Augmenting Medical Case Base Reasoning Systems With Clinical Knowledge Derived From Heterogeneous Electronic Patient Records

Syed Sbitc Raza Abidi and Selvakumar Manickam
Health Informatics Research Group, School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
{sbitc, selva}@cs.usm.my

Abstract

Development of medical Case Based Reasoning (CBR) systems necessitates an active involvement of medical experts. The featured work aims to minimize the involvement of medical experts in enhancing the knowledge content of medical CBR systems by using causal information contained in generic EPR as an alternate source of CBR compliant cases. We present an automated Case Acquisition and Transcription Info-Structure that features (a) an agent to proactively procure XML-based EPR from Internet accessible EPR repositories; and (b) a case generation methodology to automatically transform generic EPR to specialized CBR-compliant Clinical Cases (CC). EPR-CC transformation is achieved by establishing multi-level equivalence between the EPR and CC constructs—i.e. structural equivalence via metadata constructs, terminological equivalence via meta-thesaurus and conceptual equivalence via domain-specific ontologies. The transformed CC are intended to be seamlessly incorporated within CBR-based Medical Diagnostic Systems.

1. Introduction

Case-Based Reasoning (CBR) systems provide ‘analogy-based’ decision-support services to diagnostic problems by leveraging knowledge encapsulated in a corpus of previously experienced and solved real-life clinical problems (called Cases) [1, 2]. The diagnostic efficacy of medical CBR systems depend very much on the qualitative and quantitative richness of its knowledge repository—i.e. the Casebase (CB)—which is an aggregation of solved past clinical cases. Typically, such experiential clinical cases are sourced from medical practitioners, who record their clinical experiences—in terms of a list of clinical problems, diagnosis/prognosis and treatment plan—in a pre-defined case format [3]. We believe that such a knowledge engineering exercise has obvious drawbacks: (a) it imposes unrealistic time and operational constraints on medical practitioners; (b) the knowledge pool is limited to a set of medical practitioners; (c) it requires manual transcription of clinical cases to the native CB format; and (d) it requires standardization of clinical information to the CB information standard.

The featured work aims to minimize (but not eliminate) the involvement of medical experts in the CB enrichment lifecycle. This is achieved via a novel knowledge engineering strategy that purports the generation of CBR-compliant Clinical Cases (CC) from episodic Electronic Patient Records (EPR) [4]. We argue that EPR are an apt alternate source for the generation of specialized CC because EPR’s seem to encapsulate clinical problem-solving knowledge in terms of a causal relationship between the following information content:
• **Problem/situation description:** The physician-generated problem description—i.e., longitudinal patient history, illness signs and symptoms, pathological finding, etc.

• **Solution:** Physician-specified diagnosis and an optional treatment plan.

• **Outcome:** The physician-specified follow-up or rehabilitation-related information.

We present a Case Acquisition and Transcription Info-Structure (CATI) that features a multi-layer architecture comprising independent modules exhibiting the following [5]:

1. A mechanism to pro-actively seek and procure CBR-quality information from heterogeneous Internet-accessible XML-HL7 based EPR repositories.

2. A methodology to facilitate the automatic generation of CBR-system compliant cases derived from generic EPR. Our proposed CC generation methodology—leveraging both inductive and deductive techniques—advocates the establishment of structural, terminological and conceptual levels between the source EPR and the target CC representation standards.

3. A CBR engine to provide diagnostic support to clinical problems posed by the user either via a web-based or stand-alone interface.

![Figure 1. The functional architecture of CATI illustrating three task-specific layers.](image)

CATI is largely implemented using Java and is hosted on a server running Windows NT. The case-base and the intermediate database(s) are implemented using the Microsoft SQL Server 7.0 (MSSQL) database. Data exchange with donor EPR repositories is achieved using a JDBC interface. Information extraction from XML-HL7 based EPR is achieved using the Microsoft XML Parser.

**2. A Methodology for Automated Case Generation**

We believe that the transformation of EPR to specialized CC cannot be correctly achieved by a straightforward mapping of attribute-value pairs from the EPR to CC. We argue that such an approach would lead to complexities due to the presumed heterogeneous origin of EPR, whereby structural, terminological and conceptual differences may exist with respect to the a priori specified CC standards. For that matter, a correspondence between EPR and CC attribute-value pairs need to be established at the: (1) structural, (2) terminology, and (3) concept levels [6]. Our methodology, therefore, features the use of metadata constructs for structural
equivalence, meta-thesaurus for terminological equivalence and domain-specific ontologies for conceptual equivalence (see Figure 1). We explain below the three stages of case generation spanning from the establishment of structural, terminological and conceptual equivalence.

2.1 Structural Equivalence Via Metadata Maps

To determine structural equivalence between the EPR and CC attributes we propose the use of a CC metadata map describing the logical description of attributes [6]. For our purpose, EPR’s are regarded as structured XML documents, whereby their inherent structural definition is represented as Document-Type Definitions (DTD). Structural equivalence can be established at the logical level whereby the CC metadata map can be compared with the metadata map of EPR. Such metadata mediated structural equivalence will determine either a straightforward (one-one) or a more complex (many-one or many-many) mapping of EPR attributes to corresponding CC attributes. In CATI, structural equivalence is performed by the Structure Standardizer module.

2.2 Terminological Equivalence Via Meta-Thesaurus

The heterogeneous origin of the EPR leads to the inevitable usage of functionally synonymous terms to denote the same concept. Hence, it is necessary to establish terminology-level equivalence between the EPR attribute’s values with the standard CC vocabulary. We propose the use of domain-specific meta-thesaurus to achieve the necessary transformation of non-standard EPR attribute values to the standard CC vocabulary. In CATI, terminological equivalence is performed by the Terminology Standardizer module.

