

Business Case Mining and E-R Modeling Optimization

Zhaohao Sun^{1,2}, Paul Pinjik¹ & Francisca Pambel¹

¹Department of Business Studies, PNG University of Technology, Lae, Papua New Guinea

²School of Mathematical Sciences, Chongqing Normal University, Chongqing, China

Correspondence: Zhaohao Sun, Department of Business Studies, PNG University of Technology, Private Mail Bag, Lae 411, Morobe, Papua New Guinea.

Received: June 7, 2021 Accepted: June 21, 2021 Online Published: July 9, 2021

doi:10.11114/set.v8i1.5288

URL: <https://doi.org/10.11114/set.v8i1.5288>

Abstract

Business case mining and business rule discovery are at the center for entity relationship (E-R) modeling and database design to obtain E-R models. How to transform business cases through business rules into E-R models is a fundamental issue for database design. This article addresses this issue by exploring business case mining and E-R modeling optimization. Business case mining is business rule discovery from a business case. This article reviews case-based reasoning, explores business case-based reasoning, and presents a unified approach to business case mining for business rule discovery. The approach includes people-centered entity/business rule discovery and function-centered entity/business rule discovery. E-R modeling optimization aims to improve the E-R modeling process to get a better E-R diagram that reflects the business case properly. This article proposes a unified optimal method for E-R modeling. The unified optimal method includes people-centered E-R modeling, function-centered E-R modeling, and hierarchical E-R modeling. The approach proposed in this research will facilitate the research and development of E-R modeling, database design, data science, and big data analytics.

Keywords: business case mining, business rule discovery, entity relationship (E-R) modeling optimization, database design, data science, big data analytics

1. Introduction

Data science, big data analytics, and business intelligence have drawn increasing attention in academia and industries (Coronel, Morris, & Rob, 2020) (Russell & Norvig, 2010) (Sun Z. , 2019) (Ghavami, 2020) (Delena & Demirkanb, 2013). Entity relationship (E-R) modeling is an important technique for converting the physical world to a logical world through modeling entity relationships (Chen P. P., 1976) (Coronel, Morris, & Rob, 2020) (Hu, Liu, & Wang, 2020). E-R modeling is a foundational technology of data science, big data analytics, business intelligence, and database systems (Thalheim, 2000) (Coronel, Morris, & Rob, 2020). Business rules are fundamental for E-R modeling, database design, and big data analytics (Ghavami, 2020, p. 13). Many books on data management systems and big data analytics such as (Coronel, Morris, & Rob, 2020) and (Ghavami, 2020, p. 13) have mentioned business rules and discussed how to discover business rules. However, none have examined them in depth. Therefore, how to transform business cases through business rule discovery into E-R models is still a fundamental issue for a database design although many database management systems provide the built-in development tool of E-R diagram.

Without business rules as a basis for the comprehension of business cases, there would be neither data science and big data analytics nor business intelligence. Business case mining is business rule discovery from a business case, just as data mining is knowledge discovery from a large database (Kantardzic, 2011) (Ghavami, 2020). Identifying, analyzing, and refining the business rules are the first important tasks for designing relational database systems in particular and information systems and big data analytics in general (Coronel, Morris, & Rob, 2020, p. 433) (Ghavami, 2020). However, there is not a number of research publications in this direction based on our search using Google Scholar. Therefore, a big issue for designing database systems, big data analytics, and business intelligence is identified as:

1. How can we discover business rules from mining business cases?

Moreover, when we undertake E-R modeling, we usually assume that the business rules are already there or that business rules automatically exist. This is, however, not a reality in many instances. Even though one has rich experience in undertaking E-R modeling, one still does not have a unified method for dealing with business rules. There is also no

automated method for business rule discovery from mining business cases, although data mining and text mining have been developed for a few decades (Kantardzic, 2011) (Sharda, Delen, & Turba, 2018). Therefore, the second big issue for database design is identified below.

2. How can we mine business cases for business rule discovery?

E-R modeling optimization aims to improve the E-R modeling process to get better E-R diagrams that understand and translate the business cases properly. Google Scholar search for “E-R modeling optimization” found only one research publication (see Section 5). This implies that the third big issue for database design is:

3. How can we optimize E-R modeling?

This article will address these three issues by exploring business case mining and E-R modeling optimization. To address the first two issues, this article reviews case-based reasoning, explores business case-based reasoning, and presents a unified approach to business case mining for business rule discovery. To address the third issue, this article proposes a unified optimal method for E-R modeling.

The remainder of this article is organized as follows: Section 2 reviews case-based reasoning and explores business case-based reasoning. Section 3 looks at business case mining for business rule discovery. Section 4 examines E-R modeling optimization. Section 5 provides a discussion of related work and implications. The final section ends this article with some concluding remarks and future work.

It should be noted that this article limits its focus on the foundation of business case mining, business rule discovery, and E-R modeling optimization, even though the proposed approach will bring about impacts on many fields such as data science, machine learning, artificial intelligence, big data analytics, and cloud computing.

2. Business Case Based Reasoning

This section reviews case-based reasoning and explores business case-based reasoning, motivated by the business case below.

2.1 A Business Case

There are many definitions for a business case. Different disciplines have different definitions. For example,

1. Oxford defines a business case as “a justification for a proposed project or undertaking on the basis of its expected commercial benefit.”
2. A business case is a collection of the minimum information required to facilitate a decision to perform some activities. (<https://brainmates.com.au/brainrants/what-is-a-business-case/>, accessed on April 27, 2021)

From a case-based reasoning viewpoint (Sun, Han, & Dong, 2008), a business case can be defined as a description of a business scenario that includes a set of business activities and associated stakeholders. The following is a business case.

The Beem (a fictional name) Medical Center (for short, BMC) will design a database system to manage their patient appointments and billing system based on the following basic description of the system.

BMC has employed many doctors. A doctor may be scheduled for many appointments. However, each appointment is only with a single doctor. BMC records details of every doctor’s ID, first name, and surname.

A patient may have many appointments on the same day. The patient’s details include her or his first name, surname, date of birth, and the policy identification number of any insurance. BMC records the details and addresses of every patient on their system. Every patient may have multiple phone numbers including home, mobile, and work phone.

BMC records the details of every insurance company on their system, including the name of every insurance company, its postal address, a single phone number, and a webpage address. Every insurance company can have many different policy types. Each policy type includes a description of the policy and the percentage rebate amount that the insurance company will pay for each appointment.

