learning with Independent Non-Identically Distributed Data. *Proc. ICML2009*, pp. 1041–1048. Omnipress.

[18] Ping, W., Xu, Y., Ren, K., Chi, C. and Shen, F. (2010) Non-I.I.D. Multi-Instance Dimensionality Reduction by Learning a Maximum Bag Margin Subspace. *Proc. AAAI2010*, pp. 551–556. AAAI Press.

[19] Song, L., Smola, A., Gretton, A., Bedo, J. and Borgwardt, K. (2012) Feature selection via dependence maximization. *J. Mach. Learn. Res.*, **13**, 1393–1434.

[20] Wang, C. and Cao, L. (2012) Modeling and Analysis of Social Activity Process. In Cao, L. and Yu, P.S. (eds), *Behavior Computing*, pp. 21–35. Springer, Berlin.

[21] Cao, L. and Dai, R. (2008) *Open Complex Intelligent Systems*. Post & Telecom.

[22] Cao, L., Dai, R. and Zhou, M. (2009) Metasynthesis: M-space, M-interaction and M-computing for open complex giant systems. *IEEE Trans. Syst. Man Cybern.—Part A*, **39**, 1007–1021.

[23] Cao, L., Zhang C. and Zhou M. (2008) Engineering open complex agent systems: a case study. *IEEE Trans. Syst. Man Cybern.—Part C Appl. Rev.*, **38**, 483–496.

[24] Cao, L., Zhang C. and Dai, R. (2005) Organization-oriented analysis of open complex agent systems. *Int. J. Intell. Control Systems*, **10**, 114–122.

[25] Cao, L., Luo, D. and Zhang, C. (2009) Ubiquitous Intelligence in Agent Mining. *Proc. ADMI*, pp. 23–35. Springer.

[26] Cao, L., Yu, P.S., Zhao, Y. and Zhang, C. (2010) *Domain Driven Data Mining*. Springer, Berlin.

[27] Cao, L. (2010) Domain driven data mining: challenges and prospects. *IEEE Trans. Knowl. Data Eng.*, **22**, 755–769.

[28] Cao, L. (2013) Combined mining: analyzing object and pattern relations for discovering and constructing complex yet actionable patterns. *WIREs Data Mining Knowl. Disc.*, **3**, 140–155.

[29] Fishbein, M. and I. Ajzen. (2009) *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press.

[30] Fisher, W., Piazza, C. and Roane, H. (eds.) (2011) *Handbook of Applied Behavior Analysis*. The Guilford Press.

[31] Burton, E. and Shah, S. (2013) *Behavioral Finance: Understanding the Social, Cognitive, and Economic Debates*. Wiley, New York.

[32] Zhao, Y., Zhang, H., Cao, L., Zhang, C. and Bohlscheid, H. (2009) Mining Both Positive and Negative Impact-Oriented Sequential Rules From Transactional Data. *Proc. PAKDD2009*, pp. 656–663. Springer.

[33] Zhao, Y., Zhang, H., Wu, S., Pei J., Cao, L., Zhang, C. and Bohlscheid, H. (2009) Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns. *Proc. ECML-PKDD2009*, Lecture Notes in Artificial Intelligence 5782, pp. 648–663. Springer.

[34] Zheng, Z., Zhao, Y., Zuo, Z., Cao, L., Zhang, H., Zhao, Y. and Zhang, C. (2010) An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns. *PAKDD2010*, pp. 262–273. Springer.

[35] Dong, X., Zhao, Y., Cao, L., Zhao, Y., Zhang, C., Li, J., Wei W. and Ou, Y. (2011) e-NSP: Efficient Negative Sequential Pattern Mining Based on Identified Positive Patterns Without Database Rescanning. *Proc. CIKM2011*, pp. 825–830. ACM Press.

[36] Breese, J., Heckerman, D. and Kadie, C. (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *UAI-98*, pp. 43–52. AAAI Press.

[37] Wang, C. and Cao, L. *et al.* (2011) Coupled Nominal Similarity in Unsupervised Learning. *Proc. CIKM2011*, pp. 973–978. ACM Press.

[38] Scott, J. (1991) *Social Network Analysis: A Handbook*. Sage Publications, Newbury Park, CA.

[39] Kadushin, C. (2012) *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press, Oxford.

[40] Cao, L., Ou, Y. and Yu, P.S. (2012) Coupled behavior analysis with applications. *IEEE Trans. Knowl. Data Eng.*, **24**, 1378–1392.

[41] Wang, C., She, Z. and Cao, L. (2013) Coupled Attribute Analysis on Numerical Data. *Proc. IJCAI2013*. AAAI Press.

[42] Wang, C., She, Z. and Cao, L. (2013) Coupled Clustering Ensemble: Incorporating Coupling Relationships both Between Base Clusterings and Objects. *Proc. ICDE2013*, pp. 374–385. IEEE Press.

[43] Li, J., Wang, C., Cao, L. and Yu, P. (2013) Efficient Selection of Globally Optimal Rules on Large Imbalanced Data Based on Rule Coverage Relationship Analysis. *Proc. SDM2013*. SIAM / Omnipress.

[44] Yu, Y., Wang, C., Gao, Y., Cao, L. and Sun, J. (2013) A Coupled Clustering Approach for Items Recommendation. *Proc. PAKDD2013*, pp. 365–376. Springer.

[45] Resnick, P. and Iacovou, N. and Suchak, M., Bergstrom, P. and Riedl, J. (1994) GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *CSCW Proc.*, pp. 175–186. Morgan Kaufmann.

[46] Sarwar, B. and Karypis, G. and Konstan, J. and Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. *Proc. WWW2001*, pp. 285–295. ACM Press.

[47] Al Mamunur Rashid, S.K.L. and Karypis, G. and Riedl, J. (2006) ClustKNN: a Highly Scalable Hybrid Model-& Memory-Based CF Algorithm. *Proc. WebKDD 2006*. ACM Press.

[48] Huang, A. and Milne, D. and Frank, E. and Witten, I. (2009) Clustering documents using a Wikipedia-based Concept representation. *Adv. Knowl. Disc. Data Mining*, pp. 628–636.

[49] Cheng, X., Miao, D., Wang, C. and Cao, L. (2013) Coupled Term–Term Relation Analysis for Document Clustering. *Proc. IJCNN2013*. IEEE Press.