A Preference-Based Multi-Agent Data Mining Framework for Social Network Service Users’ Decision Making

Ileladewa Adeoye Abiodun, and Cheng Wai Khuen

Abstract—Multi-Agent Systems (MAS) emerged in the pursuit to improve our standard of living, and hence can manifest complex human behaviors such as communication, decision making, negotiation and self-organization. The Social Network Services (SNSs) have attracted millions of users, many of whom have integrated these sites into their daily practices. The domains of MAS and SNS have lots of similarities such as architecture, features and functions. Exploring social network users’ behavior through multi-agent model is therefore our research focus, in order to generate more accurate and meaningful information to SNS users. An application of MAS is the e-Auction and e-Rental services of the Universiti Cyber AgenT(UniCAT), a Social Network for students in Universiti Tunku Abdul Rahman (UTAR), Kampar, Malaysia, built around the Belief-Desire-Intention (BDI) model. However, in spite of the various advantages of the BDI model, it has also been discovered to have some shortcomings. This paper therefore proposes a multi-agent framework utilizing a modified BDI model- Belief-Desire-Intention in Dynamic and Uncertain Situations (BDIDUS), using UniCAT system as a case study.

Keywords—Distributed Data Mining, Multi-Agent Systems, Preference-Based, SNS.

I. INTRODUCTION

COMPUTER systems evolved from centralized monolithic computing devices supporting static applications, and over the years, metamorphosed into client-server environment that allows complex form of distributed computing, and in fact, there had been predictions of machines with humanlike intelligence software that can think and reason - tools that free us from mundane and repetitive tasks [9]. The history of computing to date has actually been marked by five important and continuing trends: Ubiquity, Interconnection, Intelligence, Delegation, and Human-orientation [24], as shown in Fig. 1.

By ubiquity, it simply means that the continual reduction in the cost of computing capability has made possible to introduce processing power into places and devices where it would hitherto have been uneconomic, and perhaps even unimaginable. By interconnection, it simply means that the earliest computers were isolated entities communicating only with their human operators, but computer systems today are usually interconnected. By intelligence, it simply means that the complexity of tasks that we are capable of automating and delegating to computers has also grown steadily, hence are relatively ever more intelligent systems. By delegation, it simply implies that we give control to computer systems. For example, we routinely delegate to computer systems such safety-critical tasks as piloting aircraft. In fact, in fly-by-wire aircraft, the judgment of a computer program is trusted over that of experienced pilots. By human orientation, it simply implies a move away from machine-oriented views of human-computer interaction towards concepts and metaphors that more closely reflect the way in which we ourselves understand the world. Hence, contrary to the earliest days of computing when programmers had no choice but to program their computers with raw machine code which implied a detailed understanding of the internal structure and operation of their machines, today, programming paradigm have progressed away from such low-level views: witness the development of assembly languages, through procedural abstraction, to abstract data types, and most recently objects. Each of these developments has allowed programmers to conceptualize and implement software in terms of higher-level-more human-oriented abstraction. Lastly, by preference-based, it means goal- or utility-based, as discussed in Section V.

Adeoye Ileladewa, Cheng Wai Khuen are with Universiti Tunku Abdul Rahman (UTAR), Department of Computer Science, Faculty of Information and Communication Technology, Bandar Barat, 31900, Kampar, Perak, Malaysia (phone: +60166034155; e-mail: ileladewaadeoye@yahoo.com).
The trends to increasing delegation and intelligence imply ability of systems to operate independently without our direct intervention, and the need to build computer systems that can act effectively on our behalf. These trends have led to the emergence of a new field in computer science: multi-agent systems. In the pursuit of the above goal and to improve our standard of living, systems around the world are evolving rapidly, which include Multi-Agent System (MAS), being widely used in many systems nowadays whether it is an online application or offline system.