A bill is generated after each appointment. For patients who have medical insurance, referred to as private patients, the insurance company may pay a certain percentage of a bill depending on the type of coverage that the patient holds. The patient and the insurance company may make multiple payments on a single bill. When a payment is made, a payment receipt number is recorded, along with the amount that has been paid, the date/time of the payment, the method of payment, and who has made the payment. The system records the different payment methods that are available including cash, credit card, cheque, and direct debit.

When a bill for an appointment is generated and if the patient has insurance, the amount of the bill that is covered by the insurance company is stored, along with an amount that is not covered by insurance. The status of each bill is also stored; these being: outstanding or paid in full.

The above business case leads to a model for business case mining driven E-R modeling optimization, illustrated in Figure 1. We will look at this model by reviewing case-based reasoning and exploring business case-based reasoning, business case mining, and E-R modeling optimization in the following sections.

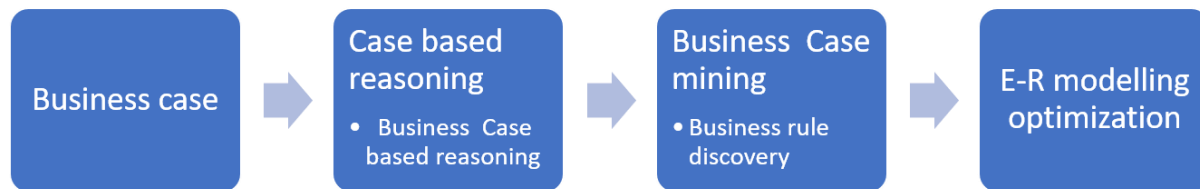


Figure 1. A model for business case mining driven E-R modeling optimization

2.2 Case Based Reasoning

Case-based reasoning (CBR) has drawn significant attention in artificial intelligence (AI), information systems, e-commerce, and data science with many successful applications in customer support, online sales, e-commerce, multiagent systems, and big data management (Sun, Han, & Dong, 2008) (Laudon & Laudon, 2016) (Weber, 2020).

CBR is a reasoning paradigm based on experiences and cases. A case is a description of previous experience that consists of the encountered problem and the successful solution to it (Sun & Finnie, 2004; 2010; 2013, p. 13). That is, CBR is a form of experience-based reasoning that solves new problems by adapting solutions that were used to solve old problems. This implies that CBR is a kind of machine learning (Russell & Norvig, 2010) (Sun, Han, & Dong, 2008) (Laudon & Laudon, 2016).

In business activities, it is usually true that “Two cars with similar quality features have similar prices”. This is a kind of experience-based reasoning, essentially, a kind of similarity-based reasoning (Sun, Han, & Dong, 2008). In other words, CBR can be considered as a kind of similarity-based reasoning.

CBR: = Similarity-based reasoning

Therefore, CBR is a kind of machine learning and similarity-based reasoning. The philosophy of CBR is that similar problems have similar solutions (Sun & Finnie, 2004; 2010; 2013) (Ghavami, 2020, p. 91).

2.3 Business Case Based Reasoning

Business case-based reasoning is a reasoning paradigm for analyzing and mining the business case and discovering business rules in particular and knowledge and intelligence in general.

Generally, we can use case-based reasoning to mine and analyze business cases. For example, case retrieval, case revise, case reuse, case retention, and case partition of CBR can be used to analyze and process the business case for assisting business decision making and improving business performance (Sun & Finnie, 2004; 2010; 2013) (Finnie & Sun, 2003).

Discovering knowledge and intelligence from mining business cases is similar to data mining, text mining, and knowledge discovery from a large database or big data (Kantardzic, 2011) (Ghavami, 2020). Therefore, business case mining is a special business case-based reasoning that mines the business case and discovers business rules for E-R modeling and database design. In this context, business rules are a kind of knowledge (Kantardzic, 2011). Therefore, business rule discovery is a kind of knowledge discovery. In the following section, we will look at business case mining.

3. Business Case Mining for Business Rule Discovery

Business case mining is fundamental for discovering detailed information on entities and their attributes as well as their relationships, the latter are the foundations for E-R modeling and database systems design. This section examines business case mining for business rule discovery.

Discovering business rules is a complex task for which many approaches have been proposed including analysis and extraction from code and data mining (Gailly & Geerts, 2013). For example, Gao, Koronios, Kennett, and Scott look at business rule discovery through data mining methods (Gao, Koronios, Kennett, & Scott, 2010). Bajić-Bizumić et al proposed an interactive, simulation-driven approach for discovering business rules with an Alloy Analyzer tool (<http://alloy.mit.edu/alloy/>). Alloy Analyzer is used as a platform for rule simulation and discovery (Bajić-Bizumić, Rychkova, & Wegmann, 2013).

Business case mining aims to mine the business case to discover business rules towards E-R modeling and database design.

3.1 Entities and Relationships

Entities, attributes, and relationships are the three main components of an E-R model (Coronel, Morris, & Rob, 2020). We will not go into attributes in this research, just as Chen did (Chen P. P., 1976).

An entity is a “thing” that can be distinctly identified (Chen P. P., 1976). A doctor, a patient, or a bill are examples of an

entity. An entity at the E-R modeling level refers to an entity set, denoted by capitated letters. For example, entity DOCTOR is an entity set corresponding to a table that consists of a number of doctor's records, that is, $DOCTOR = \{x \mid x = \text{a doctor working at BMC}\}$.

A relationship is an association among entities (Chen P. P., 1976). For instance, an appointment is a relationship between two entities, DOCTOR and PATIENT. Generally speaking, n entities have a relationship, where n is a natural number. However, in the following, we only consider binary relationships between two entities. At the E-R modeling level, we mainly consider three kinds of connectivity or the relationship classification: one-to-one (1:1), one-to-many (1:M), and many-to-many (M:N) relationships (Coronel, Morris, & Rob, 2020).

Have a look at the following excerpt from the above-mentioned business case:

A doctor may be scheduled for many appointments. However, each appointment is only with a single doctor. BMC also stores details of every doctor's ID number, first name, and surname.

In this business case, doctor, appointment, and patient correspond to three entities, DOCTOR, APPOINTMENT, and PATIENT. A doctor may be scheduled for many appointments, which is a kind of 1:M relationship between DOCTOR and APPOINTMENT. *A patient may have many appointments on the same day*, which is a 1:M relationship between PATIENT and APPOINTMENT.

The above discussion implies the following findings:

- Every relationship between entities is bidirectional. It is better to use one verb to represent these two bidirectional relationships. One is used with the active voice while another is with the passive voice.
- In the business case, a doctor may be scheduled for many appointments with patients whereas a patient may be scheduled for many appointments with doctors. Therefore, the relationship between DOCTOR and PATIENT is M:N.
- An entity is a noun while a relationship is usually represented by a verb associated with a noun, for example, appoint is associated with the appointment.

