The Knowledge Engineering Review, Vol. 36, e12, 1 of 31.
doi:10.1017/S0269888921000114

Agent mining approaches: an ontological view

EMMANUELLE GRISLIN-LE STRUGEON[1,2], KATHIA MARCAL DE OLIVEIRA[1], DORSAF ZEKRI[1,3] and MARIE THILLIEZ[1]

[1]*Univ. Polytechnique Hauts-de-France, LAMIH, CNRS, UMR 8201, F-59313 Valenciennes, France*
[2]*INSA Hauts-de-France, F-59313 Valenciennes, France*
[3]*ReDCAD Laboratory, University of Sfax, B.P. 1173, 3029 Sfax, Tunisia*
E-mails: emmanuelle.grislin@uphf.fr, kathia.oliveira@uphf.fr, dorsaf.zekri2@uphf.fr, marie.thilliez@uphf.fr

Abstract

Introduced as an interdisciplinary area that combines multi-agent systems, data mining and knowledge discovery, agent mining is currently in practice. To develop agent mining applications involves a combination of different approaches (model, architecture, technique and so on) from software agent and data mining (DM) areas. This paper presents an investigation of the approaches used in the agent mining systems by deeply analyzing 121 papers resulting from a systematic literature review. An ontology was defined to capitalize the knowledge collected from this study. The ontology is organized according to seven main facets: the problem addressed, the application domain, the agent-related and the mining-related elements, the models, processes and algorithms. This ontology is aimed at providing support to decisions about agent mining application design.

1 Introduction

Agent mining is a relatively recent area that combines approaches and techniques from the agent and data mining (DM) domains (Cao *et al.*, 2012). Agent systems are made of multiple entities, the *agents*, interacting in a shared environment. Agents are generally characterized by a form of autonomy, which is guided by the pursuit of internal goals. The different terms used, *agent-oriented*, *agent-based* and *multi-agent* systems (MAS), reflect the diversity in integration of agent-related aspects in the software engineering process. At first, the agent can be used only as a useful and abstract concept to model a member of a system, even as an isolated entity (Pulmano & Estuar, 2016) and (Sateli & Witte, 2017). However, in most cases, the interacting aspect is considered and results in the design of communication and coordination elements for the agents to share their common environment made of common resources, etc. The potential coordination issues can be solved in very different ways, from simple interaction rules (Beni, 2005) to sophisticated negotiation protocols (Gorodetsky *et al.*, 2008). Finally, the system design can be achieved using agent dedicated development frameworks (Bellifemine *et al.*, 2001; Yang *et al.*, 2010; Botelho *et al.*, 2020).

DM is part of the data analytics field and also called knowledge discovery in databases. It consists in applying statistical and artificial intelligence (AI) methods to model large data sets (Gorunescu, 2011). More precisely, DM aims at providing prediction and/or description of the data (Pruengkarn *et al.*, 2017). On the basis of the available data, the first aim is to predict which are the more plausible values of some variables; the second aim is to search for characteristics and patterns to describe the data. The AI methods used in this domain are essentially machine learning methods such as classification and clustering algorithms.

When both agents and DM are combined, the distribution and autonomy aspects of the agents can improve the DM process, such as the power of the DM techniques can enhance the agent systems. However, a thorough study of their joint use, how to design it and which methodologies to use, is still lacking.

In a previous study (Grislin-Le Strugeon *et al.*, 2021), a systematic mapping about the combination of agent and mining in the computer science studies from 2000 to 2020 was carried out. The result was a map that quantified the number of studies according to the publication source (journal, conference, workshop), the study type (theoretical, applied) and the research area (agents using DM, DM using agents, integration of agents and mining).

This article proposes to extend this previous study through a deep analysis of the set of articles that integrate both agents and mining in an interdependence way between the two research areas, namely the articles that really include agent mining solutions. The goal is to find out from those articles how the agent mining solutions were designed. More precisely, to find which models, architectures, techniques, etc. are used as approaches to design the systems that integrate agent and mining.

To that end, data related to each of those elements were extracted from the articles (i.e., methods, models, architectures and so on). As result, a lot of knowledge about agent mining approaches was obtained. This paper aims to capitalize this knowledge in some model. Class diagrams, feature models or ontologies could be used for that purpose. Among these models, ontologies offer the capability to express semantics and constraints from the concepts. According to Gruber (1993), 'an ontology is a formal and explicit specification of a shared conceptualisation' that promotes 'a shared and common understanding of some domain that can be communicated across people and computers' (Studer *et al.*, 1998). In this way, the shared knowledge obtained about the agent mining approaches could explicitly be specified and represented. To better take benefit from the result of the systematic mapping, the set of articles was divided in two subsets: one for the definition of the ontology and the other to corroborate it.

The major contributions of this study are:

- the proposition of an ontology that describes the agent mining approaches using seven main facets: the addressed problems, the application domains, the agent-related and mining-related elements, the kinds of used models, processes and algorithms;
- the analysis of 121 articles on agent mining to build the ontology;
- the definition of a reusable taxonomy of application domains related to agent mining solutions.

The remainder of this article is organized as follows. Section 2 starts with a background about the use of ontologies in both agent and DM domains, and a brief summary of the systematic mapping study that is the basis for this paper. Then, Section 3 describes the methodology followed to extract the data from the literature. Section 4 presents the ontology definition and validation. In Section 5, the threats to validity of the study are analyzed, before the conclusion in the last section.

2 Background

To the best of our knowledge, there is no work that organizes the methodological concepts of agent mining using ontologies (or any other representation). However, ontologies have been used in both agent and mining domains, essentially as shared bases among the agents and as semantic supports for mining techniques. Some works present models that include organized concepts to support their development. This section starts by an overview of the use of ontologies in agent and mining systems. The previous study that provides the first result used as a basis for the analysis is then described.

2.1 Ontologies in agent systems

Several works can be found about the use of ontologies in agents systems (Cranefield *et al.*, 2002). For instance, ontologies have been used for a long time to represent the agents' knowledge, see Huhns & Singh (1997) that first proposed to use an ontology as a 'shared virtual world' to help the agents in their communication. A similar principle is present in the MaSE extension proposed in Dileo *et al.* (2002), in the PASSI methodology (Cossentino, 2005) and in the OMNI framework (Dignum *et al.*, 2005), where domain ontologies describe and store the domain knowledge to form a shared basis for the agent interactions. The ontologies are even central in the MOBMAS methodology (Tran & Low, 2008) to support agents' reasoning with a link between the problem-solving knowledge and the domain-related knowledge. The goal is similar in the studies reviewed in Olivares-Alarcos *et al.* (2019) that analyzes the types of cognitive capabilities supported by ontologies for robot autonomy.

Ontologies have been used to support specific phases in the Agent-Oriented Software Engineering (AOSE) domain, which is the part of Software Engineering where the main concept is the software agent (Gómez-Sanz & Fuentes-Fernández, 2015). For instance, the requirements analysis phase of the agent system development approach proposed in Ribino *et al.* (2013) is based on an ontological model to discover inconsistencies and lacks in the problem description. The ontology development phase is central in the MOMA methodology for multi-agent application development (Ying *et al.*, 2013), with three steps of concept identification, ontology modelling and code generation. The ontology in Cordeiro *et al.* (2016) is aimed at providing to the MAS designer a support to select relevant models from AOSE methodologies. The ontology-based approach in Freitas *et al.* (2017) combines model-driven engineering and ontologies for the representation of the agent, environment and organization dimensions of the system, to address the phases of specification, development and verification of agent-oriented software. The Ontology for Autonomous Systems (OASys) (Rodríguez *et al.*, 2012) contains a sub-ontology that defines and organizes concepts for the design of an autonomous system, especially its structure, behaviour and function. These concepts are used in the OASys-based methodology analysis models to develop autonomous systems, like robots for instance (Pignaton de Freitas *et al.*, 2020).

In addition, some of the AOSE methods and methodologies propose concept organizations that are similar to what exist in ontologies. These concept organizations form meta-models for MAS design. For instance, the ADELFE methodology, which is dedicated to cooperative agents, includes a meta-model based on the concepts of skills, aptitudes, cooperation rules, etc. (Bernon *et al.*, 2004). The PASSI methodology (Cossentino, 2005) proposes to use an agent society meta-model to describe the agents' roles, tasks and communication capabilities in UML. The meta-models of the Tropos (Bresciani *et al.*, 2004) and Gaia (Zambonelli *et al.*, 2003) methodologies are oriented to the organizational and planning aspects of the agent system, with the concepts of goal and resource for examples.

These organizations of concepts in meta-models are close to this work, but focused only on the agent system and its components, and not using the rule expressions included in ontologies. Guizzardi & Wagner (2005) looked for organizing concepts of agents using ontologies. The study proposed an initial ontology with some concepts about roles and components of agent systems to support agent-oriented modelling. Although this work is well formalized, no evolution or use of it was found.

2.2 Ontologies and DM

Ontologies have been used in DM domain for different purposes, such as to support information retrieval, data analysis and knowledge discovery.

Regarding the information retrieval, ontologies allow semantic disambiguation with a conceptual representation of the domain. They can focus on the data of interest with precised meaning of text and multimedia data particularly, in relation to the web semantics. Ontologies are used to label data or metadata. In return, mining algorithms allow now the automatic or semi-automatic annotation of documents (Canito *et al.*, 2019; Chen *et al.*, 2020). When used in association to data sources, ontologies can drive the data extraction in linking the source content to a conceptual view, see for instance the survey by Xiao *et al.* (2018). Regarding the data analysis, ontologies can guide association discovery (Wu *et al.*, 2007), feature selection and knowledge reuse (Phillips & Buchanan, 2001). Regarding knowledge discovery, a dedicated ontology can be fully integrated in the mining methodology, like in Gottgtroy (2007). In this study, the ontology engineering process supports every step of the knowledge discovery process. However, the data analyst has still multiple choices to make, essentially regarding the data preparation, the selection of the relevant mining techniques, the parameters of these techniques and their evaluation. To support the decisions that are usually made manually during a DM process, the Data Mining OPtimization Ontology (DMOP) in Keet *et al*. (2015) proposes DM concepts such as DM tasks and DM algorithms. For instance, a DM algorithm is described by its characteristics to specify its use of a random component or its tolerance to class imbalance. However, no study that organizes all of the methodological concepts used in the mining process was found.

2.3 Systematic mapping on agent mining

To perform the systematic study on which the study presented in this paper is based on, an investigation process according to Petersen's work was followed (Petersen *et al*., 2015). The research goal (organized based on Basili *et al*. (1994)) is defined as: **to analyze** agent mining domain **with the purpose of** characterizing **regarding** the used approaches, characteristics and application domain **from the point of view of** software engineering researchers in the context of computer science domain.

Based on this goal, the research protocol was defined, including research questions, search string, inclusion and exclusion criteria, database and tool support as summarized in Table 1. As can be noted, the main research question was how are the agents and DM combined: is mining using agent(s), agent(s) using mining, integration of agent and mining, or otherwise. This protocol was executed twice: one in January 2019 and the second one in July 2020 (including papers after 2019). In total, 3384 were selected from the databases applying the research string, and from that 1634 were analyzed after applying the removal of duplicates and all inclusion and exclusion criteria. Figure 1 presents the distribution of articles in each category. The results of the systematic mapping study can be found in Grislin-Le Strugeon *et al*. (2021).

The focus in this paper concerns the 14% of the studies that present integration of agent and mining. That corresponds to 121 articles, where 96 comes from the first execution of the protocol in 2019 and 25 from the second execution. Therefore, the articles resulting from the first selection were used for the definition of the ontology and the articles from the second selection to corroborate this definition.

3 Exploring agent mining approaches: the methodology

As previously defined, the main goal of this research is to identify which approaches (model, architecture, technique and so on) have been used in the design of agent mining applications. To identify the approaches and to what purpose they are used, a deep reading of the 96 papers selected from the first execution of the research protocol (see previous section) was performed.

To guide the reading of the papers, a data extraction form (Table 2) was elaborated. In this form, inspired on open coding procedure from ground theory methodology (Strauss & Corbin, 1998), codes should be defined to a portion of data extracted from the paper for the identification of the problem being addressed by the paper and the proposed approach.

With the data form and the set of 96 papers, the researchers started the reading and extracting data from the papers individually (Figure 2). Each paper was read and data extracted by one researcher and then the extraction was peer-reviewed. Biweekly meetings were held for discussion among researchers and alignment of interpretations and definitions of codes. After the data extraction from of all the papers, the researchers started another cycle of meetings.

In a first meeting, the researchers read all filled forms to review what was extracted. The papers were grouped, firstly considering the same code approach (e.g. all papers that proposed some model or algorithm). Then, a post-it noted with the name of the main category of approach (in this example, model and algorithm) was set. Then, an analysis of all extracted data about the proposed approaches was performed. The goal was to identify the different kinds of the corresponding approach (in the example, different kinds of models or algorithms) in a way that the approaches could be refined and organized in a hierarchy. When necessary, the researchers came back to the source, re-reading the paper.

In the same way, the researchers repeated these two steps: (1) identifying and grouping by category and (2) refining the category concepts considering the data presented in the papers. At this moment, the

Table 1 Research protocol from Grislin-Le Strugeon *et al.* (2021)

Research questions	RQ1 (main RQ) – How are agents and data mining combined? In other words, is mining using agents, are agents using mining, are agents and mining integrated, or otherwise? RQ2 – What agent abilities are explored in agent mining? RQ3 – What data mining approaches are used in agent mining? RQ4 – To which fields of application does agent mining apply?
Search string	('agent' OR 'multiagent') AND ('mining' OR 'knowledge discovery' OR 'KDD').
Databases	Scopus, Web of Science
Inclusion Criteria	IC1: Articles published in the Computer Science domain IC2: Articles published after December 1999 and before January 2018 IC3: The agent term represents an agent in the Computer Science meaning, i.e., a software or a robotic agent IC4: The mining term represents data mining in the Computer Science meaning
Exclusion Criteria	EC1: Articles not written in English EC2: Articles which are not journal articles, conference papers or workshop papers EC3: Duplicate articles indexed differently EC4: Articles not peer-reviewed EC5: Articles where either agent or data mining is not included, or appears only as an example of technique EC6: Articles that concern agent in another meaning than in Computer Science EC7: Articles that concern mining in another context EC8: Books, book chapters that are not from conference proceedings, book prefaces, summaries of conferences, editorials, retracted articles EC9: Articles presenting partial results of a complete study that is selected
Acceptance Criteria	Three distinct readers: – all readers accept => paper is accepted – all readers exclude => paper is excluded – the majority of accept, others in doubt => paper is accepted else: discuss and consensus
Tool	JabRef (http://www.jabref.org)

focus was on the following parts of the extraction form: the application domains, the problem and its coding, the approach and its coding (see Table 2).

