G Model
ARTMED-1421; No. of Pages 30

ARTICLE IN PRESS

Artificial Intelligence in Medicine xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

A fuzzy-ontology oriented case-based reasoning framework for semantic diabetes diagnosis

Shaker El-Sappagh[a], Mohammed Elmogy[b,*], A.M. Riad[c]

[a] Department of Mathematics, College of Science, King Saud University, PO 2455, Riyadh, Saudi Arabia
[b] Information Technology Department, Faculty of Computers & Information, Mansoura University, PO 35516, Mansoura, Egypt
[c] Information Systems Department, Faculty of Computers & Information, Mansoura University, PO 35516, Mansoura, Egypt

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form 2 June 2015
Accepted 5 August 2015

Keywords:
Case-based reasoning
Knowledge based system
Fuzzy ontology
Semantic retrieval
Diabetes diagnosis
Standard SNOMED CT terminology

ABSTRACT

Objective: Case-based reasoning (CBR) is a problem-solving paradigm that uses past knowledge to interpret or solve new problems. It is suitable for experience-based and theory-less problems. Building a semantically intelligent CBR that mimic the expert thinking can solve many problems especially medical ones.
Methods: Knowledge-intensive CBR using formal ontologies is an evolvement of this paradigm. Ontologies can be used for case representation and storage, and it can be used as a background knowledge. Using standard medical ontologies, such as SNOMED CT, enhances the interoperability and integration with the health care systems. Moreover, utilizing vague or imprecise knowledge further improves the CBR semantic effectiveness. This paper proposes a fuzzy ontology-based CBR framework. It proposes a fuzzy case-base OWL2 ontology, and a fuzzy semantic retrieval algorithm that handles many feature types.
Material: This framework is implemented and tested on the diabetes diagnosis problem. The fuzzy ontology is populated with 60 real diabetic cases. The effectiveness of the proposed approach is illustrated with a set of experiments and case studies.
Results: The resulting system can answer complex medical queries related to semantic understanding of medical concepts and handling of vague terms. The resulting fuzzy case-base ontology has 63 concepts, 54 (fuzzy) object properties, 138 (fuzzy) datatype properties, 105 fuzzy datatypes, and 2640 instances. The system achieves an accuracy of 97.67%. We compare our framework with existing CBR systems and a set of five machine-learning classifiers; our system outperforms all of these systems.
Conclusion: Building an integrated CBR system can improve its performance. Representing CBR knowledge using the fuzzy ontology and building a case retrieval algorithm that treats different features differently improves the accuracy of the resulting systems.

1. Introduction

Diabetes is a complex, chronic illness requiring continuous medical care with multifactorial risk-reduction strategies beyond glycemic control. According to World Health Organization (WHO), diabetes will be the seventh leading cause of death in 2030 [1]. Globally, about 336 million people are living with type 2 diabetes mellitus, and this figure is set to rise to over 552 million by 2030 [2]. In 2014, 9% of adults 18 years and older had diabetes [1]. There are three main types of diabetes. The first type is type 1 diabetes mellitus or insulin dependent diabetes mellitus; this type occurs when the pancreas cannot produce sufficient insulin. The second type is type 2 diabetes mellitus or insulin-independent diabetes mellitus; this type occurs when the body cannot effectively use the produced insulin. The third type is gestational diabetes, which occurs in pregnant women. A patient of diabetes symptoms but not really diabetic is called a pre-diabetes patient.

The early diagnosis of diabetes is critical in its care process because the early care can prevent long-term microvascular complications such as retinopathy, nephropathy and neuropathy, and cardiovascular disease. Moreover, the early diagnosis can prevent the pre-diabetes patient to become a diabetic. At present, the results for early detection of diabetes are not highly accurate. Therefore, there is a need to develop a diagnosis system for diabetes that has better accuracy. Clinical decision support systems (CDSS) can help in this regard. Existing rule-based diagnose diabetes systems are mainly based on the A1C criteria or plasma glucose criteria, either the fasting plasma glucose (FPG) or the 2-h plasma glucose

* Corresponding author. Tel.: +0020 1098889791; fax: +0020 502223754.
E-mail address: melmogy@mans.edu.eg (M. Elmogy).

http://dx.doi.org/10.1016/j.artmed.2015.08.003
0933-3657/

(2-h PG) value after a 75-g oral glucose tolerance test (OGTT). For example, they take decisions using rules such as if (A1C ≥ 6.5% or FPG ≥ 126 mg/dL or 2-h PG ≥ 200 mg/dL) then the patient is diabetic [3]. However, diabetes diagnosis is more complicated than these direct decisions. Diabetes is related to other diseases including renal diseases, heart diseases, foot diseases, etc. Moreover, it has symptoms related to hyperglycemia or hypoglycemia. The *true or false* decisions about these symptoms, e.g. thirst = true, is not enough.

Diabetes diagnosis is a theory-less and unstructured problem, and it depends on the physician's experience. For experience-based problem solving, case based reasoning (CBR) is one of the most suitable AI techniques for decision support [4]. CBR imitates human reasoning, and it is suitable when we cannot formulate a problem in a set of generalized rules. It is appropriate in a medical context where symptoms represent the problem, and diagnosis and treatment represent the solution. The CBR paradigm has been successfully used in various medical fields from lung disease and eating disorders to diabetes and Alzheimer's disease [5]. Many pieces of research utilized CBR for diabetes diagnosis [6–9]. Although any CBR system relies on a set of specific previous experiences, its reasoning power can be improved by general knowledge about the domain [10]. Ontologies can enhance the capabilities of CBR by creating knowledge intensive-CBR (KI-CBR) systems [11]. It can play many roles in CBR such as background domain ontology, case-base ontology, semantic similarity measurement, and others [12]. Ontology can enhance CBR systems in many dimensions, as shown in Fig. 1. In this figure, we suggest three types of KI-CBRs paradigms. In part (a) of Fig. 1, the case-base is stored in a traditional database, and the domain knowledge is stored in an ontology. In part (b), the case-base is stored in a crisp ontology, and the domain knowledge is stored in an ontology. In part (c), the case-base is stored in a fuzzy ontology, and the domain knowledge is stored in an ontology. We have selected the most complicated and recent approach (part c). For diabetes diagnosis, researchers made efforts toward diabetes ontology development [13]. Nevertheless, the literature of ontology-based CBR for diabetes is not rich with studies [7,8].