Therefore, different verbs in the business case may lead to different relationships. The question is: how many verbs are related to business, business management, and business decision making. All these related verbs are useful to discover business rules from mining the business case to develop business rule-driven E-R diagrams.

Furthermore, intransitive verbs are not used for creating relationships between entities. Therefore, if all the business and management-related verbs are denoted as V , then, intransitive verbs are not in V .

Based on the above analysis, in order to develop business rules-driven E-R modeling, it is necessary to extract triple elements from mining a business case, that is, the nouns that are related to persons, objects, things, etc. Each of these nouns may be an entity, and all these entities are aggregated as a set, denoted as E , consisting of business entities. Every element in E is a candidate element of a business rule as a pre-condition or a sequence of the business rule. They are also the candidate entity for E-R diagrams through E-R modeling.

The business-associated verbs extracted from the business case are aggregated as a set of business verbs, denoted as V .

The numbers related to quantity and restraints are aggregated to be a set of cardinality and restraints, denoted by C , which we do not go into detail about in this research.

Formally, we have

$$E = \{x \mid x = \text{an entity}\} \quad (1)$$

$$V = \{v \mid v = \text{a business-associated activity}\} \quad (2)$$

Hu, et al, proposes two guidelines for identifying the objective uniqueness of the entity relationship (Hu, Liu, & Wang, 2020) (Chen, 1976).

- Rule 1: identifying an entity that is generally a noun object.
- Rule 2: identifying the relationship which is associated with a verb.

Rule 1 is equivalent to Equation (1) (Chen, 1976). However, a noun object is a necessary condition but is insufficient for becoming an entity. Rule 2 does not correspond to (2). Furthermore, Rule 2 is not always true, because the appointment in our business case is a relationship between Doctor and Patient, but is represented originally in the form of a noun rather than a verb. This is the reason why we mentioned a business-associated activity rather than a verb in (2). Finally, a business relationship usually is a representation of business activity. For example, appointment and payment in our business case are two business activities or processes. They associate with the verbs appoint and pay respectively.

The above-mentioned business case is mined and the following candidate entities and business-associated activities are

discovered:

$$E = \{doctor, ID, first\ name\ and\ surname, appointment, patient, address, phone, insurance\ company, system, policy, description, percentage, amount, bill, coverage, payment, method, cash, credit\ card, cheque, direct\ debit, insurance, status, home, mobile\ and\ work, name, surname, date\ of\ birth, policy\ identification\ number\}.$$

$$V = \{manage, billing, store, schedule, include, have, generate, hold, make, pay, record, cover\}.$$

Named-entity recognition (or Name's entity recognition) is a procedure to identify entities of interest such as persons, locations, organizations, and products (Ghavami, 2020, p. 73). Its input is a text, its output is entities of interest. Apache OpenNLP (<https://opennlp.apache.org/>, accessed on June 3, 2021) is a platform that supports not only named entity extraction but also the most common natural language processing tasks such as tokenization, sentence segmentation, part-of-speech tagging, and parsing. One can use OpenNLP to get the above-mentioned E . However, the above results are blind mining because only nouns and most business-associated activities are discovered. In other words, every element in E is only a candidate entity. Every element in V can be a candidate relationship. We should mine the business case based on people-centered entity discovery, and function-centered entity discovery to find the entities. People-centered entity discovery is to extract the people or organization elements from E . Function-centered entity discovery is to extract the business activity-centered elements from V .

Using people-centered entity discovery, we have $E = \{DOCTOR, PATIENT, INSURANCE\ COMPANY\}$.

The mentioned business case aims to “design a database to manage their patient appointments and billing system”. This implies that the system has mainly two functions or business activities: appointment management and billing system. The billing system is related to payment and bill directly. That is, using function-centered entity discovery, we have $E = \{APPOINTMENT, PAYMENT, BILL\}$.

Combining people-centered entity discovery and function-centered entity discovery, we have discovered six entities from the business case mining.

$$E = \{DOCTOR, PATIENT, INSURANCE\ COMPANY, APPOINTMENT, PAYMENT, BILL\}.$$

Similarly, we have $V = \{appoint, bill, schedule, have, generate, hold, make, pay, record, cover\}$.

It should be noted that the business activity *appointment* is related to *schedule*, while the *payment* is related to *generate, make and cover*. All these will be discussed in the business rule discovery.

3.2 Business Rules

Business rules have been studied from various perspectives such as an IS perspective (Bajec & Krisper, 2001) and an ontology perspective (Gailly & Geerts, 2013). Automated business rule management systems have been available in the market (Bajec & Krisper, 2001). In what follows, we discuss business rules from the perspective of E-R modeling and database design (Coronel, Morris, & Rob, 2020).

Generally speaking, a business rule is a statement that specifies policies, conditions, and knowledge in business (Bajec & Krisper, 2001). More specifically, a business rule is a brief, precise, and unambiguous description of a policy, procedure, or principle within a business organization (Coronel, Morris, & Rob, 2015, p. 37-39). For E-R modeling and database design, business rules are the basis for defining entities and relationships as well as constraints for organizations' operations. Business rules are statements that specify entities and their relationships in the form of 1:1, 1:M, and M:N (Chen P. P., 1976) (Coronel, Morris, & Rob, 2020). That is, a business rule mainly consists of four elements below:

- Entities;
- Relationships (1:1, 1:M, M:N, expressed through connectivity and cardinalities);
- Attributes;
- Constraints.

The business rules can enable the designer to fully understand how the business works and what role data plays within company operations. Business rules are derived not only from the practice of business organizations but also from a business description of organizations' operations. In our research, business rules are hidden in the business case. Business rules are discovered from the business case mining. Therefore, business case mining aims to discover business rules with 1:1, 1:M, and M:N relationships between entities under either an optional or mandatory condition from the business case.

For example, in our business case,

A doctor may be scheduled for many appointments

is a business rule, which is a summarized principle based on the business practice of many medical centers. This business rule reveals that a relationship between doctor and appointment is one-to-many (1:M). Hence, DOCTOR is an entity and

APPOINTMENT is another. The sentence from our business case below:

each appointment is only with a single doctor

is another business rule that is an unambiguous description of the relationship between DOCTOR and APPOINTMENT. This relationship is one-to-one (1:1).

Business rules can establish entities, relationships, and constraints (Coronel, Morris, & Rob, 2020). For example, the above-mentioned first business rule describes the relationship between DOCTOR and APPOINTMENT. The second business rule describes the constraint of the appointment at the medical center.