4 An ontology of agent mining approaches

There are several methodologies to design an ontology (e.g. (Grüninger & Fox, 1995; Noy & McGuinness, 2001; Lopez *et al.*, 1999). All of them consider basically the following steps: definition of the ontology *purpose*, *conceptualization*, *formalization* and *validation*. The ontology defined in this paper followed in general those steps.

The *purpose* of this ontology is based on our research goal, that is, to establish a common conceptualization regarding the approaches used in the design of agent mining applications based on the analysis

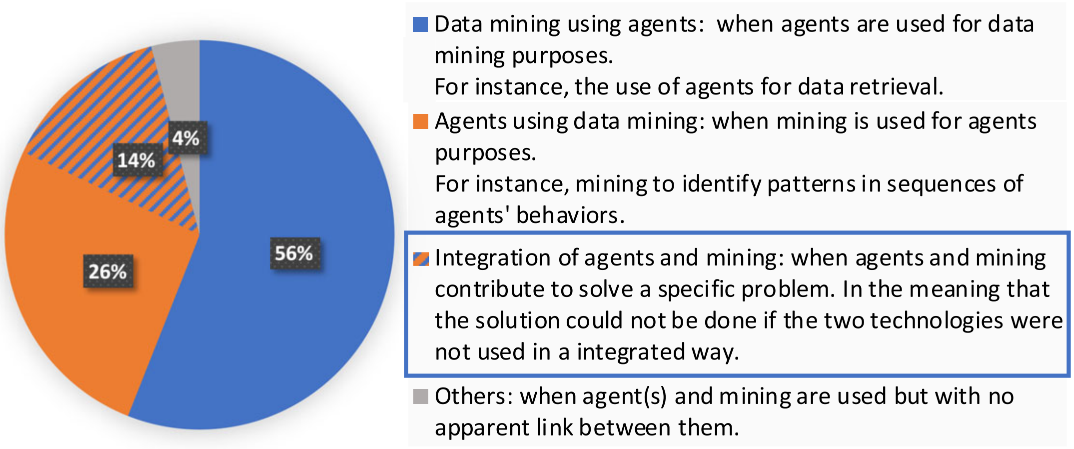

Figure 1 Answers for the main research question: How are agents and data mining combined? (adapted from Grislin-Le Strugeon *et al.* (2021)).

of literature (set of 96 papers) that present agent mining solutions. It is important to highlight that the resulting ontology represents the knowledge acquired from those papers.

The *conceptualization* represents the knowledge modelling itself focusing on the definition of concepts, relations and constraints (axioms). Considering the ontology purpose, all data extracted from the papers and recommendations from ontology methodologies (Grüninger & Fox, 1995; Noy & McGuinness, 2001). The ontology is organized according to Competency Questions (i.e., requirements in the form of questions that the ontology must answer), as follows:

- CQ1 – How are the agent mining approaches characterized?
- CQ2 – What kind of problem has agent mining been applied to?
- CQ3 – For which application domains agent mining has been used?
- CQ4 – Which are the approaches (model, architecture, technique and so on) used in agent mining?

Note that these questions refer strictly to the paper set examined in this study. For instance, the Competency Question 1 (CQ1) must be understood as: How are the agent mining approaches characterized in the papers included in the study set? This paper does not pretend to present an exhaustive representation of the agent mining studies. However, given its size, the studied set is likely to provide some knowledge of interest for the agent mining designers. Moreover, the competency questions are aligned with the information required in the data form (see Table 2) and vice versa to ensure that the ontology can be defined: CQ1 will be answered with the information from (d), (e) and (i); CQ2 with the information from (g) and (h); CQ3 from (f) and CQ4 from (i) and (j).

For the *formalization*, ontology editor and framework Protégé v5.5[1] (Musen & The Protégé Team, 2015), with OWL were used. Finally, for *validation*, a data-driven evaluation technique (Brank *et al.*, 2006) was performed by instantiating the ontology with real instances collected in the analyzed papers and new ones from the mid-2020 collection.

4.1 *Conceptualization and formalization*

In the following section, the ontology conceptualization and formalization are presented by answering each of the four competency questions on the basis of the publication set.

4.1.1 *CQ1 - How are the agent mining approaches characterized?*

The basic definitions of agent mining were considered to answer the first Competency Question (**CQ1 – How are the agent mining approaches characterized?**). Although agent mining is not a well-defined

[1] https://protege.stanford.edu/

Table 2 Extraction form

What	Explanations
(a) Paper identification	Paper identification
(b) Title	Paper title
(c) Reject?	Insert a 'X' and give the reason if you think the paper must be removed from the set. Reasons for rejecting: No approach: the document does not include any approach (method, model, architecture. . .); Redundancy: there is a similar and more complete document from the same author(s) in the set (give its identification)
(d) One agent or several agents	Information from the Syst Mapping study. If you want to add/modify it, add a line starting with your initials.
(e) Mining or KDD approach(es)	Information from the Syst Mapping study. If you want to add/modify it, add a line starting with your initials.
(f) Application domain(s)	Information from the Syst Mapping study. Precise it when possible. If you want to add/modify it, add a line starting with your initials.
(g) Problem	Short description of the applied and/or scientific problem(s), proven by text extracts (add figure when possible). The text extracts must be enclosed in quotations marks. If multiple problems, list them from the most general to the more precise one. Ex.: information retrieval to assist learners, the information is retrieved from expert social networks: 'The problem is to retrieve relevant information from social networks that provide answers to such questions. . ..'
(h) Kind of problem (coding)	Coding the problem(s) to include it or them in a wider thematic. Examples are provided in the SLR-coding-lists.xls file (not exhaustive). Ex. (related to the previous ex.): information retrieval on social networks
(i) Approach for the design of the agent mining system	Short description of the approach (architecture, method, model, framework, . . .), proven by a text extract (add figure when possible). The text extract must be enclosed in quotations marks. Ex.: a two-phase knowledge discovery process, agents retrieve the data that are then processed by a classification module: 'The process includes two phases, . . .'
(j) Approach (coding)	Coding the approach to relate it to similar approaches. Examples are provided in the SLR-coding-lists.xls file (not exhaustive). Ex. (related to the previous ex.): Knowledge discovery process

research topic, the researchers consider that a study belongs to the agent mining area if it includes both agent and mining parts. As explained in Cao *et al.* (2007), there are three types of agent-mining interaction: (i) when agent(s) uses DM solutions, (ii) when DM uses agent(s) for instance to improve parts of the mining process or (iii) when agent and mining solutions are integrated, both contributing to a common purpose. In this paper, the focus is on the third type. Therefore, any `AgentMiningApproach` is composed of two main kinds of elements (Figure 3[2]): *agent-related components* and *mining-related elements*. Moreover, any agent mining approach includes at least two elements, one of each kind.

Several object properties were set to represent the connection between classes. Part-whole relations were formalized according to W3C (11 Aug 2005). All concepts of the ontology were defined in a

[2] In this paper, OntoGraf was used for all figures of the ontology.

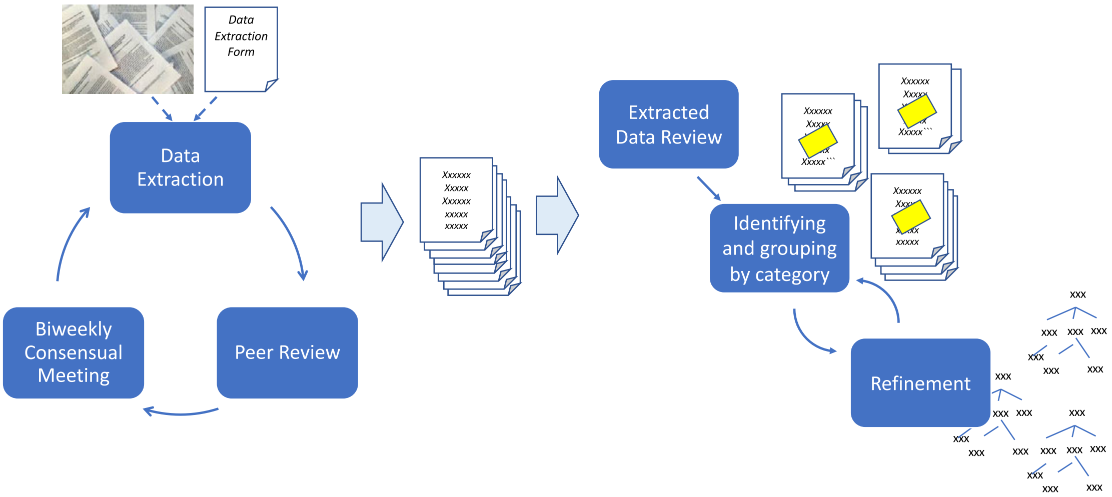

Figure 2 Qualitative analysis process.

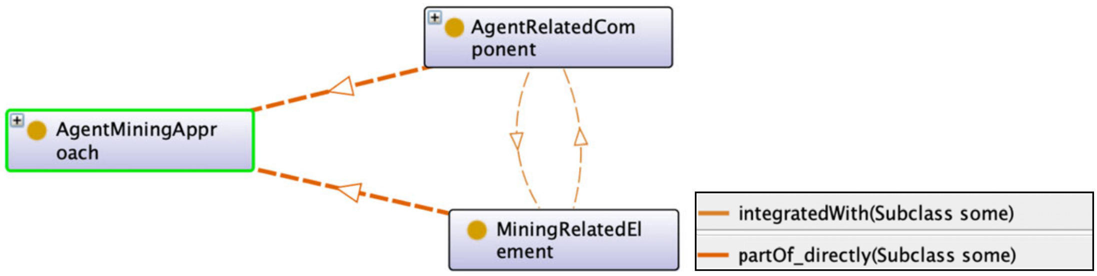

Figure 3 Elements of the agent mining approaches.

glossary based on the analyzed papers. The following (software-specific) first-order logic[3] expressions[4] were defined:

- `AgentMiningApproach` hasPart only (`MiningRelatedElement` or `AgentRelatedComponent`)
- `AgentMiningApproach` hasPart some (`MiningRelatedElement` and `AgentRelatedComponent`)
- `MiningRelatedElement` PartOf_directly some `AgentMiningApproach`
- `AgentRelatedComponent` PartOf_directly some `AgentMiningApproach`
- `AgentRelatedComponent` integratedWith some `MiningRelatedElement`
- `MiningRelatedElement` integratedWith some `AgentRelatedComponent`

Note that :

- the word `only` is equivalent to the universal quantifier, denoted by ∀. It indicates that any instance of the class, `AgentMiningApproach`, has a relationship along the specified property `hasPart` only to instances that are members of the class `AgentRelatedComponent` or `MiningRelatedElement`.
- the word `some` here is equivalent to the existential restriction, denoted by ∃. For example, it indicates that any instance of the class `AgentRelatedComponent` participates in at least one relationship along the specified object property `integratedWith` to instances that are members of the class `MiningRelatedElement`.

[3] This paper uses the same format found in Anand *et al*. (2012), Oliveira *et al*. (2013) to present the axioms.
[4] Each object property has its inverse object property defined, even if not shown in the figures. For instance, partOf_directly has its inverse named hasPart_direcly; refersTo has its inverse isReferedBy

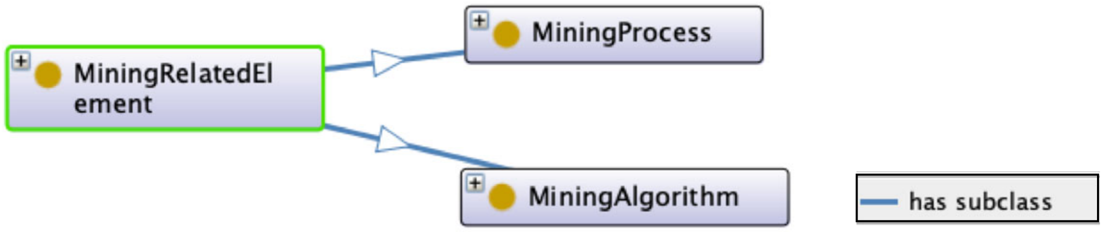

Figure 4 Concepts of the mining-related elements[5].

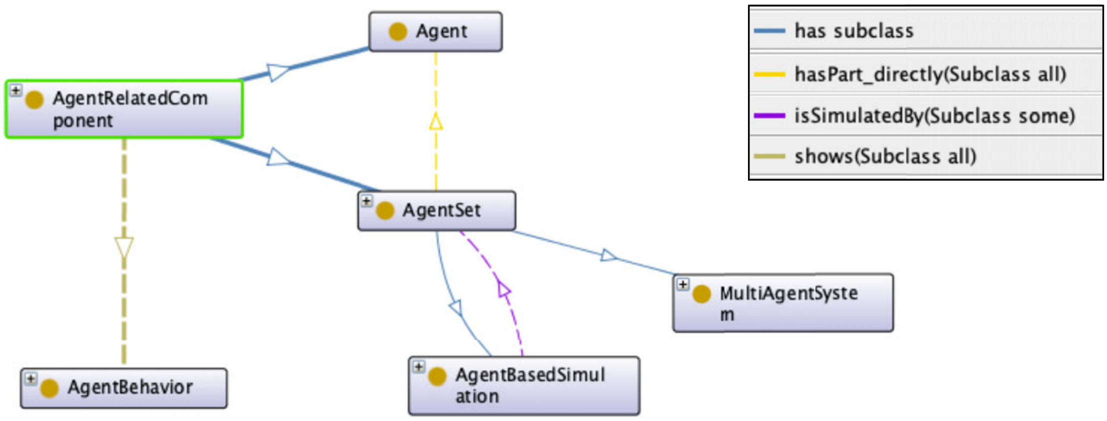

Figure 5 Concepts of the agent-related components.