A. MAS and Users’ Behaviour Capturing UniCAT

MAS, is a system where multiple intelligent agents communicate, observe and act upon the environment, which directs its activity towards achieving certain interests or goals. MAS has found its applications in so many areas like Robocups, the Second Life in Virtual World; the eBay; as well as in Social Networking. A Social Network Service (SNS) can be defined as “internet – or mobile-device-based social spaces designed and created to facilitate communication, collaboration and content sharing (information or common interests) across network of contacts”. An example also is its application in the e-Auction and e-Rental services of UniCAT (Universiti Cyber AgenT), a Social Network for students in Universiti Tunku Abdul Rahman (UTAR), Perak Campus, Kampar, built around the Belief-Desire-Intention (BDI) model. The system consists of multiple interacting agents capable of: i) performing autonomous action-decide function on what they need to do for satisfying their specific goals; and ii) interacting with other agents - exchange information, cooperation, coordination, and Negotiation [9].

II. PROBLEM STATEMENT

However, in spite of the various advantages of the BDI model, it has also been discovered to have some shortcomings, which include (i) its partial address of the issue of temporal persistence (ii) the architecture does not have (by design) any look-ahead deliberation or forward planning, and (iii) the model, also does not support the simulation of unpredictability within a human-being, rather its main concern is projecting the attribute so the user within the system in an attempt to assist in social networking. In order to build a robust MAS, it has been argued that the BDI architecture should incorporate an appropriate computational model, and to bring about the required level of human orientation needed in agents to represent users, intelligent systems which use Artificial Intelligence (AI), also referred to as machine intelligence or computational intelligence, in dynamic and uncertain situations, are inevitable [20]. Hence, this paper therefore aims at providing a hybrid multi-agent framework utilizing a modified BDI - Belief-Desire-Intention model in Dynamic and Uncertain Situations (BDIDUS), using the UniCAT system as a case study, in order for the agents to take better and more informed decisive actions on behalf of the SNS users, to increase the accuracy in decision making. That is, this paper proposes a Framework that captures social network users’ behavior, and based on the user’s utility or preference for a goal, enables the agents to take better and more informed decisive actions that best suit the goal on behalf of the SNS’s users in dynamic and uncertain situations.
III. RELATED WORKS

Several behaviour capturing methods in mobile application had been studied, for example, PeopleNet, SmartNet, etc. Facebook, Myspace, Friendster, Plazes, and Jambo Networks are just a few examples of commercialized operations [26]. Other works that have deployed multi-agent to address data capturing and decision-making challenges of SNS users' behavior using different frameworks include the work of Peter in [32] on “Layered Learning in Multi-Agent Systems”, which contributes several techniques for building agents in the robotic soccer domain, only focused or centers on the particularly prevailing class of real-world domains: real-time, noisy, collaborative and adversarial multi-agent environments.

The work did not consider uncertainties in the dynamism and unpredictability of future situations. For example, if the soccer server based on a 2-dimensional moved to a 3-dimensional simulation; the work in [33] on “g-BDI: A Graded Intentional Agent Model for Practical Reasoning”, it was considered that making the BDI architecture more flexible will allow for designing and developing intentional agents potentially capable to have a better performance in uncertain and dynamic environments; “Multi-Agent System framework to support the decision-making in complex real-world domains” [35], a Rule-based Reasoning (RBR) tool was used to build a domain-independent framework for decision-making; “Multi-AgentSystem for Supply Chain Modeling: Methodological Frameworks” [36]; “A Framework for BDI Agent-based Software Engineering” [37], the BDI Agent Software Development Process (BDI-ASDP) was innovated with the aid of the Use Case, the method of OO, to find the Desires and Intentions, and the Data Flow Diagram, the method of functionality, to find the Beliefs; “A BDI Agent Based Framework for Modeling and Simulation of Cyber Physical Systems (CPS)” [38], a BDI agent based software framework was proposed to enable efficiently, modeling and simulation of heterogeneous CPS systems in an integrated manner. Most of the works, especially the ones mentioned above lack the look-ahead deliberation or forward planning to get more reliable and useful information needed in decision-making, which is inevitable especially in dynamic and uncertain situations.