2.3 Conceptual Equivalence Via Ontologies

Concept-level equivalence of EPR attribute values with respect the standard CC concept vocabulary is necessitated when a terminological equivalence between the corresponding EPR and CC attribute’s values cannot be established. The presence of this situation implies that the EPR attribute’s value is not necessarily terminologically divergent, rather equivalence need to be established by considering the conceptual specialization/generalization of the EPR attribute’s value. In CATI, conceptual equivalence is performed by the Concept Standardizer module.

Figure 2. A system-generated ontology from the MeSH (MSH99) coding scheme. We illustrate the transversal path (shown in solid) for the concept Fever.
We propose the use of medical ontologies to establish conceptual equivalence via a two step process: (1) Determine the location of the EPR attribute's value in the specified medical ontology; (2) Establish conceptual equivalence by transversing the ontology—upward transversal yields a generalization whereas downward transversal yields a specialization of the concept (see Figure 2). We not only perform a one-one conceptual matching, but we also take into account transitive ontological matching i.e. an EPR attribute's value may match with any intermediate concept B which in turn is related to a legitimate CC value.

3. Exemplar EMR-CCS Transformation

Below we present an exemplar transformation of a generic EPR to a specialized CC based on our above-mentioned methodology. The data entities involved in EPR-CC transformation are as follows: (a) The set of CC Attributes (say CCA); (b) The range of allowed CC Values (say CCV) for each CCA; A set of EPR Attributes (say EA) and corresponding EPR Values (say EV). The EPR-CC transformation process is carried out according to the following scheme:

3.1 EMR-CC Structural Equivalence

Prior to initiating the structural equivalence exercise the acquired EA set is minimized by removing certain undesirable information, for instance patient name, address, next-of-kin, etc. Next, using the CC metadata map, each CCA is mapped to an EA. Two possibilities may exist: (1) an exact match between CCA and EA, e.g. the CCA referred to as 'symptoms' is mapped to the similar EA of 'symptoms'; or (2) an indirect match involving a structural/terminological mapping of a CCA to an EA, for example the CCA of 'diagnosis' is mapped to the corresponding EA of 'diagnostic outcome'. In Figure 3 we show the establishment of structural equivalence for an EPR in an XML-HL7 format; the underlined attributes in the intermediate CC are the ones that need to be standardized and in the right-hand frame we show the standardized CC.

3.2 EMR-CC Terminological Equivalence

In the second stage, we seek to convert the EV vocabulary to the pre-specified CC terminology standards. In Figure 4 we illustrate terminological equivalence whereby the EV available in Figure 3 are subsequently standardized to legitimate CCV.

We use the Unified Medical Language System (UMLS) meta-thesaurus to map EV to standard CCV. Two mapping possibilities exist: (1) an EV directly matches with the allowed list of CCV; and (2) an EV is indirectly equated to a synonymous CCV via the UMLS meta-thesaurus. For instance, the EV 'rapid heartbeat' is mapped to its synonymous CCV 'tachycardia'.

3.3 EMR-CC Conceptual Equivalence

Conceptual equivalence is achieved by employing two types of medical ontologies: (1) standard medical ontologies and (2) ontologies derived by us from medical coding schemes such as ICD10, MSH99 and so on (see Figure 2). Ontology mediated conceptual equivalence ensures that an EV is matched as a specialization or generalization of a standard CCV, for instance an EV 'orthopnea' can be mapped as a generalization of the CCV 'hypercapnia'. At the completion of the conceptual equivalence stage, all the EV are transformed to standard CCV, hence the right-hand frame in Figure 5 can be regarded as the final output—i.e. a functionally capable CC.
4. Concluding Remarks

A resource-intensive knowledge engineering exercise tends to undermine the incorporation of CBR systems to medical diagnostic-support tasks. The work presented here suggests an alternative to the practice of sourcing experiential information directly from expert medical practitioners [7]. We propose to leverage ‘information rich’ EPR, nowadays mostly accessible over the Internet, to enrich the ‘functional-knowledge’ of traditional medical CBR systems. In fact, for testing purposes, we have arranged for the proactive procurement of CBR-quality information from selected EPR repositories accessible via the Internet (of course subject to prior arrangement with the institutions holding the EPR repositories). This is achieved via an EPR procurement agent that proactively and periodically checks a pre-defined set of EPR repositories for ‘new’ case-quality data. Any addition, on a cohort basis, of CBR-quality data at these EPR repositories is reported to the EPR procurement module that initiates the download of the ‘new’ case-quality data to the EPR procurement server for subsequent transformation to CC.

The efficacy of our methodology was assessed by presenting the transformed CC (from a set of EPR) to medical experts for verification. For XML-based EPR containing information restricted to selected illnesses, the transformed CS was found to be technically correct and usable for down-stream decision-support activities. Notwithstanding the positive reviews from domain experts, we plan to conduct a broad range of tests to identify and rectify transformation inaccuracies, if any, resulting from our proposed EPR-CC transformation methodology. But at present the work is restricted due to two factors (a) the inability to acquire medical ontologies defining a broad range of medical knowledge spanning a variety of illnesses; and (b) the unavailability of diverse EPR structures to test the efficacy of the EPR-CC transformation methodology in varying operational conditions. Nevertheless, we believe that we have put forth an interesting utility of routinely collected EPR that may interest medical institutions seeking to operationalize their EPR towards the generation of diagnostic-support services. Currently, we are engaging various medical institutions to provide us ‘blind’ access to their EPR for testing purposes.

References