3.3 A Unified Process of Mining Business Cases for Business Rule Discovery

A unified process of mining business case for business rule discovery consists of the following six steps.

Step 1. Identify and construct the set of entities E , using people-centered entity discovery, and function-centered entity discovery from mining the business case.

Step 2. Identify and construct the set of relationships V , and the set of constraints C from mining the business case, using people-centered relationship discovery, and function-centered relationship discovery.

Step 3. Build mapping between elements of E taking into account the element of V .

These steps are an iterative process (Coronel, Morris, & Rob, 2020), like the agile process of software development (Pressman & Maxim, 2014).

All the mappings over E together constitute a set of business relationships, denoted as R , that is,

$$R = \{r \mid r = (e_i, v_k, e_j), e_i, e_j \in E, v_k \in V\} \quad (3)$$

Note that any verb describing a business relationship cannot exist independently and must associate with at least one-entity.

Now we look at these three steps using an example. If doctor and appointment are in E , and schedule in V , then there is a candidate relationship between doctor and appointment. This candidate relationship might be *schedule*, that is,

$$r = (\text{doctor}, \text{schedule}, \text{appointment})$$

Furthermore, taking into account the connectivity of association relationship 1:1, 1:M, and M:N (Coronel, Morris, & Rob, 2020, p. 122), we can represent the above formula in (3), $r = (e_i, v_k, e_j, \text{connectivity})$, for example, $r = (\text{doctor}, \text{schedule}, \text{appointment}, 1:M)$.

Step 4. Traverse the business case iteratively and look at which elements $r = (e_i, v_k, e_j)$ of R most fit the business case and then select these most fitting elements to form a set of R . We still use R as the business relationships of the business case.

This step is related to optimization because it is related to “most fitting”.

Step 5. Attach the connectivity and cardinality (Coronel, Morris, & Rob, 2020, p. 122) to the element of R taking into account the business case in the form $r = (e_i, v_k, e_j, \text{connectivity}, \text{cardinality})$.

Step 6. Use the business rule definition language to translate $(e_i, v_k, e_j, \text{connectivity}, \text{cardinality})$ into a natural language description or in a semi-structured language.

Finally, we have the following set of business rules with respect to the business case.

$$(e_i, v_k, e_j, \text{connectivity}, \text{cardinality}).$$

Coronel, Morris, and Rob suggested the following method for developing business rules (Coronel, Morris, & Rob, 2020):

1. Create a detailed narrative of the organization’s description of operations
2. Identify business rules based on the business descriptions
3. Identify main entities and relationships from the business rules

Item 1 corresponds to the business case mentioned in 2.1. Item 2 corresponds to Steps 1-6. Item 3 corresponds to Steps 1-3. Therefore, the above-proposed unified method is an abstraction and extension of what Coronel, Morris, and Rob mentioned (Coronel, Morris, & Rob, 2020).

In the next section, we will illustrate the proposed unified process of business case mining for business rule discovery.

3.4 Business Rule Discovery from Base Case Mining

Business rule discovery aims to analyze information about organization units (Bajec & Krisper, 2001). Bajec & Krisper stated that the entire business rule life cycle consists of six stages: business rule discovery, analyses, classification,

articulation, formalization, and documentation. Each of the stages is related to E-R modeling and database design. Business rule discovery in this research consists of people-centered, function-centered, and people and function combined entity/relationship/business rule discovery. In other words, when the business case is mined, people-centered, function-centered entities, relationships, and business rules will be discovered. Then they will be combined to obtain the business rules for developing E-R diagrams. Based on this principle, the business case is mined and all possible corresponding business rules are discovered sequentially as follows:

Based on people-centered entity discovery, DOCTOR and PATIENT are the most important entities in the business case.

“*A doctor may be scheduled for many appointments, however, each appointment is only with a single doctor*” is extracted from the business case. Then the relationship between doctor and appointment is 1:M optionally, that is:

Business rule 1. *A doctor may be scheduled for many appointments.* The relationship between DOCTOR and APPOINTMENT is 1:M optionally.

Note that doctor belongs to people, therefore, this business rule discovery belongs to people-centered business rule discovery.

“*A patient may have many appointments*” is extracted from the business case. Then we have

Business rule 2. *A patient may have many appointments.* The relationship between PATIENT and APPOINTMENT is 1:M optionally.

Business rule 1 and Business rule 2 implies that

Business rule 3. The relationship between DOCTOR and PATIENT is M:N optionally (Coronel, Morris, & Rob, 2020).

In other words, one doctor might see many patients via appointments. One patient might see many doctors via appointments. APPOINTMENT (corresponding to APPOINT) is a composite entity that decomposes Business rule 3 into Business rules 1 and 2 (Coronel, Morris, & Rob, 2020).

“*BMC records basic details of each insurance company on their system, including the name of the insurance company, their postal address, a single phone number, and a webpage address.*” is extracted from the business case. This implies that INSURANCE COMPANY is an entity, which directly associates with the patient as an entity, that is, a patient might associate with many insurance companies.

Business rule 4. A patient may associate with many insurance companies. The relationship between PATIENT and INSURANCE COMPANY is 1: M optionally.

“*Each insurance company can have a number of different policy types. Each policy type includes a description of the policy and the percentage rebate amount that the insurance company will pay for each appointment*” is extracted from the business case. This implies that

Business rule 5. *The insurance company will pay for each appointment.* The relationship between INSURANCE COMPANY and APPOINTMENT is 1:M optionally.

In reality, an insurance company has no direct relationship with appointments but with the bill and payment of an associated patient. Therefore, this business rule should be removed in the stage of the E-R Modeling optimization.

“*The patient’s details including her or his ... policy identification number of any insurance are stored within the database*” is extracted from the business case. This implies that the policy identification number of any insurance is an attribute of PATIENT and:

Business rule 6. *A patient may have many insurances.* The relationship between PATIENT and INSURANCE COMPANY is 1:M optionally.

“*The patient and the insurance company may make multiple payments on a single bill*” is extracted from the business case. This implies:

Business rule 7. *Multiple payments on a single bill.* The relationship between BILL and PAYMENT is 1:M optionally.

Business rule 8. *The patient may make multiple payments.* The relationship between PATIENT and PAYMENT is 1:M optionally.

Business rules 7 and 8 imply that the relationship between the BILL and PATIENT is M:N optionally, which leads to a new composite entity PAYMENT (Coronel, Morris, & Rob, 2020). This is true in theory. However, in our business case, we use Business rule 12 (see below).

Business rule 9. *The insurance company may make multiple payments.* The relationship between INSURANCE COMPANY and PAYMENT is 1:M optionally.