The `MiningRelatedElement` and `AgentRelatedComponent` differ in nature. Indeed, `MiningRelatedElement` is defined by its functional aspect, that is, the description of a process or an algorithm related to the mining function, whereas `AgentRelatedComponent` is fundamentally defined by its compositional aspect, that is, the presence of one or several agents.

Every `MiningRelatedElement` is either a `MiningProcess` or a `MiningAlgorithm`, see Figure 4. A `MiningProcess` is a process that contributes to the discovery of data. A `MiningAlgorithm` is an algorithm that contributes to the discovery of data, generally using a more detailed or formal expression than a process. These concepts are detailed in Section 4.1.4.

Every `AgentRelatedComponent` is either a single agent or a set of agents (see Figure 5), where `Agent` is a software component of the system that shows a certain autonomy in its behaviour (`AgentBehavior`). Several works on agent mining are developed using only one `Agent`. For instance, in Pulmano & Estuar (2016) one agent performs continuous learning to provide diagnostic results in a medical care service. In another work, the single agent in Sateli & Witte (2017), called 'semantic research agent', learns the user's topics of interest and formulates queries on databases. However, a large majority of the works includes two or more agents composing an `AgentSet` (86 of 96 papers). In these papers (such as Moemeng *et al*. (2008), Gerardo & Lee (2009), Gorodetsky *et al*. (2005), Pinzón *et al*. (2013) Saadi & Chaoui (2017) and Garcia-Sanchez *et al*. (2008)), multiple agents interact and contribute to the DM. Several agents grouped in an `AgentSet` can perform a simulation (`AgentBasedSimulation`). In this case, the agents are used to simulate the dynamics of a distributed system, generally to produce emergent states that are difficult to anticipate by other methods due to the interaction between interdependent entities. `AgentBasedSimulation` can thus be used as an *agent mining model*, described in Section 4.1.4, to produce predictions (De Arruda *et al*. 2017; Kaschesky *et al*. 2011).

[5] Note, in all figures about the ontology 'has subclass' relationship corresponds to the inverse of the 'is-a' relationship for taxonomies in ontology. For instance, in this figure the relationship means '`MiningProcess` is-a `MiningRelatedElement`'

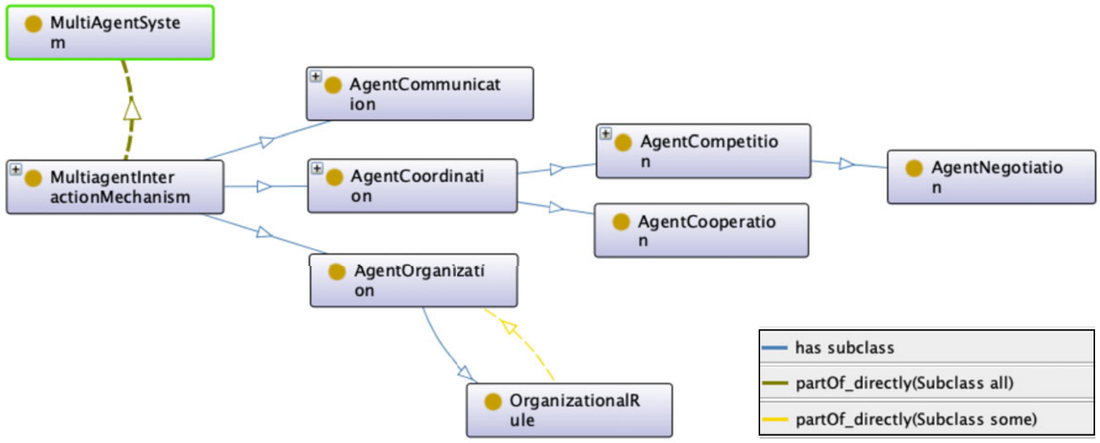

Figure 6 Concepts to define multi-agent interaction mechanisms in agent mining approaches.

Several object properties were set to represent the connection between those defined classes (Figure 5), as follows:

- `AgentRelatedComponent` shows only `AgentBehavior`
- `AgentSet` hasPart min 2 `Agent`
- `AgentBasedSimulation` isSimulated by only `AgentSet`
- `AgentBasedSimulation` isSimulated by some `AgentSet`

When the interaction is structured according to an interaction mechanism, the agent set is often called a MAS (`MultiAgentSystem`). For example, in Gorodetsky (2013) a multi-agent architecture is proposed for anomaly detection, including two agents: Agent of network traffic-based Alert (NTA-agent) and Alert Correlation agent (AC-agent). Another example of `MultiAgentSystem` is in Ioniţă & Ioniţă (2013) where several agents contribute to the DM process (IDS Data Gathering Agent, Classification Agent, Decision/Action Agent, Alarm Generation Agent, Configuration Agent). The interaction method can be specified according to defined *communication* (`AgentCommunication`), *coordination* (`AgentCoordination`) and/or *organization* (`AgentOrganisation`) structures (see Figure 6).

When specified, the `AgentCommunication` is often based on message passing, where agents communicate by sending individual messages to each other. This can be formalized by a define protocol (`CommunicationProtocol`), as shown, for instance, in Sinnappan *et al.* (2001), Ponte *et al.* (2015) and Liu *et al.* (2009)). The papers present two kinds of `AgentCoordination` as follows:

- *competition* (`AgentCompetition`), when the agents have distinct goals and compete for resources. In some of the competition cases, the agents negotiate to make joint decisions, while they pursue different goals. One example of `AgentNegotiation` is given in Gorodetsky *et al.* (2008) where seven agents communicate asynchronously by message passing according to a protocol and apply a strategic negotiation.
- *cooperation* (`AgentCooperation`), when the agents contribute to a common goal. For example in Ghedini Ralha & Sarmento Silva (2012), Haider *et al.* (2009) and Kotova (2017), the global task is divided into subtasks to be distributed among the agents, taking into account the resources available and the skills of the agents.

Note that the specific mechanism or technique is not always detailed in the paper. In such cases, instead of adding a ‘miscellaneous’ category, it was decided that the higher level category should be used (e.g. `AgentCoordination` when the exact kind of coordination is not known).

Lee & Park (2005), Pinzón *et al.* (2013), Santos Jr. & Johnson Jr. (2004), Xue *et al.* (2009), Bajo *et al.* (2016) and Gago *et al.* (2005) propose an `AgentOrganisation` defined by several `Rules`. The

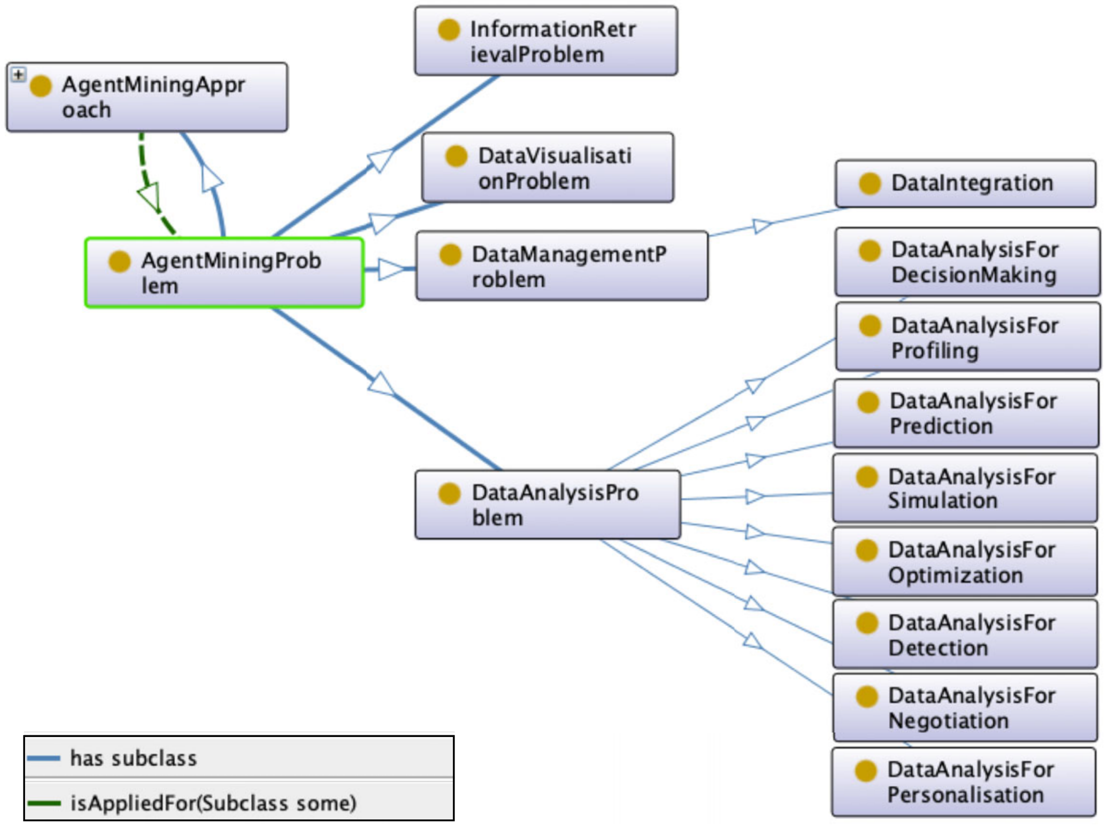

Figure 7 Taxonomy of problems addressed by agent mining approaches.

organizational rules define constraints on agent behaviour and their interactions. At the analysis level, this equates to restrictions on the roles an agent may play or how an agent may interact with other agents. For example, in Gago *et al*. (2005) inference rules called 'bridge rules' defining the interaction among the system components (the agents) describe the agents' reactions to events occurring in their environment. Several object properties were set to represent the connection between these defined classes (Figure 6), as follows:

- `MultiAgentInteractionMechanism` partOf_directly some `MultiAgentSystem`
- `OrganisationalRule` partOf_directly some `AgentOrganisation`

4.1.2 CQ2 – What kind of problem has agent mining been applied to?

To answer the second Competency Question (**CQ2 – What kind of problem has agent mining been applied to?**), the problem description and its coding (see Table 2 – (g) and (h)) were analyzed for all papers, following the procedure presented in Figure 2 right side. As a result, a taxonomy of *problem domain* (`AgentMiningProblem`) was defined in Figure 7. Four main kinds of problem were identified: `DataVisualisationProblem`, `InformationRetrievalProblem`, `DataManagementProblem` and `DataAnalysisProblem`. `DataIntegration` was defined as a specific kind of `DataManagementProblem`.

Moreover, `DataAnalysisProblem` is applied for different purposes, as follows:

- *for profiling* (`DataAnalysisForProfiling`)—to record and analyze behavioural characteristics, so as to assess or predict their capabilities in a certain sphere, or to assist in identifying categories of people (see, for instance, Mgbemena & Bell (2016), Matei *et al*. (2016) and Lee & Park (2005)),
- *for decision making* (`DataAnalysisForDecisionMaking`)—to identify and choose alternatives based on the values and preferences. Every decision-making process produces a final choice, which may prompt action (e.g. Islam (2007), Kularbphettong *et al*. (2012) and Kwon *et al*. (2004)),

- *for simulation* (DataAnalysisForSimulation)—to test, in a modelling environment, scenarios with varying starting conditions assess and characterize their likely outcome (e.g. Izumi *et al.* (2007), Jayyousi & Reynolds (2014) and Li & Lin (2012)),
- *for prediction* (DataAnalysisForPrediction)—to extract information from data and use it to predict trends and behaviour patterns (e.g. Haider *et al.* (2009), Kwon *et al.* (2004) and Pulmano & Estuar (2016)),
- *for optimization* (DataAnalysisForOptimization) – to take the most efficient decisions by way of determining a satisfactory solution (e.g. Chen & Chen (2017), Eugin Lilly & Venkataraman (2013) and Liu *et al.* (2009)),
- *for detection* (DataAnalysisForDetection)—to identify and filter noises, frauds and anomalies (e.g. Kusumura *et al.* (2004), Esfandi (2010) and Dahl *et al.* (2010)),
- *for negotiation* (DataAnalysisForNegotiation)—to generate future arguments, to evaluate alternative arguments and to improve process of argumentation using contextual information (e.g. Islam (2007)),
- *for personalization* (DataAnalysisForPersonalisation)—to adapt the output of a system to a specific user, her context and profile (e.g. Cobos *et al.* (2007), Loia *et al.* (2007) and Oskouei & Moradi-Kor (2016)).

The following properties apply to the problem-related concepts:

- AgentMiningApproach isAppliedFor only AgentMiningProblem
- AgentMiningApproach isAppliedFor some AgentMiningProblem

4.1.3 CQ3 – For which application domains agent mining has been used?

The same procedure was followed to answer the third Competency Question (**CQ3 – For which application domains agent mining has been used?**) resulting in a taxonomy of *application domains*. Indeed, even if few papers are purely theoretical and if application system or result is not always presented, almost all of the papers propose at least one type of application for the proposed approach.

The AgentMiningApplicationDomain represents the domain the approach is meant for. There are previous taxonomies of application domains, for example, the taxonomy by Kotonya *et al.* (2003). The taxonomy presented here is specific to the agent mining area and non-exhaustive: it includes all and only the domains found in the studied set of documents (Figure 8).

SoftwareEngineering is the most represented domain. Indeed, when an approach was proposed to improve a DM method or a new system architecture, it was classified in the SoftwareEngineering domain, subclass of the ComputerScienceDomain. Otherwise, the most represented domains are the Education, the ISSecurity and the Marketing domains. The Education, classified in the HumanitiesDomain, appears mainly in studies where the approach is applied to e-Learning and tutoring systems as in Oskouei & Moradi-Kor (2016) and Rosé & Ferschke (2016). The ISSecurity (Information System Security), which combines elements from the ComputerScienceDomain and the Security domain, includes, for example, all the studies where the approach is applied to intrusion detection (Silva & Ralha, 2011). The Marketing domain, subclass of the CommercialDomain, includes for instance the approaches used to increase the sales on the basis of customer profiling (Kularbphettong *et al.*, 2012).