Other works on Data Mining includes a project, Reality Mining. Eagle in [28] tries to model user behaviours and the complex social network via mobile phones which capture data on users’ location, proximity, communication and device usage behaviour. By continually logging and time-stamping information about a user's activity, location, and proximity to other users, the dynamics of large-scale human behaviour can be measured. As also pointed out in [34,82], building a monolithic database, in order to perform non-distributed data mining, may be infeasible or simply impossible in many applications. The cost of transferring large blocks of data may be prohibitive and result in very inefficient implementations. Surveys [34,56] and [34,77] provide a broad, up-to-date overview of DDM, touching on issues such as: clustering, association rule mining, basic statistics computation, Bayesian network learning, classification, and the historical roots of DDM. The collection of papers in Kargupta and Chan [34,54] describes a variety of DDM algorithms (ARM, clustering, classification, preprocessing, etc.), systems issues in DDM (security, architecture, etc.), and some topics in parallel data mining. Survey [34,110] discusses parallel and distributed association rule mining in DDM. Survey [34,111] discusses a broad spectrum of issues in DDM and parallel data mining and provides a survey of distributed and parallel association rule mining and clustering. Other DDM applications [52, 89] deal with continuous data streams. An overview of the data stream mining literature can be found in [34,10].

Most DDM methods in the literature operate over an abstract architecture which includes multiple sites having independent computing power and storage capability. Local computation is done on each of the sites and either a central site communicates with each distributed site to compute the global models or a peer-to-peer architecture is used [34,87].

IV. MAS AND DATA MINING

The use of MAS provides a radical alternative approach to DM where collections of data mining agents (of various types) can be used to address traditional DM problems under decentralized control. Autonomy and decentralized control are, arguably, the most significant features of MAS that serve to distinguish such systems from distributed or parallel approaches to computation. Autonomy and decentralized control control [34,107] imply that individual agents, within MAS, operate in an autonomous manner and are (in some sense) self-deterministic. Robustness, in turn, is a feature of the decentralized control, where the overall system continues to operate even though a number of individual agents have disconnected (crashed). Decentralized control also supports extendibility, in that additional functionality can be added simply by including further agents. The advantages of sharing expertise and resources are self-evident.

MAS offer some advantages which are entirely applicable to Knowledge Discovery in Data (KDD) where a considerable collection of tools and techniques are current. MAS also have some particular advantages to offer with respect to KDD, and particularly data mining, in the context of sharing of resources and expertise. KDD is concerned with the extraction of hidden knowledge from data. Very often data relevant to one search is not located at a single site, it may be widely-distributed and in many different forms.

A. MAS and Distributed Data Mining

The term Multi-Agent Data Mining (MADM) refers to Data Mining within a Multi-Agent Environment. Naturally, data mining is often applied to sensitive data. MAS allow data to be mined remotely. Similarly, with respect to data mining algorithms, MAS can make use of algorithms without necessitating their transfer to users, thus contributing to the preservation of intellectual property rights. The vision that the work espouses is that of an anarchic and dynamic agent
community where agents interact with one another to address DM problems posted by users and where data sources (agents) can be made available by individuals as desired by those individuals [34].

B. Reactive Intelligent System (RIS)

Fig. 4 Single Intelligent within a single loop [20]

Fig. 5 Intelligent system within the multiple event loops [20]

RIS model is strongly situated in its environment and it is highly responsive to changes in its environment. They are designed as event-based systems which loop through the basic operations of sensing events, filtering events and then triggering system actions, and the pre-set actions may be directly triggered from sensor inputs without any conditions, as shown in Fig. 4, as discussed in [20]. In a simple reactive system, where there is only one event loop, only one event can trigger an action at any time, meant for simple (context-aware) control system (as in Fig. 4). Whereas, more complex designs (as in Fig. 5) is a design needed to consider how to respond to multiple simultaneous events and how to handle multiple actions which, when triggered, may conflict. In practice, many systems are designed not to be purely reactive but to combine reactive behavior with other types of behavior such as model-based, goal-based, or utility-based behavior. These are Hybrid reactive systems.

V. THE PROPOSED FRAMEWORK - BIDIDUS

As mentioned earlier in section 2, this research requires providing a hybrid multi-agent framework that utilizes a modified BDI - the Belief-Desire-Intention in Dynamic and Uncertain Situations (BIDIDUS) model for capturing SNS users’ behavior promptly anytime, anywhere to reduce users’ frequent intervention in online SNS data capturing, reorganizing and structuring the captured data into beliefs, based on users’ preferences or priorities, performing Data Mining using an Agent-based Distributed Data Mining (ADDM) model, to increase the reasoning horizons of agents, in order for the agents to take better and more informed decisions on behalf of the Social Network Service (SNS) users in dynamic and uncertain situations, to increase the accuracy in decision making, using the UniCAT as a case study.