Business rule 8 and Business rule 9 implies that (Coronel, Morris, & Rob, 2020)

Business rule 10. The relationship between PATIENT and INSURANCE COMPANY is M:N optionally.

That is, one patient can have many insurance (Company) policies, and an insurance company can provide many patients with insurance policies.

“After each appointment, a single bill is generated” is extracted from the business case. This implies that

Business rule 11. Each appointment generates a single bill. The relationship between APPOINTMENT and BILL is 1:1.

Business rule 12. One bill is generated for one patient. The relationship between PATIENT and BILL is 1:1

Business rule 13. One patient can have many insurance (company) policies, and an insurance company can provide many patients with insurance policies. Therefore, the relationship between BILL (PATIENT) and INSURANCE COMPANY (insurance policy) is M:N, which leads to a new composite entity PAYMENT.

Therefore, the discovered all possible business rules from the above business case mining can be summarized in Table 1 below.

Table 1. The brief representation of discovered business rules

Business Rule No.	Entity	Relationship	Connectivity	Entity
1	DOCTOR	schedule	1:M	APPOINTMENT
2	PATIENT	schedule	1:M	APPOINTMENT
3	DOCTOR	see	M:N	PATIENT
4	PATIENT	associate	1:M	INSURANCE COMPANY
5	INSURANCE COMPANY	pay	1:M	APPOINTMENT
6	PATIENT	have	1:M	INSURANCE COMPANY
7	BILL	have	1:M	PAYMENT
8	PATIENT	make	1:M	PAYMENT
9	INSURANCE COMPANY	make	1:M	PAYMENT
10	PATIENT	have	M:N	INSURANCE COMPANY
11	APPOINTMENT	generate	1:1	BILL
12	BILL	pay	1:1	PATIENT
13	INSURANCE COMPANY	make	M:N	BILL

Noted that APPOINTMENT, BILL, and PAYMENT are three composite entities discovered from the business case mining. They are the core functions of the database system that are designed by BMC, corresponding to appoint, bill, and pay in the form of verbs or business activities.

From a viewpoint of people-centered business rule discovery, the primary relationships are between people, not between people and things. For example, DOCTOR and PATIENT have a primary relationship, whereas DOCTOR and APPOINTMENT have a secondary relationship because APPOINTMENT is a composite entity. However, from a viewpoint of function-centered business rule discovery, DOCTOR and APPOINTMENT have a primary relationship. Therefore, people-centered business rule discovery and function-centered business rule discovery must be combined to discover business rules from the business case mining.

4. E-R Modeling Optimization

There are many methods for developing an E-R diagram from business rules. For example, Coronel, Morris, & Rob suggested the following method for developing an E-R diagram (Coronel, Morris, & Rob, 2020):

- Develop the initial E-R diagram;
- Revise and review E-R diagram;
- E-R to relational mapping for database design the initial E-R diagram.

The drawback of the above-mentioned method is how to develop the initial E-R diagram. This section addresses this issue by presenting a unified optimal method for E-R modeling. The unified optimal method includes people-centered E-R modeling, function-centered E-R modeling, and hierarchical E-R modeling.

4.1 People-centered E-R Modeling

People-centered modeling means that the initial E-R diagram starts from a people-related entity or two. DOCTOR and PATIENT are two people-related entities discovered from our business case mining. Our initial E-R diagram includes DOCTOR and PATIENT and their M:N relationship (see Business rule 3 in Section 3). This M:N relationship has been converted into two 1:M relationships (see business rule 1 and business rule 2 in Section 3), which has been realized by the first E-R diagram, illustrated in Figure 2.



Figure 2. The first E-R diagram

Furthermore, INSURANCE COMPANY is a people-related entity, therefore, the next step is to add the INSURANCE COMPANY-related business rules to the first E-R diagram, taking into account Table 1. This implies that business rules 6 and 8 are added to the initial E-R diagram, as illustrated in Figure 3.

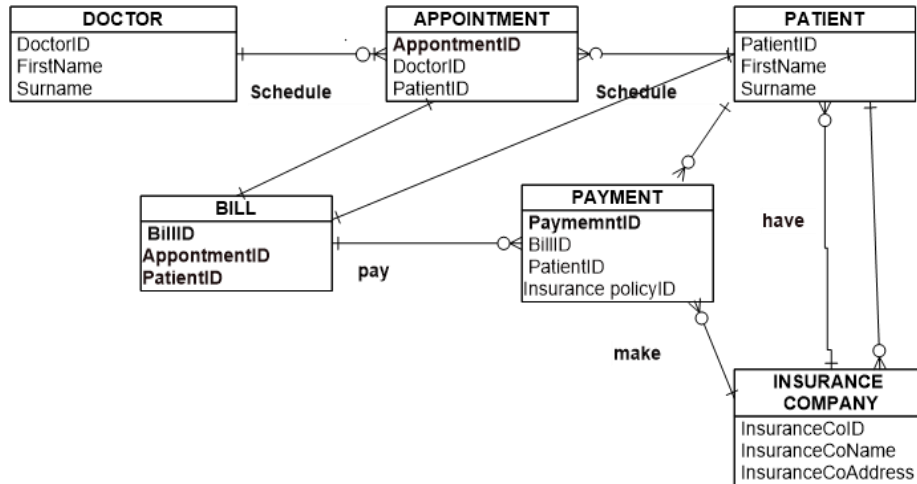


Figure 3. An initial E-R diagram-2

Similarly, we can continue the development of the E-R diagram using people-centered E-R modeling. We will now turn our focus on function-centered E-R modeling.

4.2 Function-centered E-R Modeling

Function-centered E-R modeling means that the initial E-R diagram starts from a functioning entity such as APPOINTMENT, BILL, or PAYMENT. We will continue our development of the initial E-R diagram by first adding BILL and then adding PAYMENT, as illustrated in Figure 4. Consequently, we have developed the initial E-R diagram based on the discovered 13 business rules.