Most applications domains are disjoint but some concepts combine different elements, such as the ISSecurity previously cited and the SocialNetworks defined as subclass of both the DistributedSystems to represent its technical aspects and the HumanitiesDomains to represent its social aspects. In addition, several papers mention more than one application domain. For example, Lee & Park (2005) combine ComputerScienceDomain and CommercialDomain.

The following properties apply to the application domain concepts:

- AgentMiningApproach isAppliedIn only AgentMiningApplicationDomain
- AgentMiningApproach isAppliedIn some AgentMiningApplicationDomain

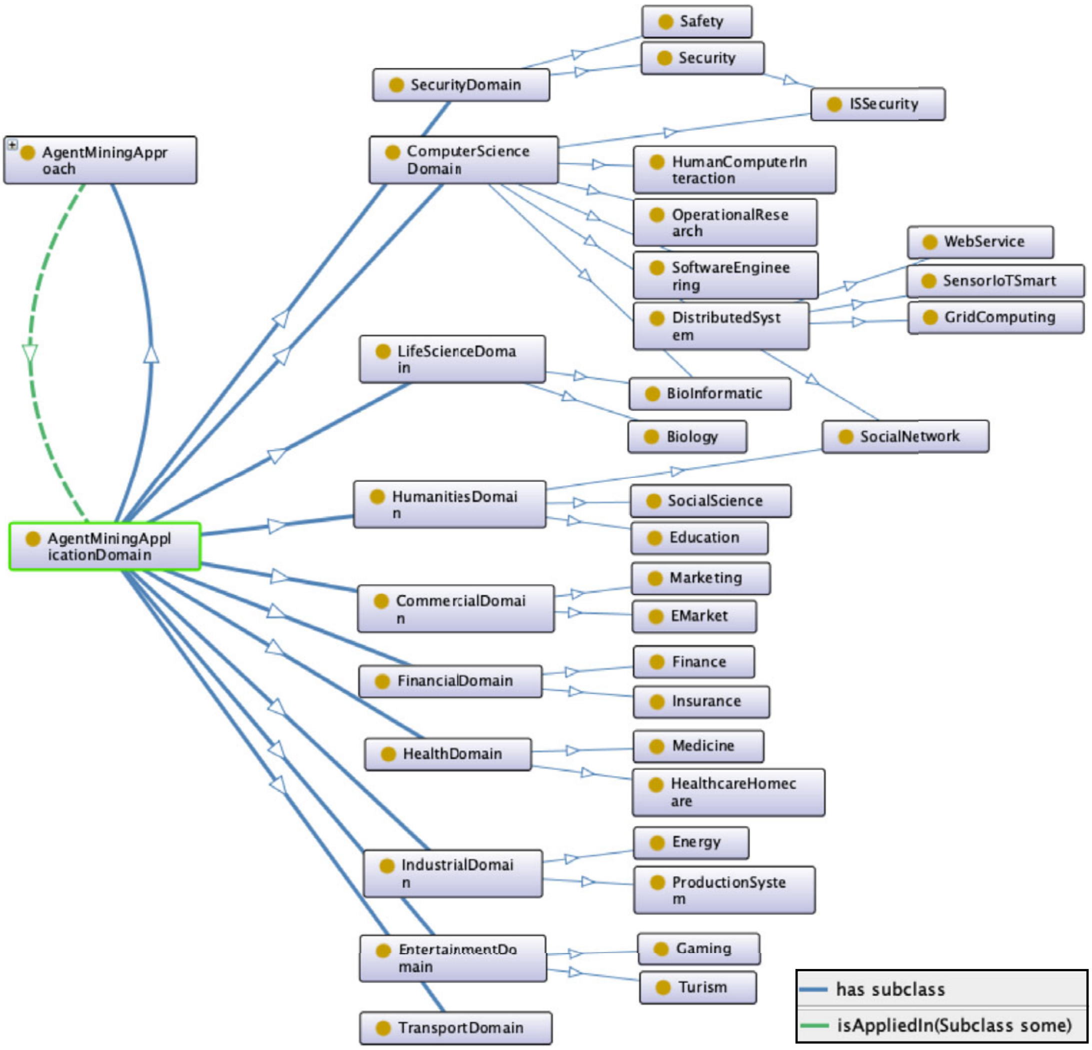

Figure 8 Taxonomy of agent mining application domains.

4.1.4 CQ4 – Which are the approaches (model, architecture, technique and so on) used in agent mining?

Finally, to answer the fourth Competency Question (**CQ4 – Which are the approaches (model, architecture, technique and so on) used in agent mining?**) the data extracted from the papers were analyzed and organized first in main categories and then refined in a specific taxonomy. Three main concepts that support the description of the approaches (Figure 9): `AgentMiningProcess`, `AgentMiningModel` and `AgentMiningAlgorithm` were identified. For each of these concepts, a specific taxonomy was defined.

The following properties apply to describe the approach concepts:

- `AgentMiningApproach` follows only `AgentMiningProcess`
- `AgentMiningApproach` follows some `AgentMiningProcess`
- `AgentMiningApproach` isRepresentedBy only `AgentMiningModel`
- `AgentMiningApproach` isRepresentedBy some `AgentMiningModel`
- `AgentMiningApproach` applies only `AgentMiningAlgorithm`
- `AgentMiningApproach` applies some `AgentMiningAlgorithm`

`AgentMiningProcess`. A `AgentMiningProcess` is an orderly set of steps to follow to achieve the mining operation. Figure 10 presents the different kinds of *processes* identified, according to the element(s) concerned by the process:

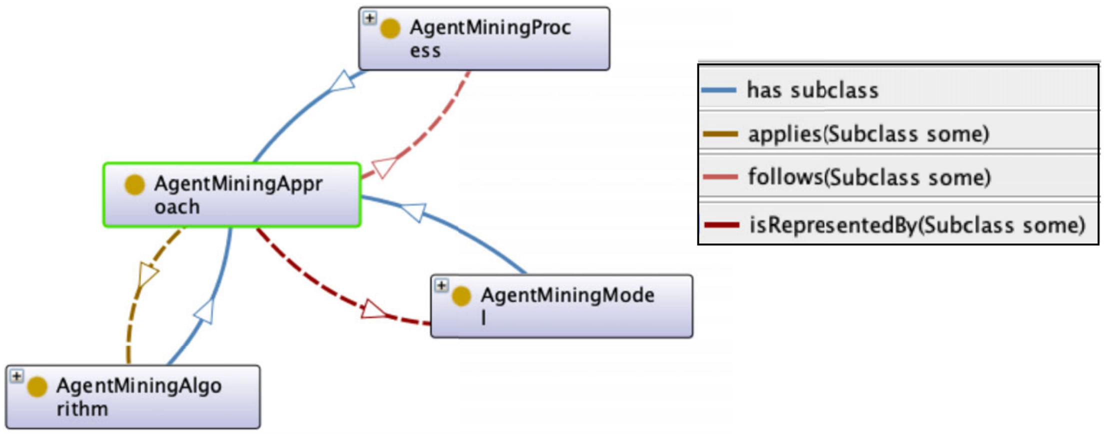

Figure 9 Main concepts to describe the agent mining approaches.

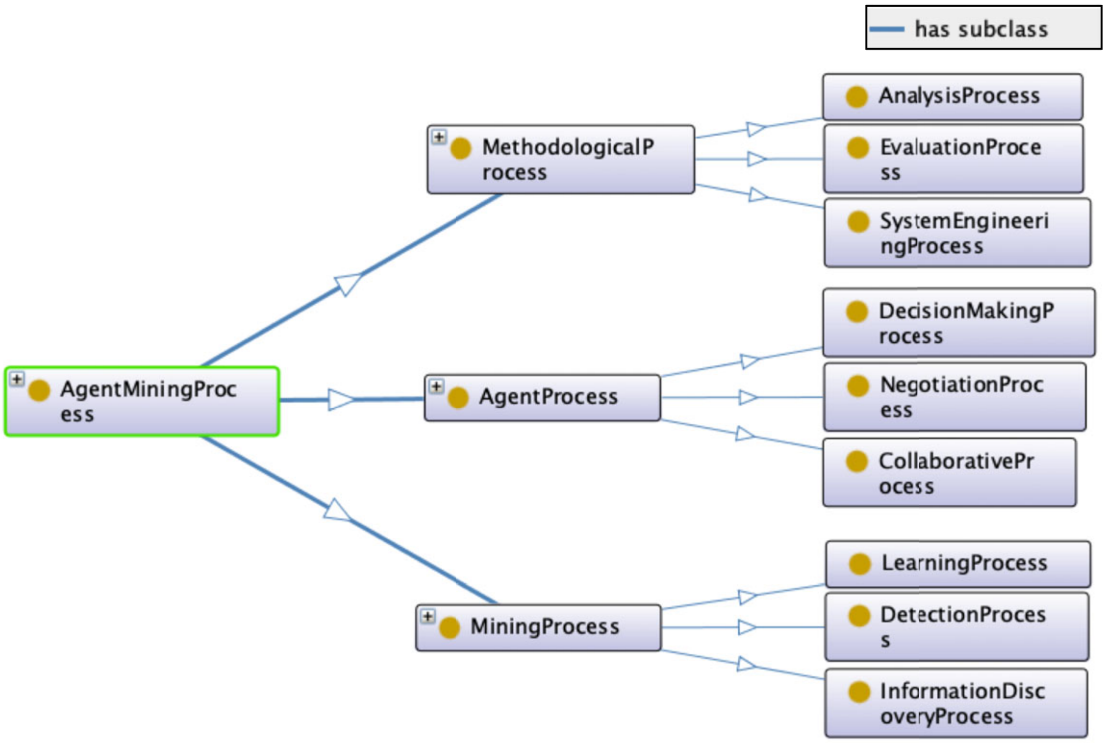

Figure 10 AgentMiningProcess taxonomy.

- *methodological* (`MethodologicalProcess`) can represent a methodology either to perform the data processing (`AnalysisProcess`) (Mgbemena & Bell, 2016)), to evaluate the performance of the system (`EvaluationProces`) (Dimou *et al.*, 2007) or to design the agent system (`SystemEngineeringProcess`) (Cobos *et al.*, 2007; Liu *et al.*, 2009);
- *process related to agents* (`AgentProcess`) describing the steps conducted by agents, either for *decision making* (`DecisionMakingProcess`) (Wang *et al.*, 2012; Liu *et al.*, 2009) *negotiation* (`NegotiationProcess`) (Garcia-Sanchez *et al.*, 2008; Dos Santos & Bazzan, 2005) and/or *collaboration* (`CollaborativeProcesses`) (Shih *et al.*, 2008; Pedrycz & Rai, 2008), (Pi *et al.*, 2005);
- *process for the mining* (`MiningProcess`) describing the procedures for mining, mainly *learning* (`LearningProcess`) (Kadhim *et al.*, 2016; Duarte & Julia, 2017), *detection* (`DetectionProcess`), *information discovery* (`InformationDiscovery`) (Aridor *et al.*, 2002; Hamdi, 2007).

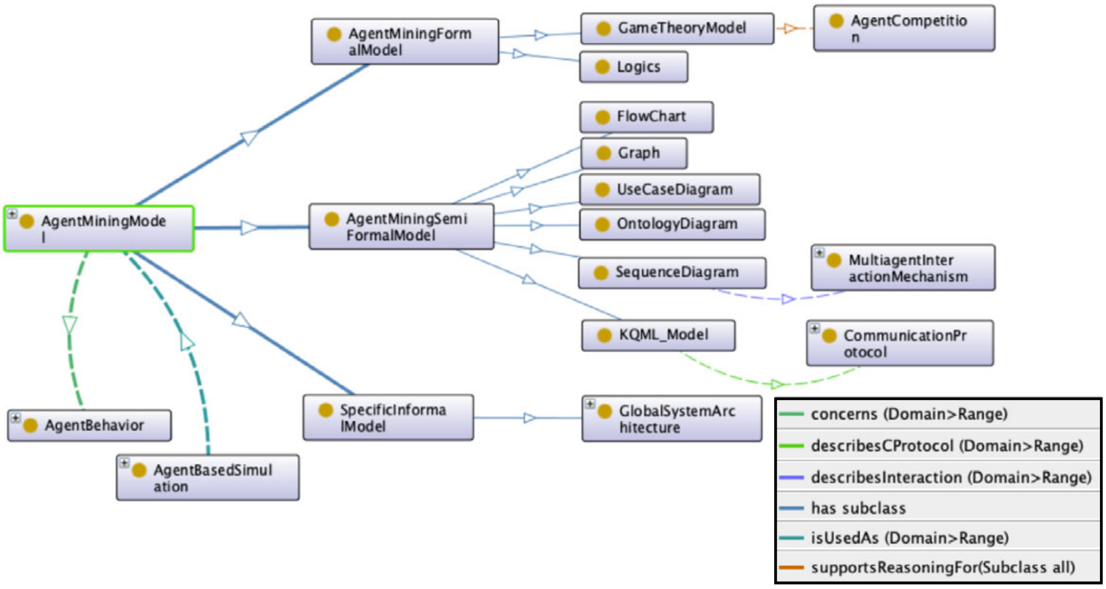

Figure 11 AgentMiningModel taxonomy.

The following properties apply to the process concept:

- `AgentProcess` describesBehaviorOf only `AgentRelatedComponent`
- `MethodologicalProcess` refersTo some (`AgentRelatedComponent` or `MiningRelatedElement`)
- `MethodologicalProcess` disjoint with[6] `AgentProcess` and `MiningProcess`

`AgentMiningModel`. Figure 11 presents the taxonomy of *models* with the `AgentMiningModel` concept. A `AgentMiningModel` can be a representation of a `AgentMiningElement`, an object used by a `AgentMiningElement` or an interaction between several `AgentMiningElement`. The models are distinguished according to their form, more or less formal.

Formal models (`AgentMiningFormalModel`) are mostly represented using `Logics` (e.g. Wang *et al.* (2012), Dahl *et al.* (2010)); sometimes in the context of `GameTheory` (e.g. Abdelrahman *et al.* (2011), Liu *et al.* (2017)), used by agent systems to perform reasoning. `GameTheory` supports especially agent competitions.