<table>
<thead>
<tr>
<th>Strong or weak intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical (embodied) hardware, e.g. robots or virtual software, e.g. software agents</td>
</tr>
<tr>
<td>Fundamental properties such as autonomous, social, reactive, proactive, etc.</td>
</tr>
<tr>
<td>Thinking (cognitive) or acting (behavior)</td>
</tr>
<tr>
<td>Human or rational</td>
</tr>
<tr>
<td>Complex organisms (explicit, high-level, knowledge-based action selection) or simple cellular organisms (implicit, low-level action selection)</td>
</tr>
<tr>
<td>Type of design architecture: reactive, model-based, goal-based, utility-based, etc.</td>
</tr>
<tr>
<td>Learning or non-learning</td>
</tr>
<tr>
<td>Certainty or Uncertainty</td>
</tr>
<tr>
<td>The environments in which intelligent systems operate: observable, deterministic, sequential, etc.</td>
</tr>
<tr>
<td>Individual intelligent entities or as multiple, collective, intelligent entities</td>
</tr>
</tbody>
</table>

The agent-oriented paradigm proposed to be adopted is highly suited for behaviour capturing applications dealing with complex and dynamic environments, that are usually unpredictable (it is not possible to predict the future state of the environment), and unreliable (the action of the agent can fail because of factors beyond its control), and the pervasive environments are really dynamic and enable software
applications to have access to large amounts of information from anywhere [25]. The ultimate objective of which is to provide technology solutions that enable anytime, anywhere information and data exchange and open access to services and application, whether used for business or pleasure. Hence, Intelligent Systems (IS) are systems which use artificial intelligence (AI), also referred to as machine intelligence, computational intelligence and include agent-based systems, software agents and robots. Dimensions along which intelligent systems can be classified are as shown in Table I.

A. The Framework Model

![Fig. 6 Hybrid Goal / Utility/Preference model-based IS [20]](image)

A goal-based IS, also referred to as planning-based IS, defines an internal plan or sequence of actions to achieve a future system goal (Fig. 5). Unlike the environment model-based IS, the action selection here depends on which is the next best system action to take the system towards a future goal state, regardless of its utility (a quantifiable measure of the performance or worth or usefulness of a specific goal among a set of possible goals). Whereas, a Utility-based IS design compares the utilities of different possible outcomes, and selects the action outcome with the highest utility. Sometimes it may require some experience or training phases in order to set and tune the utility function values, hence, goal-based action selection may be preferred over utility-based action selection in a new environment.

B. The Hybrid Multi-Agent System Model

Even though horizontally layered architecture has a great advantage of conceptual simplicity in design: n different types of behavior to be exhibited by an agent needs n layers, the danger of the overall decision of the agent not being coherent is high. For consistent decision making, a mediator (central control system) is problematic. For m possible outcomes of each layer, m^2 of such interactions among layers are to be considered in the design – practically difficult. This system also introduces a bottleneck into agent’s decision making. Whereas in vertical, there are n-1 interfaces between n layers, and there are m^2(n-1) interactions to be considered between layers. The complexity of interactions between layers is reduced. Clearly much simpler but at a cost of flexibility: Decision taking by agent causes control to pass between each different layer – not fault tolerant because failures in any one layer are likely to have serious consequences for agent performance (e.g. Touring Machines and Muller’s InterRap) [24]. Therefore, the Hybrid Goal or Utility / Preference-based Intelligent System model, suitable, and that is adopted for this framework is as shown in Fig. 6.

C. The New BDIDUS MAS Design

From the discourse in sections A and B, the proposed framework can be divided into steps highlighted below:

- The Users’ Behaviour Capturing Phase
- The Data Mining Phase
- The Mining-Based Decision Taking Phase

* * *

Users’ Behaviour Capturing Phase: This phase comprises of the boxes painted pink, designed for the capturing of the SNS users’ behaviours.

* Data Mining Phase: This phase comprises of the boxes painted yellow, designed for the reorganizing and mining of the captured SNS users’ behaviours.

* Decision-Making Phase: This phase comprises of the boxes painted ash, designed for decision-making of the mined SNS users’ behaviours based on preferences / utilities.