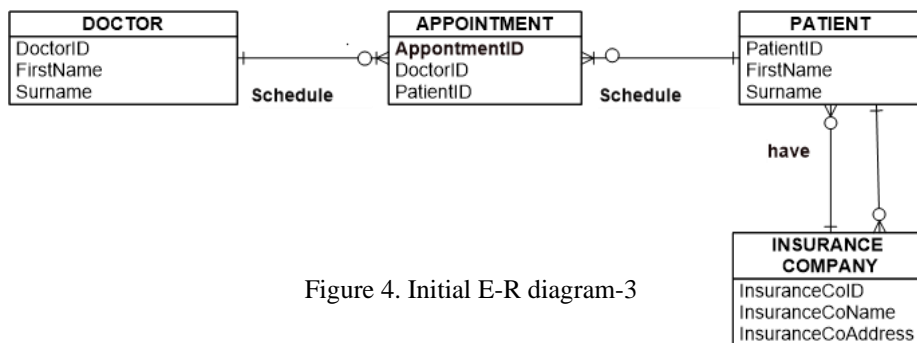


Figure 4. Initial E-R diagram-3

4.3 Hierarchical ER Modeling

From the discovered business rules in Section 3, the relationship between DOCTOR and PATIENT is M:N. We introduce a new composite entity: APPOINTMENT, which leads to two relationships: the relationship between DOCTOR and APPOINTMENT is 1:M; the relationship between PATIENT and APPOINTMENT is 1:M. This implies that hierarchical E-R modeling should be introduced (Pressman & Maxim, 2014). A hierarchical E-R modeling aims to develop first-level E-R diagrams, second-level E-R diagrams, and a detailed E-R diagram that includes the attributes of the entities (Coronel, Morris, & Rob, 2020). This research will not cover detailed E-R modeling.

4.3.1 First Level E-R diagram

The first level E-R diagram can be considered as a primary entities-based E-R diagram. At this level, we only focus on the primary entities and their relationships. Primary entities are only people-associated entities and people-related organization-associated entities. Therefore, the first level E-R diagram is a people-centered E-R diagram. Based on the discovered business rules from the business case mining, DOCTOR, PATIENT, and INSURANCE COMPANY are three primary entities. Therefore, we have the first level E-R diagram, illustrated in Figure 5, based on business rules 3 and 10 discovered in Section 3.4.

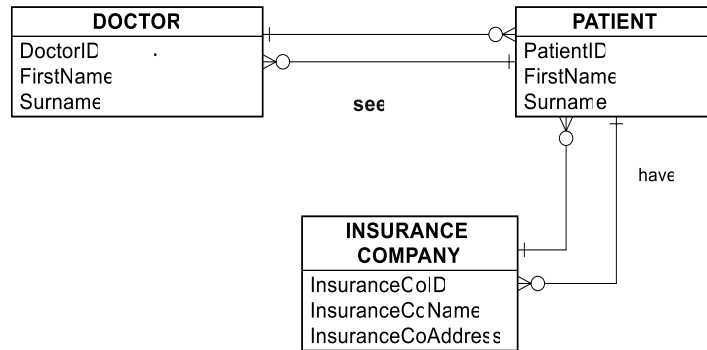


Figure 5. A segment of the first level E-R diagram

A key idea behind developing a first-level E-R diagram is to realize the principle of top-down, stepwise refinement used in software engineering (Pressman & Maxim, 2014). Therefore, the second level E-R diagram can be considered as a refinement of the first level E-R diagram.

4.3.2 Second level E-R Diagram

At the second level, we must first decompose the M:N relationship between primary entities by creating a composite entity and corresponding two 1:M relationships between primary entities and the composite entity. For example, we find that the M:N relationship between DOCTOR and PATIENT is to “schedule an appointment”. APPOINTMENT is then created to be a composite entity to decompose the M:N relationship into two 1:M relationships that occurred in business rules 1 and 2. Similarly, PAYMENT is created as a composite entity to decompose the M:N relationship between PATIENT and INSURANCE COMPANY into two 1:M relationships that occurred in business rules 8 and 9. The results are illustrated in Figure 6.

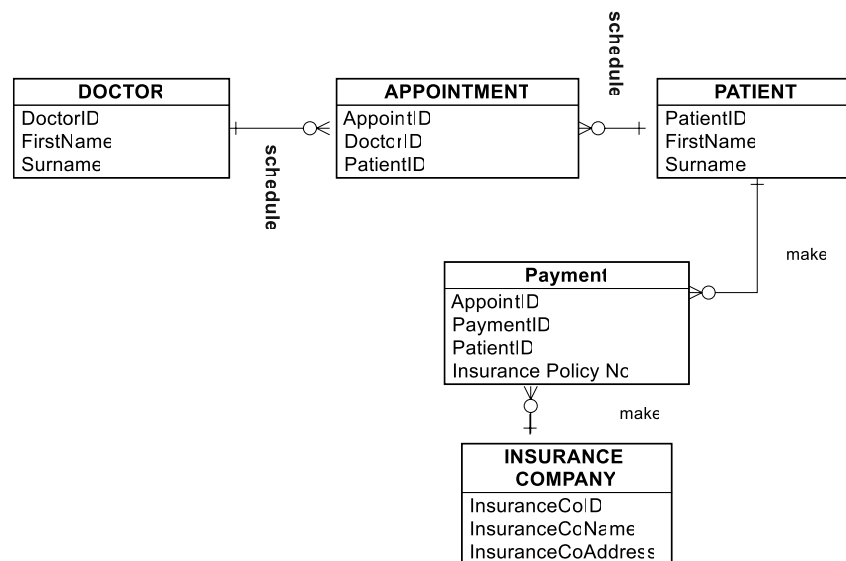


Figure 6. A second level E-R diagram

Figure 6 can be considered as a people-centered E-R diagram. Similarly, based on business rules 7, 9, and 13 discovered in Section 3, the relationship between INSURANCE COMPANY and BILL is M:N. This M:N relationship can be decomposed by creating the PAYMENT entity. Furthermore, BILL is a primary entity from a function-centered

development of an E-R diagram. This has been illustrated in Figures 3 and 7.

4.3.3 Finalizing E-R Diagram

After these two-level developments of the E-R diagram, we further consider business rule 12, discovered from “After each appointment, a single bill is generated” and add entity BILL to the latest updated E-R diagram. This entity is associated with business rules 7, 11, 12, and 13 in order. From the viewpoint of function-centered development of an E-R diagram, we first add entity BILL to link entity APPOINTMENT. Then we add the relationship between APPOINTMENT and BILL. Finally, we have an E-R diagram, illustrated in Figure 7.