Semi-formal models (`AgentMiningSemiFormalModel`) often use UML diagrams, mostly `UseCaseDiagram` (e.g. Samoylov & Gorodetsky (2005), Gerardo & Lee (2009)) and `SequenceDiagram` (e.g. Xue *et al.* (2009), Dos Santos & Bazzan (2005)). Among the semi-formal models, `Graph` (e.g. Thanawala & Joshi (2017), Šlapák & Neruda (2014)), `Flowchart` (e.g. Sadhasivan & Balasubramanian (2017); Moemeng *et al.* (2012)) and `OntologyDiagram` (e.g. Sateli & Witte (2017), Gorodetsky (2013)) can be also found. Agent communication can be described using the well-known language *KQML* and also included as a semi-formal model (`KQML_Model`) (e.g. Jiang *et al.* (2007), Yang *et al.* (2007)).

The less formal models are `SpecificInformalModel` proposed by the authors with specific representations. For instance, some papers present a generic illustration, here named `GlobalSystemArchitecture`, where all elements of the approach (agents and mining elements) are organized to show how they are associated or how they interact (see, for instance, Silva & Ralha (2011), Pultar *et al.* (2008), Fowler & Hammell (2015)). This illustration is represented by a subclass of the `SpecificInformalModel`.

Some of the `AgentMiningModel` represent agent(s) behaviour (`AgentBehavior`). The `AgentBehavior` defines the behaviour of one or several agents: the activities performed and sometimes

[6] Note that Disjoint With is equivalent to the ⊕ operator in formal logic.

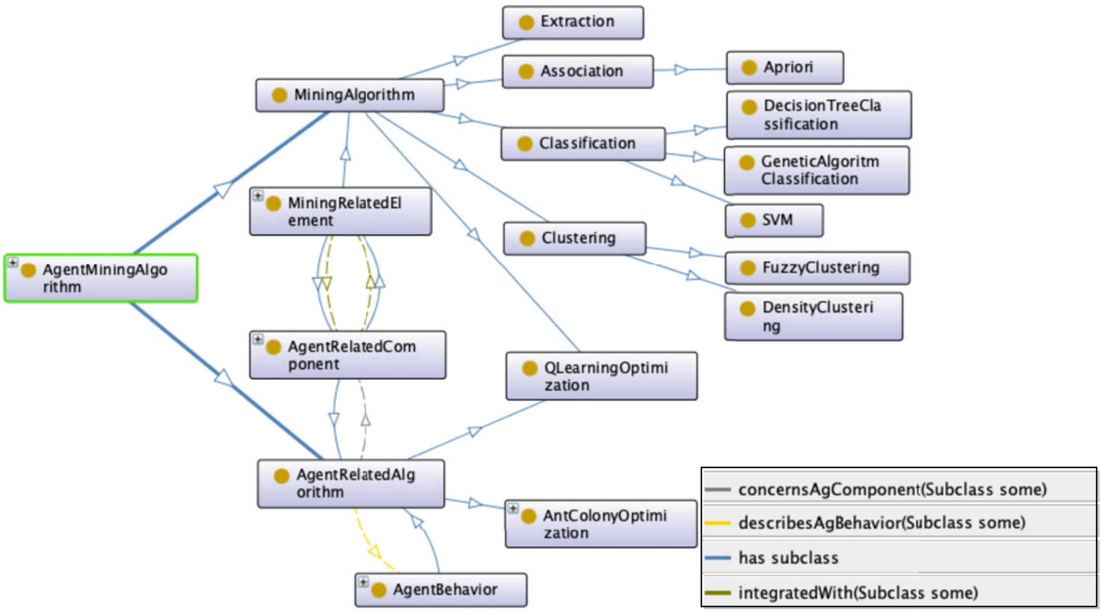

Figure 12 AgentMiningAlgorithm taxonomy.

the activity sequence or life cycle of the agent(s). The `AgentBehavior` is explained in Liu *et al.* (2017), Pedrycz & Rai (2008) and Santos Jr. & Johnson Jr. (2004) and can also be more precisely described by an algorithm as in Menczer (2003).

The following properties apply to the model concept:

- `AgentMiningModel` concerns only `AgentBehavior`
- `GameTheoryModel` supportsReasoningFor only `AgentCompetition`
- `GlobalSystemArchitecture` organizes some (`AgentRelatedComponent` and `MiningRelatedElement`)
- `KQML_Model` describesCProtocol exactly[7] 1 `CommunicationProtocol`
- `CommunicationProtocol` formalizes exactly 1 `AgentCommunication`
- `AgentMiningFormalModel` disjoint with `AgentMiningSemiFormalModel` and `SpecificInformalModel`
- `AgentBasedSimulation` isUsedAs min 1 `AgentMiningModel`

`AgentMiningAlgorithm`. An `AgentMiningAlgorithm` is an orderly set of actions that must be performed to achieve one or several mining steps. Figure 12 presents the taxonomy of `AgentMiningAlgorithm` that can be of two kinds according to the concerned `AgentMiningElement`, as in the process taxonomy.

The first one represents the *algorithms related to the mining parts* (`MiningAlgorithm`) (e.g. Fowler & Hammell (2015), Ioniţă & Ioniţă (2013)). These algorithms were classified according to their objective: `Extraction` (i.e. access and retrieval of data) (e.g. De Arruda *et al.* (2017), Chen & Plale (2015)), `Classification`, `Clustering` and `Association`. Three well-known types of algorithms are mostly used for the *classification* objective: `SVM` (Pinzón *et al.*, 2013), `DecisionTree` (Oskouei & Moradi-Kor, 2016) and `GeneticAlgorithms` (Gaya & Giráldez, 2011). *Fuzzy techniques* (`FuzzyClustering`) (Loia *et al.*, 2007) and *density-based clustering* (`DensityClustering`) (Saadi & Chaoui, 2017) are mostly

[7] Note that the word *exactly* represents a cardinality restriction. It specifies the exact number that an individual must participate in for a given property. For example, an instance of the class `KQML_Model` participates in one relationship along the specified object property `describesCProtocol` to exactly one instance of the class `CommunicationProtocol`. In other words, ‘a `KQML_Model` describes exactly one `CommunicationProtocol`’.

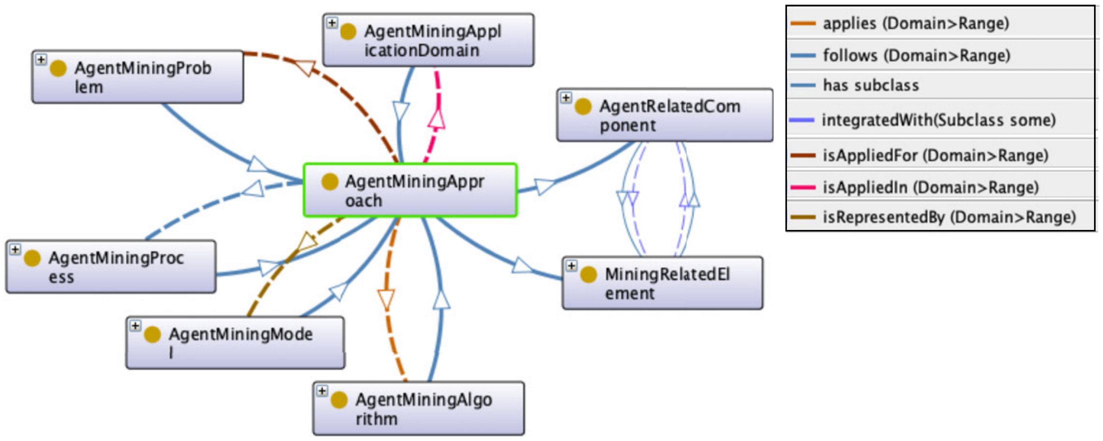

Figure 13 General view of the ontology on agent mining approaches.

used for the clustering objective. When detailed, the `APriori` algorithm is generally the *association* algorithm that is mentioned (e.g. Wang *et al*. (2012), Mateo *et al*. (2008)).

The second kind of *agent mining algorithm* concerns the agents (`AgentRelatedAlgorithm`), indeed they specifically describe agent behaviours (e.g. Jayyousi & Reynolds (2014), Qiu *et al*. (2009)). The *ant colony* (`AntColonyOptimization`) algorithms (Xu & Yan, 2009) and the *Q-Learning* (`QLearningOptimization`) algorithms (Xu *et al*., 2012) are two kinds of *agent mining algorithms* that are used for optimization. The latter appears both as a `MiningAlgorithm` and an `AgentRelatedAlgorithm`.

The following properties apply to the `AgentRelatedAlgorithm` concept:

- `AgentRelatedAlgorithm` concernsAgComponent some `AgentRelatedComponent`
- `AgentRelatedAlgorithm` concernsAgComponent only `AgentRelatedComponent`
- `AgentRelatedAlgorithm` describesAgBehavior some `AgentBehavior`
- `AgentRelatedAlgorithm` describesAgBehavior only `AgentBehavior`
- `AntColonyOptimization` isPerformedBy some `AgentSet`
- `AntColonyOptimization` isPerformedBy only `AgentSet`

4.1.5 General view of the ontology

Finally, a general view of the ontology is presented in Figure 13 and described as follows:

- the problem that is tackled by the approach (`AgentMiningApproach`);
- the domain(s) of application for the approach (`AgentMiningApplicationDomain`);
- the approach elements: agent-related components and mining-related elements (`AgentRelatedComponent` and `MiningRelatedElement`);
- the model(s) (`AgentMiningModel`), process(es) (`AgentMiningProcess`) and algorithm(s) (`AgentMiningAlgorithm`) in some of the approaches.

Considering the knowledge acquired during the analysis, some additional object properties and constraints presented in Table 3 were defined. All concepts were defined in a data glossary. Table 4 presents an extract of this glossary for the concepts presented in Figure 13.

4.2 Validation

To validate the defined ontology, the articles on agent mining selected from the systematic mapping were considered to create instances for all defined ontology classes, which means all elements presented in the

Table 3 Other constraints (axioms)

Constraint	Formalization
An agent competition (kind of `AgentCoordination`), when represented, is described by formal models	`AgentCompetition` isDescribeByFModel only `AgentMiningFormalModel`
A sequence diagram is used to describe the interaction among agents in a multi-agent system	`SequenceDiagram` describesInteraction exactly 1 `MultiagentInteractionMechanism`
An agent process concerns one or more elements related to an agent-related component	`AgentMiningApproach` and `AgentProcess` SubClassOf ((hasPart_directly some `AgentRelatedComponent`) and (follows some `AgentProcess`))
An agent algorithm concerns one or more elements regarded to an agent component (part of an agent mining approach)	`AgentMiningApproach` and `AgentRelatedAlgoritm` SubClassOf ((hasPart_directly some `AgentRelatedComponent`) and (applies some `AgentRelatedAlgorithm`))
A collaborative process describes agent components interactions	`AgentCoordination` and `AgentMiningApproach` SubClassOf (follows only `CollaborativeProcess`)

Table 4 Extract of Data Glossary

Concept	Definition
AgentMiningProblem	Problem addressed by the proposed agent mining system to be developed
AgentMiningApplication Domain	Application domain for which the agent mining system will be developed
AgentMiningApproach	Approach used for the design of the agent mining system to be developed
AgentMiningModel	Model defined to represent the agent mining system to be developed
AgentMiningProcess	Sequence of steps performed in the agent mining system to be developed
AgentMiningAlgorithm	Algorithm applied by the agent mining system to be developed
AgentRelatedComponent	Representation of an agent component in the agent mining approach
MiningRelatedElement	Representation of a mining process or algorithm

article (i.e. its domain application, algorithm, models, process and so on) were considered to instantiate the related concepts in the ontology.

Firstly, the concepts have been instantiated considering the 96 articles from the first execution of the protocol (see Section 2.3) and that were used for the definition of the ontology (see column #1st in Table 5). In this way, instances are defined for all classes, what was expected since the ontology is defined based on the 96 articles. This instantiation process enables to confirm that all the defined classes are really based on a published research.

Then, the 25 papers selected from the second execution of the protocol (after 2019) have been used to instantiate the concepts (see column #2nd in Table 5). Of course, in that case, all the defined ontology classes are covered since the instantiation is based on a limited set of papers. However, it confirmed that all approaches used in those 25 papers could be instantiated with the ontology, which corroborates that it represents a good conceptual representation of the approaches used for agent mining design and

Table 5 Number of instances for each concept from articles selected in the first(1st) and (2nd) execution of the protocol

Class				#1st	#2nd
AgentMiningApproach				96	25
MiningRelatedElement	MiningProcess			21	5
	MiningAlgoritm			86	21
AgentRelatedComponent	Agent			10	5
	AgentSet			8	2
		MultiAgentSystem		78	18
AgentMiningProblem	InformationRetrievalProblem			11	5
	DataAnalysisProblem			19	–
		DataAnalysisForNegotiation		1	–
		DataAnalysisForOptimization		17	5
		DataAnalysisForDecisionMaking		7	6
		DataAnalysisForDetection		14	6
		DataAnalysisForPrediction		5	4
		DataAnalysisForPersonalization		10	3
		DataAnalysisForProfiling		8	2
		DataAnalysisForSimulation		5	1
	DataVisualizationProblem			2	1
	DataManagementProblem			22	–
		DataIntegration		14	3
MultiAgentInteractionMechanism	AgentCommunication			37	11
	AgentOrganization			29	4
	AgentCoordination			2	2
		AgentCompetition	AgentNegotiation	10	2
		AgentCooperation		19	3
CommunicationProtocol				9	3

Table 5 Continued.

Class				#1st	#2nd
AgentBehavior				12	3
AgentBasedSimulation				9	5
AgentMiningApplicationDomain	SecurityDomain	Safety		2	–
		Security		4	3
			ISSecurity	9	3
	EntertainmentDomain	Tourism		1	2
		Gaming		4	2
	HealthDomain	HealthcareHomecare		4	1
		Medicine		3	1
	HumanitiesDomain	SocialSciences		4	–
		Education		9	2
		SocialNetworks		3	–
	ComputerScienceDomain	BioInformatics		2	–
		DistributedSystems	SensorsIoTSmart	3	–
			GridComputing	2	–
			SocialNetworks	3	–
			WebServices	4	1
		ISSecurity		9	3
		SoftwareEngieering		11	2
		HMI		6	2
		OperationalReasearch		1	–
	FinancialDomain	Insurance		1	1
		Finance		3	2
	CommercialDomain	Marketing		9	1
		E-Markets		5	1

Table 5 Continued.