The most critical steps in CBR paradigm are the *case representation* and *case retrieval*. We concentrate on these two main steps to improve the performance of medical CBR. The case base building process reduces the efforts and time to build the system's knowledge base compared to rule-based systems. No generalized knowledge is required to build a successful CBR system. However, the collection of cases for patients requires the integration between the CDSS system and the distributed electronic health record (EHR) environment. As a result, the standardization of CBR knowledge and data is critical to achieving interoperability. Interoperability between EHR systems and CDSS facilitates the automatic collection of knowledge from patients' EHRs, supports the integration of CDSS in the healthcare environment, and eases the physician's querying process. EHR uses standards as Health Level 7s reference information model (HL7 RIM) [14] and systematized nomenclature of medicine-clinical terms (SNOMED-CT) [15], SCT for short, ontology for data storage and exchange, which can be utilized in CBR. RIM can be used as a standard case-base structure, and SCT can be used as background knowledge to enhance semantic retrieval [16,17]. El-Sappagh et al. [9] proposed a standard data model for diabetes case-base. SCT is a huge ontology, which affects the performance of the CBR retrieval algorithm. Creating a reference set from SCT for diabetes is required. El-Sappagh et al. [18] proposed a diabetes diagnosis OWL2 standard ontology from an SCT reference set. As far as we know, there are no studies utilize SCT reference sets in CBR systems for diabetes diagnosis, which is considered as a required issue for semantic retrieval and integration of CDSS in EHR environment. Using the created SCT-based OWL2 for semantic retrieval requires the encoding of the case-base unstructured knowledge with the same code. The encoding process is not a straightforward process, and it requires a methodology. El-Sappagh et al. [19] proposed an encoding methodology and utilized it to encode the case-base contents.

Physicians often describe patients using imperfect and linguistic data, and their knowledge and natural language have a great deal of imprecision and vagueness. As Zadeh [20] argued much of the knowledge that humans acquire through experience is perception-based and thus subject to imprecision and inaccuracy. Such knowledge, when not treated in some suitable way that can consider and convey its inherent imprecision, usually leads to the poor effectiveness of the knowledge-based systems that use it. As a result, KI-CBR paradigm must handle the imprecise knowledge representation and reasoning [21]. The existing fuzzy CBR systems utilize imprecise knowledge through the use of fuzzy logic for case representation and relevant fuzzy pattern matching techniques for similarity assessment [22]. A survey of existing systems of fuzzy CBR in diabetes diagnosis indicates that there are few works in this field. However, the lack of representation of this knowledge in ontological restricts the effectiveness of these systems because they did not take advantage of the reasoning capabilities that ontologies provide. The fuzzy ontology focuses on assigning a meaning to the fuzziness of the ontology's components. It is an important characteristic as it makes the fuzzy ontology's imprecision explicit, thus facilitating more efficient knowledge acquisition and ontology reuse. Moreover, it enables the definition of more effective semantic similarity measures, which facilitate case retrieval. For diabetes, the existing fuzzy CBR systems have not used fuzzy ontology or even crisp ontology as background domain knowledge or case-base ontologies [8]. On the other hand, ontologies and fuzzy logic have been utilized in diabetes in other reasoning methods such as rule-based expert systems [23].

In this paper, we present a fuzzy KI-CBR framework that handles and exploits imprecise knowledge through the effective integration of fuzzy logic in the ontology-based CBR paradigm. Fuzzy case-base ontology and a fuzzy semantic retrieval algorithm are proposed and integrated to build an intelligent CBR for diabetes diagnosis. This approach introduces fuzzy semantics to CBR in two places. The first is the representation of imprecise knowledge itself, and the second is case retrieval. In particular, our proposed framework

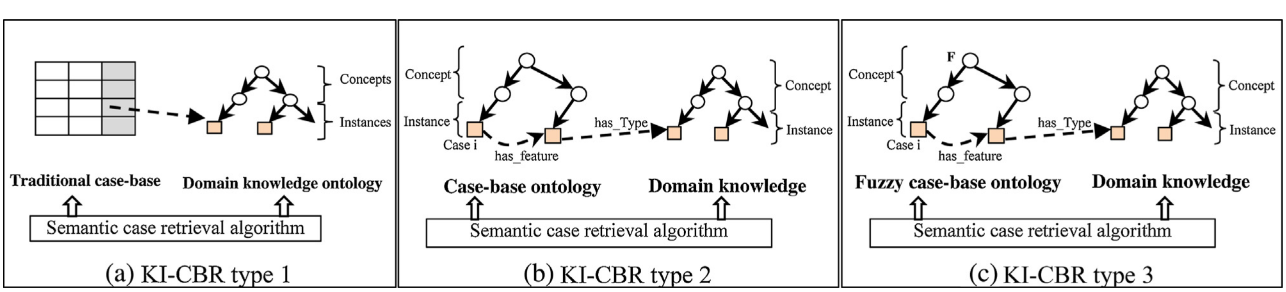

Fig. 1. KI–CBR frameworks.