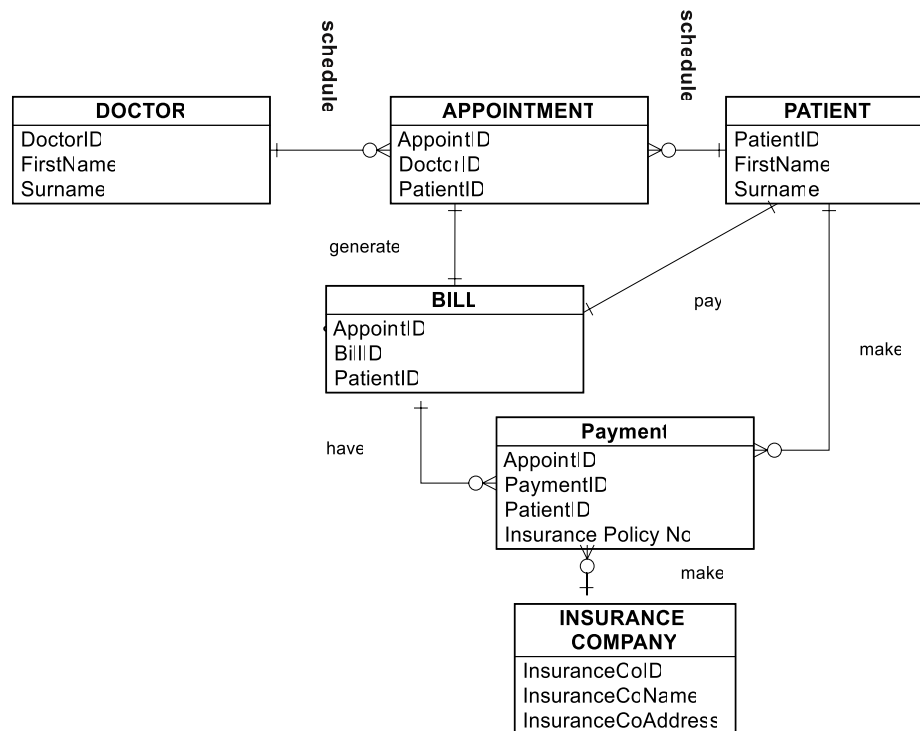


Figure 7. A final E-R diagram based on the discovered 13 business rules from BMC

It should be noted that the 1:1 relationship between BILL and PATIENT can be eliminated by adding BillID as an attribute in the entity PAYMENT (Coronel, Morris, & Rob, 2020). We leave it there in order to keep the nature of business case-based reasoning.

4.3.4 Merging 1:M relationships

Some consider the relationship between INSURANCE COMPANY and INSURANCE POLICY to be 1:M because an insurance company has many insurance policies. This is at a level. At a higher level, it is not necessary to introduce this 1:M relationship, instead, we have only one entity, INSURANCE COMPANY. The insurance policy is only an attribute of the INSURANCE COMPANY. This is also related to the development of hierarchical E-R diagrams. Similarly, in this research, phone and address are attributes of PATIENT. It is not necessary to introduce phone and address as entities either. This is also the reason why we propose people-centered and function-centered business rule discovery to discover primary entities.

More generally, for any 1:M relationship between two entities A and B, we can remove B (at M side) and put B as an attribute of A and keep A there, taking into account people-centered and function centered development of E-R diagrams. In other words, if entity A and B have 1:M relationship, then B can be merged into A, all B’s attributes can be considered as attributes of A after removing some repetitive attributes.

5. Discussion and Implications

We have mentioned several scholarly publications on business case mining and E-R modeling optimization. In what follows, this section will discuss the related work on business case mining, business rule discovery, E-R modeling optimization, based on big data-driven small data analysis (Sun Z. , 2019). It also examines the theoretical and technical implications of this research and its limitations.

5.1 Discussion

Google Scholar searches for “business case based reasoning” found 8 results (0.07 sec) (retrieved on July 3, 2021). However, these results are not related to business case-based reasoning as a keyword but are related to two keywords: One is business and the other is case-based reasoning. Case-based reasoning has become a research field since the end of the 1970s (Sun & Finnie, 2004; 2010; 2013). It is this research that makes business case-based reasoning the first to emerge in a research publication.

Google Scholar searches for “business case mining” found 16 results (0.08 sec) (retrieved on July 2, 2021). Preliminary analysis of the discovered results shows that no studies are related to databases nor related to E-R modeling. Even so, according to Zhan, business case mining is necessary for obtaining detailed information on knowledge nodes, business nodes, and logistic properties (Zhan, 2010). Our research demonstrates that business case mining is fundamental for obtaining detailed information on entities and their attributes as well as their relationships. The latter is the foundation for E-R modeling and database systems design. More generally, business case mining should be a part of text mining or text analytics (Sharda, Delen, & Turba, 2018). Text mining is a process of knowledge discovery from textual data. Google Scholar searches for “business rule text mining” found zero results (retrieved on June 4, 2021). This implies that text mining has not yet used to discover business rules.

Google Scholar searches for “business rule discovery” found 45 results (0.09 sec) (retrieved on April 27, 2021,). This implies that business rule discovery has appealed to academia to some extent. In particular, the following two research publications are worthy of some detailed analysis.

1. Gailly and Geerts presented a novel approach to generate business rules for an enterprise model based on the semantics of a domain ontology (Gailly & Geerts, 2013), while our business rule discovery is for E-R modeling and database design.
2. Bajec & Krisper discussed management of business rules in an enterprise (Bajec & Krisper, 2001). The activities for managing business rules in the enterprise include business rule discovery, business rule analysis, business rule consistency, and conflict validation, business rule modeling, and business rule implementation. How to manage each of the above-mentioned is still a big issue for business rule management in enterprises.

Google Scholar searches for “E-R modeling optimization” found only one research publication. In that publication, Hu, et al, proposed the four empirical rules for optimizing the E-R modeling based on their experience of teaching database systems (Hu, Liu, & Wang, 2020). We have mentioned and discussed the first two in Section 3. The other two empirical rules are as follows:

Rule 3. There must be no isolated entities in the E-R diagram.

Rule 4: The primary key of one entity set should not be an attribute of another entity set. The attribute implies a relationship between two entity sets.

The third rule is right because an E-R diagram should be a connected graph, from a graph theory viewpoint. The fourth rule seems not right for the E-R modelling because of the confusing attribute with the relationship. The attribute is a part of an entity (Coronel, Morris, & Rob, 2020).

5.2 Theoretical and Technical Implications

This research is the first endeavor for revealing business case mining and its impact on E-R modeling optimization. Just as data mining is fundamental for data science (Ghavami, 2020), so business case mining will play a pivotal role in the development of not only database management systems but also data science and big data analytics (Weber, 2020) (Ghavami, 2020). This is the first theoretical implication of business case mining.

The technical implication of this research is that the proposed approach on business case mining and E-R modeling optimization can appeal to more researchers and practitioners to develop automatic generators of business rules from business case mining to optimize E-R modeling and effectively design database systems.

5.3 Limitations

A limitation of this research is that it should integrate the proposed approach with advanced tools to provide a practical user guide to make business case mining and E-R modeling optimization become a data technique for database systems, big data analytics, and data science.