Class			#1st	#2nd
	TransportationDomain		4	2
	IndustrialDomain	ProductionSystems	2	1
		Energy	3	–
	LifeSciencesDomain	Biology	2	–
		BioInformatics	2	–
AgentMiningProcess	MethodologicalProcess	AnalysisProcess	1	1
		EvaluationProcess	1	–
		SystemEngineeringProcess	1	–
	MiningProcess		1	1
		DetectionProcess	8	1
		InformationDiscoveryProcess	8	–
		LearningProcess	4	3
	AgentProcess		1	2
		CollectiveProcess	6	1
		DecisionMakingProcess	5	1
		NegotiationProcess	2	1
AgentMiningModel	SpecificInformationModel		52	5
		GlobalSystemArchitecture	46	13
	AgentMiningSemiFormalModel	KQML_Model	5	–
		FlowChart	8	5
		SequenceDiagram	2	–
		Graph	9	1
		UseCaseDiagram	2	1
		OntologyDiagram	18	–
	AgentMiningFormalModel		1	–
		LogicModel	3	3
		GameTheoryModel	2	–

Table 5 Continued.

Class				**#1st**	**#2nd**
AgentMiningAlgorithm	MiningAlgorithm	Extraction		9	1
		Classification		13	3
			DecisionTreeClassification	12	5
			SVM	6	3
			GeneticAlgorithmClassification	5	3
		Clustering		11	2
			FuzzyClustering	2	–
			DensityClustering	2	–
		AssociationRules		12	2
			Apriori	4	1
		QLearningOptimization		3	–
	AgentRelatedAlgorithm			6	1
		AntColonyOptimization		1	–
		QLearningOptimization		3	–

implementation. This validation step confirms also that the main concepts of the ontology (at the top of the taxonomies presented in Figures 3, 4, 5 and 13) are instantiated.

Some results that can be highlighted from Table 5 are :

- regarding `MiningRelatedElement`, mining algorithms are largely presented in the papers (86/96 from the first set of papers and 21/25 from the second set). Moreover several papers present both `MiningAlgorithm` and `MiningProcess`, for instance Yu *et al*. (2019), Zhao *et al*. (2018) and Zhang (2018);
- regarding `AgentRelatedComponent`, most of the papers present the use of several agents (86 papers in the first set and 20 papers in the second set), while some papers present the use of a unique agent (for instance, Atto & Kotova (2019), De Souza *et al*. (2019) and Yu *et al*. (2019)). When the study includes several agents, these are generally described as a MAS, with 8 exceptions in the first set of papers and 2 in the second, mainly in agent-based simulation (e.g. Mgbemena & Bell (2016), Liu *et al*. (2017));
- the main problem addressed is the data analysis (named `DataAnalysisProblem`) in both sets of articles. In particular, the data analysis for detection (`DataAnalysisForDetec- ion`) and the data analysis for optimization (`DataAnalysisForOptimization`) are well represented: respectively 14 and 17 instances. For example, in Sethi *et al*. (2019), an instance of `DataAnalysisForDetection` is presented to prevent and detect intrusions in enterprise network. In Chen & Chen (2017), an instance of `DataAnalysisForOptimization` is described to optimize scheduling;
- almost all articles are applied to a specific domain (`AgentMiningApplicationDomain`). The application domain that is the most represented in the articles from the 1st set is some sub-domain from `ComputerScienceDomain`. However, in the 2nd set the `SecurityDomain` is the most applied domain. In `ComputerScienceDomain,Calpur2018DomainSC` presents an example of solution focused on human computer interaction, specially on natural language processing. `SecurityDomain` is well represented in Tangod & Kulkarni (2020), Thanudas *et al*. (2019), Yang *et al*. (2019);
- `AgentMiningAlgorithm`, `AgentMiningModel` and `AgentMiningProcess` can be found in several papers. Some papers even include instances for each of these three concepts (e.g. Larsen *et al*. (2019), Roldán *et al*. (2018), Sethi *et al*. (2019));
- finally, the use of an illustrative and informal figure of the global system (named `GlobalSystemArchitecture`) is very common: 46 instances from 96 papers during the first round and 13 from 25 in the more recent papers. Good examples of `GlobalSystemArchitecture` can be found in Aksvonov & Antonova (2018) and Xu *et al*. (2020).

For a better view, the article (Tounsi *et al*., 2020) can be used as a complete example including instances of the ontology concepts, from the addressed problem to the elements used for the system design. The paper (Tounsi *et al*., 2020) presents a software system to address a credit scoring problem, based on specialized agents that perform data pre-processing and mining. To that end, the authors describe the solution using the following elements represented in the ontology:

- one instance of the concept `AgentMiningApplicationDomain` and more specifically of the concept `FinancialDomain`;
- instances for the concepts `DataIntegration`, `DataAnalysisForPrediction` and `DataAnalysisForDetection`;
- one instance of the concept `MethodologicalProcess`: a CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology extension for Credit Scoring is described;
- one instance of the concept `AgentCoordination`: several types of agent and the coordination between them are detailed;
- instances of the concept `SpecificInformalDiagram`: informal diagrams are used to explain the CRISP-DM framework, in addition to a `GlobalSystemArchitecture` detailing the different layers of the system general architecture.

5 Threats to validity

According to Wohlin *et al*. (2012), four threats to validity should be considered in software engineering studies: construct validity, internal validity, conclusion validity and external validity.

Threats to the **construct validity** illustrate the relation between theory and observation. It must be assured that (a) the treatment (information extracted from the articles) adequately reflects the construct (in this study: the articles) and (b) the outcome (final model obtained from the analysis of the competency questions introduced in Section 4) adequately reflects the effects of the construct. This threat is associated with the appropriate extraction of the information from the articles and is limited by the used set of articles. To address (a) a specific data extraction form was developed to extract the information, and the exact transcription of the content of the article was required to support the comprehension and analysis by all researchers. Periodical meeting to discuss what was extracted for each required information assured the correct identification of the treatment. To address (b), the competency questions (CQ) and the data form used for data extraction (see Section 3) were defined consistently, that is, the selected data should answer the CQs and these in turn should correspond to the collected data required in the data form as explained in Section 4. Moreover, after having defined the complete ontological model, all articles were used to instantiate the model assuring that the final model (b) represents the construct.

Threats to the **internal validity** draw from influences that can affect the independent variables with respect to causality without the researchers' knowledge. In a study based on a systematic mapping, this threat is related to the possibility to miss some studies in the set of paper considered for this analysis. Moreover, the natural bias the researchers can have in selecting the papers can influence also the results. First of all, since only two databases were used for the systematic mapping, some studies may have been missed, which may imply incomplete results. Nevertheless, SCOPUS and Web of Science are recognized for indexing different journals, conferences and workshops in computer science. Moreover, considering the large number of articles analyzed in the systematic mapping (3384) that allow to select the articles for this study, the set of 121 articles can be considered quite representative of the domain; therefore, this risk was accepted. Researcher bias may have appeared during the selection and data extraction not only because of differences in interpretation but also because the researchers have different levels of experience in the two fields of study (agents and DM). Several actions were undertaken to mitigate this threat. First, in the article selection step, several trials were performed to harmonize understanding about what should be collected. Periodical discussions about the doubtful articles were also held. Moreover, a well-defined protocol and process can lead to consistency studies. Finally, peer-review of the data extraction was performed, contributing to mitigate this risk.

Threats to the **conclusion validity** are those that affect the ability to draw the correct conclusion about the relation between the treatment and the outcome of the study. In this study, this threat concerns the final model itself (the concepts and relationship between concepts) and the treatment (the information collected from the papers). As presented in the previous section, the conclusion was drawn from data analysis, systematically retrieved from data collected from a systematic mapping study, ensuring a high degree of traceability between data and conclusions as shown in the validation section. However, like any qualitative study, a significant threat is the interpretation bias. To reduce such bias, several meetings with the participation of the four researchers had been accomplished aiming to discuss the interpretation of the different concepts extracted from the data as described in the methodological section. From those meetings has emerged the organization of concepts and the choice of how to name these concepts in the ontology model. There is always a risk that the researchers missed some information, however considering that four researchers working on that and several consensual meetings were organized, this risk was accepted. Finally, the meaning of each defined concept was clarified using a data dictionary to avoid any further misinterpretation, mitigating the interpretation bias.

Finally, threats to the **external validity** are conditions that limit our ability to generalize the results of the experiment outside the scope of the study. This study relies on data analysis from the literature, and the results represent synthesized concepts related to agent mining, as found in the primary sources. Although the results come from the set of articles, the main concepts of the ontology (concerning the top concepts of taxonomies and main concepts presented in Figures 3, 4, 5 and 13) can be used outside of this study, as confirmed by the analysis of new articles (see Section 4.2). However, new concepts can be defined from new papers. Indeed, the 2nd set of articles (25 against to 96 used for the definition) is not large to confirm the generalizability of all concepts since it considered only papers of half a year. Therefore, further instantiation of the ontology may include new sub-classes in the taxonomies, and as consequence, make the model evolve.

6 Conclusion

Emerged from the integration of two large domains (software agents and DM), agent mining software systems have been developed for diverse purposes and applied for several domains. In the absence of a proper methodology, designers have used different approaches to the development of these applications. In this paper, we analyzed 121 articles on agent mining with the goal of identifying the approaches used for the development of agent mining systems. The analysis was performed from an ontological perspective, organizing the approaches and elements of the approaches in an ontology.

The main contributions of this paper are:

- an ontology describing the agent mining approaches according to the addressed problems, the application domains, the agent-related and mining-related elements, the models, processes and algorithms. This ontology is formalized in Protégé and is composed of 108 classes, 51 relationships ('object properties') and 255 logical axioms;
- a set of instances collected from the agent mining studies presented in the 121 analyzed articles; and
- a taxonomy of application domains for agent mining solutions. This taxonomy can be further used to complete the largely used classification from Kotonya *et al*. (2003) with current application domains not considered in 2003.

As future work, we plan to develop a recommendation system to support agent mining designers in choosing the right approach considering the problem and the application domain to tackle. This recommendation system will be developed on top of the ontology and its instances collected from published studies. The ontology and the instances would be used as a data repository to query the best approach for the problem domain on hand. The ontology can also be used to support the definition of guidelines for agent mining design. Finally, there is no specific model for the design of agent mining systems. This could be explored in a further research considering the 46 propositions of general architecture model presented in the studies.

Competing interests

The authors declare none.

Acknowledgments

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

References

Abdelrahman, S., Bahgat, R. & Farag, G. 2011. Order statistics Bayesian-mining agent modelling for automated negotiation. *Informatica (Ljubljana)* **35**(1), 123–137.

Aksvonov, K. & Antonova, A. 2018. Development of a hybrid decision-making method based on a simulation-genetic algorithm in a web-oriented metallurgical enterprise information system. In *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 197–202. doi: 10.1109/ICUFN.2018.8436676.

Anand, N., Yang, M., van Duin, J. & Tavasszy, L. 2012. Genclon: an ontology for city logistics. *Expert Systems with Applications* **39**(15), 11944–11960. doi: 10.1016/j.eswa.2012.03.068.

Aridor, Y., Carmel, D., Maarek, Y. S., Soffer, A. & Lempel, R. 2002. Knowledge encapsulation for focused search from pervasive devices. *ACM Transactions on Information Systems* **20** (1), 25–46. doi: 10.1145/503104.503106.

Atto, K. & Kotova, E. E. 2019. Data mining agents as means of communicating users in an e-learning environment. In *Proceedings of the 2019 IEEE Communication Strategies in Digital Society Seminar, ComSDS 2019*, 39–42.

Bajo, J., Campbell, A. T. & Zhou, X. 2016. Mobile sensing agents for social computing environments. *Advances in Intelligent Systems and Computing* **473**, 157–167.

Basili, V. R., Caldiera, G. & Rombach, H. D. 1994. The goal question metric approach. In *Encyclopedia of Software Engineering*. Wiley.

Bellifemine, F., Poggi, A. & Rimassa, G. 2001. JADE: a FIPA2000 compliant agent development environment. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 216–217.

Beni, G. 2005. From swarm intelligence to swarm robotics. In *Swarm Robotics. SR 2004*, Şahin, E. & Spears, W. (eds.), 3342. *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 1–9. doi: 10.1007/978-3-540-30552-1_1.

Bernon, C., Camps, V., Gleizes, M.-P. & Picard, G. 2004. Tools for self-organizing applications engineering. In *Engineering Self-Organising Systems, Nature-Inspired Approaches to Software Engineering*, 2977. *LNCS*. Springer, 283–298.

Botelho, W. T., Marietto, M. D. G. B., Mendes, E. D. L., Sousa, D. R. D., Pimentel, E. P., da Silva, V. L. & dos Santos, T. 2020. Toward an interdisciplinary integration between multi-agents systems and multi-robots systems: a case study. *The Knowledge Engineering Review* **35**, e35. doi: 10.1017/S0269888920000375.

Brank, J., Mladenić, D. & Grobelnik, M. 2006. Gold standard based ontology evaluation using instance assignment. In *Proceedings of 4th International EON Workshop'06 Evaluation of Ontologies for the Web @WWW2006*, CEUR Workshop Proceedings, Edinburgh, UK.

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F. & Mylopoulos, J. 2004. Tropos: an agent-oriented software development methodology. *Autonomous Agent and Multi-Agent Systems* **3**, 203–236.

Calpur, M. C., Tatlidede, M. U. & Cataloglu, I. 2018. Domain specific conversational intelligent agents: natural language processing in real world applications. In *Proceedings of the 12th Turkish National Software Engineering Symposium UTMS 2018*. CEUR Workshop Proceedings.