The block diagram illustrating the sequence of the above three phases is as shown in Fig. 7, which is actually in a cyclic form but broken down into its three respective phases as shown Fig. 6a, Fig. 6b, and Fig. 6c, because of the specified given format for this paper.
The combined diagram of Fig. 8, Fig. 9, and Fig. 10 represent the flow of the 3 phases of the BDIDUS MAS Design.

**D. System Development Tool**

The ability of the various development toolkits in handling the implementation and architecture of the provided communication and agent functionality differs. As a result, when selecting a toolkit to build some desired MAS developers should make their decision based on the MAS goals and services that are desired. Any potential toolkit should also be evaluated for potential problems related to the toolkits, strengths, and weaknesses prior to any decision being made [34].

JADEX has been chosen for the proposed BDIDUS framework development. This is because JADEX is a software environment, implemented in the JAVA programming language, directed at the development of MAS. It is a FIPA compliant middleware that enables development of peer to peer applications based on the agent paradigm. JADE defines an agent platform that comprises a set of containers, which may be distributed across a network (as desired in the case of EMADS). The goal of JADEX is to simplify the development of MAS while at the same time ensuring FIPA compliance through a comprehensive set of system services and agents. While appearing as a single entity to the outside observer, a JADEX agent platform can be distributed over several hosts each running an instance of the JADE runtime environment.

Summarily, JADEX been selected for a variety of reasons as follows:
- JADEX is both popular and regularly maintained (for bug fixes and extra features).
- It was developed with industry quality standards.
- The tool kit covers many aspects of MAS, including agent models, interaction, coordination, organization, etc.
VI. DISCUSSION ON ADOPTION OF BDIDUS INTO UNICAT

There are three types of components needed to implement decentralized Data Mining [39]. However, because of the peculiarity of this work—a modified Data Mining Framework, the fourth one is necessary to be added for the framework. Hence, the four types of components needed to implement BDIDUS framework into UniCAT are:

- Data Providers: The applications that generate the data to be mined.
- Mobile Clients: The applications that require the execution of data mining computations on remote data.
- Mining Servers: Server nodes used for storing the data generated by data providers and for executing the data mining tasks submitted by mobile clients.
- The UniCat server: The application for updating an overall SNS server, to which all other mining server nodes are connected, from which decisions are made because it contains the past historic and current mined data, periodically updated to cater for any prevailing or future situations that need decisive actions.

As shown in Fig. 11, data generated by data providers is collected by mining servers, whose main role is to allow the mobile clients to perform the analysis of remote data by using a set of data mining algorithms. Once connected to a given server, the mobile client allows a user to select the remote data to be analyzed and the algorithm to be run. When data mining task has been completed on the mining server, the results of the computation are available on the user device either in textual or visual form. However depending on the application requirements, data coming from a given provider could be stored in more than one mining server. By this arrangement, SNS users are connected to one another, and also made to have access to the servers online, so as to perform any mining operation deemed fit at any time, update the server regarding any present or anticipated future conditions, and hence the server is always made to contain current information or updated informed action plans, needed in a dynamic environmental situation.

From the diagram in Fig. 11, e-rental, e-auction and various other application services intended to be on the social network (UniCAT), are being equipped with data mining capabilities, before any decisive operation can be accessed or shared by the UniCAT SNS users. In effect, it follows therefore that as many applications as possible, can be integrated into a Social Network using the framework, with the provision of an improved infrastructural facility sophisticated enough to meet the changing needs in the dynamic environment.

VII. CONCLUSION

With the BDIDUS framework provided, and if the implementation process is guided jealously, and made operational, the framework with the modified BDI model should be of tremendous benefits to SNS users by: (i) increasing the SNS capability by allowing users to share/broadcast real time status information or progress with other SNS users. (ii) Widening the knowledge-base and reasoning horizons of the intelligence of the agents, taking into consideration the past historical and current events in planning for alternative solutions to achieving agents’ goals as a result of the data mining. (iii) allowing agents to taking a course of action that best suits the users’ desires or goals based on preference, and (iv) above all, it increases the accuracy of the overall decision making process for SNS users. However, future works should take care of, and spell out the various data mining algorithms needed by this framework for decision-making.

REFERENCES