Another limitation of this research is that it should have provided more practical examples for readers to better understand the proposed approach. The third limitation is that we did not elaborate the development from the second level to the final E-R diagram owing to the space limitation.

6. Conclusion

As we mentioned at the beginning of this article, how to transform business cases through business rules into E-R models is a fundamental issue for a database design. This article provides a solution to this issue by highlighting three research questions:

1. How can we mine business cases for business rule discovery?
2. How can we discover business rules from mining business cases?
3. How can we optimize E-R modeling?

To address the first two research questions, this article reviewed case-based reasoning, explored business case-based reasoning and treated business case mining as a kind of business case-based reasoning. Then it presented a unified process of business case mining for business rule discovery, which has been underpinned by people-centered entity/relationship/business rule discovery, and function-centered entity/relationship/business rule discovery. To address the third research question, this article proposed a unified optimal method for E-R modeling. The unified optimal method includes people-centered E-R modeling, function-centered E-R modeling, and hierarchical E-R modeling. The approach proposed in this research might facilitate E-R modeling, database design, data science, and big data analytics.

The research demonstrated that business case mining can be considered as a special case of data mining. Business rule discovery is a special case of knowledge discovery. Therefore, business case mining and business rule discovery will play an important role in business intelligence, big data analytics, and data science. We will further delve into business case mining and business rule discovery in big data analytics. We will also develop a business rule description language (BRDL) to represent business rules in either logical form or semi-structured form.

There are at least two main relationships explored in database systems as its foundation. The first relationship is on entity-relationship, introduced by Peter Chen in 1976 (Chen P. P., 1976). The second relationship is the normalization of database tables, introduced by E.F. Codd (Codd, 1970). More generally, a relationship theory is a basis for data science and big data analytics. In future work, we will develop a relationship theory in database systems and big data analytics, which will underpin the development of database systems and data science.

7. Acknowledgments

The authors thank the editor and the anonymous reviewers for their constructive comments on the early manuscript of this article, which helped us to improve the manuscript. Dr. Zhaohao Sun would like to express his sincere gratitude to all his students in the universities of China, Australia, and Papua New Guinea (PNG) where he taught subjects of data systems, data management, intelligent systems, business intelligence and analytics as well as related subjects. The original ideas for this research were motivated in the class discussion with those students.

References

- Bajec, M., & Krisper, M. (2001). Managing business rules in enterprises. *Elektrotehnikski vestnik*, 68(4), 236-241.
- Bajić-Bizumić, B., Rychkova, I., & Wegmann, A. (2013). Simulation-Driven Approach for Business Rules Discovery. In Franch X., Soffer P. (eds) *Advanced Information Systems Engineering Workshops. CAiSE 2013. Lecture Notes in Business Information Processing, 148*, 111-123. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-38490-5_9
- Chen, P. P. (1976). The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1(1), 9-36. <https://doi.org/10.1145/320434.320440>
- Codd, E. F. (1970). A relational model of data for large shared data banks. *The Communications of ACM*, 13(6), 377-387. <https://doi.org/10.1145/362384.362685>
- Coronel, C., Morris, S., & Rob, P. (2020). *Database Systems: Design, Implementation, and Management (14th edition)*. Boston: Course Technology, Cengage Learning.
- Delena, D., & Demirkanb, H. (2013). Data, information, and analytics as services. *Decision Support Systems*, 55(1), 359-363. <https://doi.org/10.1016/j.dss.2012.05.044>
- Finnie, G., & Sun, Z. (2003). R5 model of case-based reasoning. *Knowledge-Based Systems*, 16(1), 59-65. [https://doi.org/10.1016/S0950-7051\(02\)00053-9](https://doi.org/10.1016/S0950-7051(02)00053-9)
- Gailly, F., & Geerts, G. L. (2013). Ontology-Driven Business Rule Specification. *Journal of Information Systems*, 27(1), 79-104. <https://doi.org/10.2308/isys-50428>
- Gao, J., Koronios, A., Kennett, S., & Scott, H. (2010). Business Rule Discovery Through Data Mining Methods. *Definitions, Concepts and Scope of Engineering Asset Management, 1*, 159-172. London, Scott: Springer. https://doi.org/10.1007/978-1-84996-178-3_9

- Ghavami, P. (2020). Analytics Techniques in Data Mining, Deep Learning and Natural Language Processing (2nd edition). *Big Data Analytics Methods*, Boston/Berlin: de Gruyter. <https://doi.org/10.1515/9781547401567>
- Hu, S., Liu, Y., & Wang, S. (2020). Teaching Exploration of Case-Based Data Modeling Optimization for Database System. *Open Journal of Social Sciences*, 8, 514-521. <https://doi.org/10.4236/jss.2020.83044>
- Kantardzic, M. (2011). Concepts, Models, Methods, and Algorithms. *Data Mining*. Hoboken, NJ: Wiley & IEEE Press. <https://doi.org/10.1002/9781118029145>
- Laudon, K. G., & Laudon, K. C. (2016). Management Information Systems: Managing the Digital Firm (14th Ed). Harlow, England: Pearson.
- Pressman, R., & Maxim, B. (2014). Software Engineering: A Practitioner's Approach 8th Edition. McGraw-Hill Education.
- Russell, S., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach (3rd edition). Prentice Hall.
- Sharda, R., Delen, D., & Turba, E. (2018). Business Intelligence and Analytics: Systems for Decision Support (10th Edition). Boston, MA: Pearson.
- Sun, Z. (2019). Intelligent Big Data Analytics: A Managerial Perspective. *Managerial Perspectives on Intelligent Big Data Analytics*, 1-19. USA: IGI-Global. <https://doi.org/10.4018/978-1-5225-7277-0.ch001>
- Sun, Z., & Finnie, G. (2004; 2010; 2013). A Case-based Reasoning Perspective. *Intelligent Techniques in E-Commerce*. Heidelberg Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-540-40003-5_2
- Sun, Z., Han, J., & Dong, D. (2008). Five Perspectives on Case Based Reasoning. LNAI 5227. ICIC2008, Shanghai (pp. 410-419). Berlin: Springer. https://doi.org/10.1007/978-3-540-85984-0_50
- Thalheim, B. (2000). Entity-Relationship Modeling Foundations of Database Technology. Berlin: Springer. <https://doi.org/10.1007/978-3-662-04058-4>
- Weber, H. (2020). Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data, and Machine Learning. USA: ICGtesting.
- Zhan, H. (2010). Research on Building Method of Knowledge Resource Cooperation Complex Network for Regional Manufacturing. *Applied Mechanics and Materials*, 37-38, 985-989. <https://doi.org/10.4028/www.scientific.net/AMM.37-38.985>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.