Canito, A., Marreiros, G. & Corchado, J. M. 2019. Automatic document annotation with data mining algorithms. In Rocha, A., Adeli, H., Reis, L. & Costanzo, S., editors, *New Knowledge in Information Systems and Technologies. WorldCIST'19 2019. Advances in Intelligent Systems and Computing*, 930. Springer.

Cao, L., Luo, C. & Zhang, C. 2007. Agent-mining interaction: an emerging area. In *2nd International Workshop Autonomous Intelligent Systems: Agents and Data Mining, AIS-ADM 2007*, 4476. LNAI, 60–73.

Cao, L., Weiss, G. & Yu, P. S. 2012. A brief introduction to agent mining. *Autonomous Agents and Multi-Agent Systems* **25** (3), 419–424.

Chen, C. & Chen, X. 2017. Scheduling optimization in restricted channels based on the agent technology and Bayesian network. In *2017 4th International Conference on Transportation Information and Safety, ICTIS 2017 – Proceedings*, Yan, X., Zhong, M., Lu, M., Wu, C., Qiu, Z. & Hu, Z. (eds.). Institute of Electrical and Electronics Engineers Inc., 291–295. doi: 10.1109/ICTIS.2017.8047779.

Chen, P. & Plale, B. A. 2015. Big data provenance analysis and visualization. In *Proceedings – 2015 IEEE/ACM 15th International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2015*. Institute of Electrical and Electronics Engineers Inc., 797–800. doi: 10.1109/CCGrid.2015.85.

Chen, Y., Zeng, X., Chen, X. & Guo, W. 2020. A survey on automatic image annotation. *Applied Intelligence* **50**(10), 3412–3428.

Cobos, C., Niño, M., Mendoza, M., Fabregat, R. & Gomez, L. 2007. Learning management system based on SCORM, agents and mining. In *8th International Conference on Web Information Systems Engineering, WISE 2007*, **4831**. LNCS, 298–309.

Cordeiro, F., Werneck, V., Santos, N. & Cysneiros, L. 2016. Mas ontology: ontology for multiagent systems. In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016)*, **1**. SCITEPRESS, 536–543.

Cossentino, M. 2005. *From Requirements to Code with the PASSI Methodology*, chapter IV, 79–106. Idea Group Publishing.

Cranefield, S., Willmott, S. & Finin, T. 2002. Introduction to the special issue on ontologies in agent systems. *The Knowledge Engineering Review* **17**(1), 1–5. doi: 10.1017/S0269888902000310.

Dahl, V., Barahona, P., Bel-Enguix, G. & Krippahl, L. 2010. Biological concept formation grammars: a flexible, multiagent linguistic tool for biological processes. In *ICAART 2010 – 2nd International Conference on Agents and Artificial Intelligence, Proceedings*, 2, 388–394.

De Arruda, H. F., Silva, F. N., Costa, L. d. F. & Amancio, D. R. 2017. Knowledge acquisition: a Complex networks approach. *Information Sciences* **421**, 154–166. doi: 10.1016/j.ins.2017.08.091.

De Souza, R. M. M., Vilasbôas, F. G., Notargiacomo, P. & De Castro, L. 2019. Integrating an association rule mining agent in an ERP system: a proposal and a computational scalability analysis. In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, **2**, 778–786.

Dignum, V., Javier, V.-S. & Frank, D. 2005. *OMNI: Introducing Social Structure, Norms and Ontologies into Agent Organizations*, **3346**. LNCS. Springer-Verlag, 181–198.
Dileo, J., Jacobs, T. & DeLoach, S. 2002. Integrating ontologies into multiagent systems engineering. In *Proceedings of 4th International Bi-Conference Workshop on Agent Oriented Information Systems (AOIS)*, 15–16.
Dimou, C., Symeonidis, A. L. & Mitkas, P. A. 2007. Evaluating knowledge intensive multi-agent systems. In *2nd International Workshop Autonomous Intelligent Systems: Agents and Data Mining, AIS-ADM 2007*, **4476**. LNAI, 74–87.
Dos Santos, C. T. & Bazzan, A. L. C. 2005. Integrating knowledge through cooperative negotiation – a case study in bioinformatics. In *International Workshop on Autonomous Intelligent Systems: Agents and Data Mining, AIS-ADM 2005*, **3505**. LNAI, 277–288.
Duarte, V. A. R. & Julia, R. M. S. 2017. Improving netfeaturemap-based representation through frequent pattern mining in a specialized database. In *Proceedings – 2016 IEEE 28th International Conference on Tools with Artificial Intelligence, ICTAI 2016*, Esposito, A., Alamaniotis, M., Mali, A. & Bourbakis, N. (eds.). Institute of Electrical and Electronics Engineers Inc., 954–961 doi: 10.1109/ICTAI.2016.0147.
Esfandi, A. 2010. Efficient anomaly intrusion detection system in adhoc networks by mobile agents. In *Proceedings – 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, **7**, 73–77. doi: 10.1109/ICCSIT.2010.5563804.
Eugin Lilly, M. M. & Venkataraman, S. 2013. Predictive load balancing for data mining in distributed systems. *Journal of Theoretical and Applied Information Technology* **53**(1), 13–23.
Fowler, C. A. & Hammell, R. J. 2015. Mining information assurance data with a hybrid intelligence/multi-agent system. In *2015 IEEE/ACIS 14th International Conference on Computer and Information Science, ICIS 2015 – Proceedings*. Institute of Electrical and Electronics Engineers Inc., 23–28 doi: 10.1109/ICIS.2015.7166564.
Freitas, A., Bordini, R. H. & Vieira, R. 2017. Model-driven engineering of multi-agent systems based on ontologies. *Applied Ontology* **12**(2), 157–188.
Gago, P., Santos, M. F., Silva, S. A., Cortez, P., Neves, J. M. & Gomes, L. 2005. INTCare: a knowledge discovery based intelligent decision support system for intensive care medicine. *Journal of Decision Systems* **14**(3), 241–259. doi: 10.3166/jds.14.241-259.
Garcia-Sanchez, F., Tomas Fernandez-Breis, J., Valencia-Garcia, R., Miguel Gomez, J. & Martinez-Bejar, R. 2008. Combining semantic web technologies with multi-agent systems for integrated access to biological resources. *Journal of Biomedical Informatics* **41**(5, SI), 848–859. doi: 10.1016/j.jbi.2008.05.007.
Gaya, M. C. & Giráldez, J. I. 2011. Merging local patterns using an evolutionary approach. *Knowledge and Information Systems* **29**(1), 1–24. doi: 10.1007/s10115-010-0332-x.
Gerardo, B. D. & Lee, J. 2009. A framework for discovering relevant patterns using aggregation and intelligent data mining agents in telematics systems. *Telematics and Informatics* **26**(4), 343–352. doi: 10.1016/j.tele.2008.05.003.
Ghedini Ralha, C. & Sarmento Silva, C. V. 2012. A multi-agent data mining system for cartel detection in Brazilian government procurement. *Expert Systems with Applications* **39**(14), 11642–11656. doi: 10.1016/j.eswa.2012.04.037.
Gómez-Sanz, J. & Fuentes-Fernández, R. 2015. Understanding agent-oriented software engineering methodologies. *The Knowledge Engineering Review* **30**(4), 375–393.
Gorodetsky, V. 2013. Agents and distributed data mining in smart space: challenges and perspectives. In *8th International Workshop on Agents and Data Mining Interaction, ADMI 2012*, **7607**. LNAI, 153–165.
Gorodetsky, V., Karsaev, O. & Samoilov, V. 2005. Multi-agent and data mining technologies for situation assessment in security-related applications. In *International Workshop "Monitoring, Security and Rescue Techniques in Multiagent Systems" (MSRAS 2004)*, **28**, 411–422.
Gorodetsky, V., Karsaev, O., Samoylov, V. & Serebryakov, S. 2008. Interaction of agents and data mining in ubiquitous environment. In *Proceedings – 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology – Workshops, WI-IAT Workshops 2008*, 562–566. doi: 10.1109/WIIAT.2008.354.
Gorunescu, F. 2011. *Data Mining. Concepts, Models and Techniques*, **12**. Intelligent Systems Reference *Library*. Springer.
Gottgtroy, P. 2007. Ontology driven knowledge discovery process: a proposal to integrate ontology engineering and KDD. In *PACIS 2007 – 11th Pacific Asia Conference on Information Systems: Managing Diversity in Digital Enterprises*.
Grislin-Le Strugeon, E., Oliveira, K. M., Thilliez, M. & Petit, D. 2021. A systematic mapping study on agent mining. *Journal of Experimental & Theoretical Artificial Intelligence*, 1–26. doi: 10.1080/0952813X.2020.1864784.
Gruber, T. R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* **5**(2), 199–220.
Grüninger, M. & Fox, M. S. 1995. Methodology for the design and evaluation of ontologies. Technical report, University of Toronto, Toronto, Canada.
Guizzardi, G. & Wagner, G. 2005. Towards ontological foundations for agent modelling concepts using the unified foundational ontology (UFO). In *Agent-Oriented Information Systems II*, Bresciani, P., Giorgini, P., Henderson-Sellers, B., Low, G. & Winikoff, M. (eds.). Springer Berlin Heidelberg, 110–124. ISBN 978-3-540-31946-7.

Haider, K., Tweedale, J. & Jain, l. 2009. An intelligent decision support system using expert systems in a MAS. *Studies in Computational Intelligence* **199**, 213–222.

Hamdi, M. S. 2007. Masacad: a multi-agent approach to information customization for the purpose of academic advising of students. *Applied Soft Computing Journal* **7**(3), 746–771. doi: 10.1016/j.asoc.2006.02.001.

Huhns, M. N. & Singh, M. P. 1997. Ontologies for agents. *IEEE Internet Computing* **1**(6), 81–83.

Ioniţă, I. & Ioniţă, L. 2013. An agent-based approach for building an intrusion detection system. In *Proceedings – RoEduNet IEEE International Conference*. IEEE Computer Society. doi: 10.1109/RoEduNet.2013.6714184.

Islam, K. S. 2007. An approach to argumentation context mining from dialogue history in an e-market scenario. In *AIDM '07 Proceedings of the 2nd International Workshop on Integrating Artificial Intelligence and Data Mining*, Ong K.-L., Li, W. & Gao J. (eds.), 84. *Conferences in Research and Practice in Information Technology Series*. Australian Computer Society, 73–81.

Izumi, K., Matsui, H. & Matsuo, Y. 2007. Socially embedded multi agent based simulation of financial market. In *Proceedings of the International Conference on Autonomous Agents*, 1110–1112. doi: 10.1145/1329125.1329337.

Jayyousi, T. W. & Reynolds, R. G. 2014. Extracting urban occupational plans using cultural algorithms [application notes]. *IEEE Computational Intelligence Magazine* **9** (3), 66–87. doi: 10.1109/MCI.2014.2326103.

Jiang, P., Mair, Q. & Feng, Z.-R. June 2007. Agent alliance formation using ART-networks as agent belief models. *Journal of Intelligent Manufacturing* **18** (3), 433–448. doi: 10.1007/s10845-007-0032-x.

Kadhim, M. A., Alam, A. & Kaur, H. 2016. A multi-intelligent agent system for automatic construction of rule-based expert system. *International Journal of Intelligent Systems and Applications* **8**(9), 62–68. doi: 10.5815/ijisa.2016.09.08.

Kaschesky, M., Sobkowicz, P. & Bouchard, G. 2011. Opinion mining in social media: modeling, simulating, and visualizing political opinion formation in the web. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, 2011*, 317–326. doi: 10.1145/2037556.2037607.

Keet, C. M., Ławrynowicz, A., d'Amato, C., Kalousis, A., Nguyen, P., Palma, R., Stevens, R. & Hilario, M. 2015. The data mining optimization ontology. *Journal of Web Semantics* **32**, 43–53.

Kotonya, G., Sommerville, I. & Hall, S. 2003. Towards a classification model for component-based software engineering research. In *Proceedings 29th Euromicro Conference*, 43–52. doi: 10.1109/EURMIC.2003.1231566.

Kotova, E. E. 2017. Intellectual data analysis in the educational process. In *Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017*, Shaposhnikov, S. (ed.). Institute of Electrical and Electronics Engineers Inc., 757–759 doi: 10.1109/SCM.2017.7970714.

Kularbphettong, K., Meesad, P. & Clayton, G. 2012. A hybrid system based on multi-agent systems in case of e-WeddingThailand. In *7th International Workshop on Agents and Data Mining Interaction, ADMI 2011*, **7103**. LNAI, 344–359.

Kusumura, Y., Hijikata, Y. & Nishida, S. 2004. NTM-agent: text mining agent for net auction. *IEICE Transactions on Information and Systems* **E87-D** (6), 1386–1396.

Kwon, J., Kim, S. & Yoon, Y. 2004. Just-in-time recommendation using multi-agents for context-awareness in ubiquitous computing environment. In *International Conference on Database Systems for Advanced Applications DASFAA 2004*, 2973, 656–669.

Larsen, J. B., Burattin, A., Davis, C. J., Hjardem-Hansen, R. & Villadsen, J. 2019. A data driven agent elicitation pipeline for prediction models. *Lecture Notes in Business Information Processing* **362**, 570–582.

Lee, J. H. & Park, S. C. 2005. Intelligent profitable customers segmentation system based on business intelligence tools. *Expert Systems with Applications* **29**(1), 145–152. doi: 10.1016/j.eswa.2005.01.013.

Li, C.-T. & Lin, S.-D. 2012. Evaplanner: an evacuation planner with social-based flocking kinetics. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1568–1571. doi: 10.1145/2339530.2339782.

Liu, X., Merrick, K. & Abbass, H. 2017. Designing artificial agents to detect the motive profile of users in virtual worlds and games. In *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/SSCI.2016.7850036.

Liu, Z., Niu, D., Yang, X. & Sheng, W. 2009. Research on intelligent decision support system for power system. In *2009 IEEE International Conference on Information and Automation, ICIA 2009*, 412–417. doi: 10.1109/ICINFA.2009.5204959.

Loia, V., Pedrycz, W., Senatore, S. & Sessa, M. I. 2007. Interactive knowledge management for agent-assisted web navigation. *International Journal of Intelligent Systems* **22** (10), 1101–1122. doi: 10.1002/int.20239.

Lopez, M. F., Gomez-Perez, A., Sierra, J. P. & Sierra, A. P. 1999. Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent Systems & Their Applications*, **14** (1), 37–45.

Matei, O., Di Orio, G., Jassbi, J., Barata, J. & Cenedese, C. 2016. Collaborative data mining for intelligent home appliances. *IFIP Advances in Information and Communication Technology* **480**, 313–323.

Mateo, R. M. A., Yoon, I. & Lee, J. 2008. Data-mining model based on multi-agent for the intelligent distributed framework. In *2nd KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications, KES-AMSTA 2008*, **4953**, LNAI, 753–762. doi: 10.1007/978-3-540-78582-8_76.

Menczer, F. 2003. Complementing search engines with online web mining agents. *Decision Support Systems* **35**(2), 195–212. doi: 10.1016/S0167-9236(02)00106-9.

Mgbemena, C. & Bell, D. 2016. Data-driven customer behaviour model generation for agent based exploration. In *Proceedings of the 49th Annual Simulation Symposium*, 48. *Simulation Series*, 155–161. The Society for Modeling and Simulation International.

Moemeng, C., Wang, C. & Cao, L. 2012. Obtaining an optimal MAS configuration for agent-enhanced mining using constraint optimization. In *7th International Workshop on Agents and Data Mining Interaction, ADMI 2011*, 7103. LNAI, 46–57. doi: 10.1007/978-3-642-27609-5_5.

Moemeng, P., Cao, L. & Zhang, C. 2008. F-trade 3.0: an agent-based integrated framework for data mining experiments. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT Workshops*, 612–615. doi: 10.1109/WIIAT.2008.230.

Musen, M. A. & The Protégé Team. 2015. The Protégé project: a look back and a look forward. *AI matters* **1**(4), 4–12.

Noy, N. F. & McGuinness, D. L. 2001. Ontology development 101: a guide to creating your first ontology. Technical report, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Olivares-Alarcos, A., Beßler, D., Khamis, A., Goncalves, P., Habib, M. K., Bermejo-Alonso, J., Barreto, M., Diab, M., Rosell, J., Quintas, J., Olszewska, J., Nakawala, H., Pignaton, E., Gyrard, A., Borgo, S., Alenyà, G., Beetz, M. & Li, H. 2019. A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review* 34, e29. doi: 10.1017/S0269888919000237.

Oliveira, K. M., Bacha, F., Mnasser, H. & Abed, M. 2013. Transportation ontology definition and application for the content personalization of user interfaces. *Expert Systems with Applications* **40**(8), 3145–3159. doi: https://doi.org/10.1016/j.eswa.2012.12.028.

Oskouei, R. J. & Moradi-Kor, N. 07 2016. Proposing a novel adaptive learning management system: an application of behavior mining & intelligent agents. *Intelligent Automation and Soft Computing*, 1–7. doi: 10.1080/10798587.2016.1186429.

Pedrycz, W. & Rai, P. 2008. A multifaceted perspective at data analysis: a study in collaborative intelligent agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **38**(4), 1062–1072. doi: 10.1109/TSMCB.2008.925728.

Petersen, K., Vakkalanka, S. & Kuzniarz, L. 2015. Guidelines for conducting systematic mapping studies in software engineering: an update. *Information and Software Technology* **64**, 1–18.

Phillips, J. & Buchanan, B. G. 2001. Ontology-guided knowledge discovery in databases. In *Proceedings of the First International Conference on Knowledge Capture*. Association for Computing Machinery (ACM), 123–130.

Pi, D., Wang, Q., Li, W. & Lv, J. 2005. APA: an interior-oriented intrusion detection system based on multi-agents. In *Third International Conference on Computer Network and Mobile Computing, ICCNMC 2005*, **3619**, 1227–1233.

Pignaton de Freitas, E., Bermejo-Alonso, J., Khamis, A., Li, H. & Olszewska, J. I. 2020. Ontologies for cloud robotics. *The Knowledge Engineering Review* **35**, e25. doi: pignaton2020.

Pinzón, C. I., De Paz, J. F., Herrero, Á., Corchado, E., Bajo, J. & Corchado, J. M. 2013. IdMAS-SQL: intrusion detection based on MAS to detect and block SQL injection through data mining. *Information Sciences* **231**, 15–31. doi: 10.1016/j.ins.2011.06.020.

Ponte, B., De La Fuente, D., Pino, R. & Rosillo, R. 2015. Real-time water demand forecasting system through an agent-based architecture. *International Journal of Bio-Inspired Computation* **7** (3), 147–156. doi: 10.1504/IJBIC.2015.069559.

Pruengkarn, R., Wong, K. W. & Fung, C. C. 2017. A review of data mining techniques and applications. *Journal of Advanced Computational Intelligence and Intelligent Informatics* **21**(1), 31–48.

Pulmano, C. E. & Estuar, M. R. J. E. 2016. Towards developing an intelligent agent to assist in patient diagnosis using neural networks on unstructured patient clinical notes: Initial analysis and models. *Procedia Computer Science* **100**, 263–270. doi: 10.1016/j.procs.2016.09.153.

Pultar, E., Raubal, M. & Goodchild, M. F. 2008. GEDMWA: Geospatial Exploratory Data Mining Web Agent. In *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 499–502. doi: 10.1145/1463434.1463519.

Qiu, J., Luo, L. & Zhang, Y. 2009. The design and assessment of web data analysis system. In *Proceed. International Conference on E-Business and Information System Security, EBISS 2009*. doi: 10.1109/EBISS.2009.5138109.

Ribino, P., Cossentino, M., Lodato, C., Lopes, S., Sabatucci, L. & Seidita, V. 2013. Ontology and goal model in designing BDI multi-agent systems. *In CEUR Workshop Proceedings*, **1099**, 66–72.

Rodríguez, M., Bermejo-Alonso, J., Hernandez Corbato, C. & Sanz, R. 2012. Ontology-driven description and engineering of autonomous systems: application to process systems engineering. In *22nd European Symposium on Computer Aided Process Engineering*, Bogle, I. D. L. & Fairweather, M. (eds.), 30. *Computer Aided Chemical Engineering*. Elsevier, 717–721.

Roldán, J. J., Olivares-Méndez, M. A., del Cerro, J. & Barrientos, A. 2018. Analyzing and improving multi-robot missions by using process mining. *Autonomous Robots* **42**(6), 1187–1205.

Rosé, C. P. & Ferschke, O. 2016. Technology support for discussion based learning: from computer supported collaborative learning to the future of massive open online courses. *International Journal of Artificial Intelligence in Education* **26**(2), 660–678. doi: 10.1007/s40593-016-0107-y.

Saadi, C. & Chaoui, H. 2017. Intrusion detection system based interaction on mobile agents and clust-density algorithm 'IDS-AM-Clust'. In *Colloquium in Information Science and Technology, CIST*, El Mohajir, M., El Mohajir, B., Chahhou, M. & Al Achhab, M. (eds.), 681–684. doi: 10.1109/CIST.2016.7804973.

Sadhasivan, D. K. & Balasubramanian, K. 2017. A Fusion of Multiagent Functionalities for Effective Intrusion Detection System. *Security and Communication Networks.* doi: 10.1155/2017/6216078.

Samoylov, V. & Gorodetsky, V. 2005. Ontology issue in multi-agent distributed learning. In *International Workshop on Autonomous Intelligent Systems: Agents and Data Mining, AIS-ADM 2005*, 3505. LNAI, 215–230.

Santos Jr., E. & Johnson Jr., G. 2004. Toward detecting deception in intelligent systems. In *Proceedings of SPIE – The International Society for Optical Engineering*, Trevisani D.A. & Sisti, A. F. (eds.), 5423, 130–141. doi: 10.1117/12.547296.

Sateli, B. & Witte, R. 2017. Personal research agents on the web of linked open data. In *Language, Data, and Knowledge: Proceedings of International Conference on Language, Data and Knowledge LDK 2017*, McCrae, J., Hellmann, S., Buitelaar, P., Gracia, J., Bond, F. & Chiarcos, C. (eds.), 10318. LNAI, 10–25.

Sethi, K., Sai Rupesh, E., Kumar, R., Bera, P. & Madhav, Y. V. 2019. A context-aware robust intrusion detection system: a reinforcement learning-based approach. *International Journal of Information Security* **19**, 657–678.

Shih, D.-H., Chiang, H.-S. & Lin, B. 2008. Collaborative spam filtering with heterogeneous agents. *Expert Systems with Applications* **35**(4), 1555–1566. doi: 10.1016/j.eswa.2007.08.062.

Silva, C. V. S. & Ralha, C. G. 2011. Agmi – an agent-mining tool and its application to Brazilian government auditing. In *WEBIST 2011 – Proceedings of the 7th International Conference on Web Information Systems and Technologies*, 535–538.

Sinnappan, S., Williams, M.-A. & Muthaly, S. 2001. Agent based architecture for internet marketing. In *6th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2000*, **2112**. Springer Verlag, 158–169.

Šlapák, M. & Neruda, R. 2014. Multiobjective genetic programming of agent decision strategies. In *5th International Conference on Innovations in Bio-Inspired Computing and Applications, IBICA 2014*, **303**. Springer Verlag, 173–182. doi: 10.1007/978-3-319-08156-4_18.

Strauss, A. & Corbin, J. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications. ISBN 9780803959408.

Studer, R., Benjamins, V. R. & Fensel, D. 1998. Knowledge engineering: principles and methods. *Data and Knowledge Engineering* **25**(1), 161–197. doi: https://doi.org/10.1016/S0169-023X(97)00056-6.

Tangod, K. & Kulkarni, G. 2020. Secure communication through multiagent system-based diabetes diagnosing and classification. *Journal of Intelligent Systems* **29**(1), 703–718.

Thanawala, P. & Joshi, M. 2017. Autorec: intelligent agent for rapid recommendation to engender or sustain blogbased virtual community. In *Proceedings – 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICCUBEA.2016.7860027.

Thanudas, B., Shridharan, S., Raj, V. C., Sairam, A. P., Gajmoti, V. & Joshi, P. 2019. A novel architecture for an integrated enterprise network security system. *International Journal of Security and Networks*, **14** (1), 47–60. doi: 10.1504/IJSN.2019.098919.

Tounsi, Y., Anoun, H. & Hassouni, L. 2020. Csmas: improving multi-agent credit scoring system by integrating big data and the new generation of gradient boosting algorithms. In *Proceedings of the 3rd International Conference on Networking, Information Systems & Security NISS2020*, 1–7. doi: 10.1145/3386723.3387851.

Tran, Q.-N. N. & Low, G. 2008. Mobmas: a methodology for ontology-based multi-agent systems development. *Information and Software Technology* **50** (7–8), 697–722.

W3C. 11 Aug 2005. Simple part-whole relations in owl ontologies, w3c editor's draft. Technical report, W3C.

Wang, X., Niu, W., Li, G., Yang, X. & Shi, Z. 2012. Mining frequent agent action patterns for effective multi-agent-based web service composition. In *7th International Workshop on Agents and Data Mining Interaction, ADMI 2011*, 7103. LNAI, 211–227. doi: 10.1007/978-3-642-27609-5_14.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B. & Wesslén, A. 2012. *Experimentation in Software Engineering*. Springer-Verlag Berlin Heidelberg.

Wu, C.-A., Lin, W., Tseng, M.-C. & Wu, C.-C. 2007. Ontology-incorporated mining of association rules in data warehouse. *Journal of Internet Technology* 8(4), 477–485.

Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R. & Zakharyaschev, M. 2018. Ontology-based data access: a survey. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 5511–5519.

Xu, D.-S. & Yan, S.-L. 2009. Multi-agent in ant colony algorithm approach for solving traveling salesman problem. In *2009 International Workshop on Intelligent Systems and Applications, ISA 2009*. doi: 10.1109/IWISA.2009.5072962.

Xu, J., Yao, L., Li, L., Ji, M. & Tang, G. 2020. Argumentation based reinforcement learning for meta-knowledge extraction. *Information Sciences* **506**, 258–272.

Xu, Y., Zhang, W., Liu, W. & Ferrese, F. 2012. Multiagent-based reinforcement learning for optimal reactive power dispatch. *IEEE Transactions on Systems Man and Cybernetics Part C - Applications and Reviews* 42(6), 1742–1751. doi: 10.1109/TSMCC.2012.2218596.

Xue, H., Guo, P., Zhang, H. & Kang, B. 2009. Study and realization of supplier business intelligence system for chain supermarket. In *Proceedings – 2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009*. doi: 10.1109/CISE.2009.5366538.

Yang, J., Li, T., Liang, G., He, W. & Zhao, Y. 2019. A hierarchy distributed-agents model for network risk evaluation based on deep learning. *CMES – Computer Modeling in Engineering and Sciences* **120**(1), 1–23.

Yang, Q., Feng, B. & Song, P. 2007. Study on anti-money laundering service system of online payment based on union-bank mode. In *2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007*, 4986–4989. doi: 10.1109/WICOM.2007.1223.

Yang, Y. J., Sung, T.-W., Wu, C. & Chen, H.-Y. 2010. An agent-based workflow system for enterprise based on fipa-os framework. *Expert Systems with Applications* **37**(1), 393–400.

Ying, W., Ray, P. & Lewis, L. 2013. A methodology for creating ontology-based multi-agent systems with an experiment in financial application development. In *46th Hawaii International Conference on System Sciences*, 3397–3406.

Yu, H.-X., Zheng, W.-S., Wu, A., Guo, X., Gong, S. & Lai, J.-H. 2019. Unsupervised person re-identification by soft multilabel learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2019-June**, 2143–2152.

Zambonelli, F., Jennings, N. R. & Wooldridge, M. 2003. Developing multiagent systems: the gaia methodology. *ACM Transactions on Software Engineering and Methodology* **12**(3), 317–370.

Zhang, Y. 2018. Key technologies of confrontational intelligent decision support for multi-agent systems. *Automatic Control and Computer Sciences* **52** (4), 283–290.

Zhao, M., Cai, W. & Turner, S. 2018. Clust simulating realistic crowd behaviour by mining pattern from crowd videos. *Computer Graphics Forum* **37**(1), 184–201.