Artificial Intelligence in Medicine 51 (2011) 117–123

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

Integrating case-based reasoning with an electronic patient record system

Martijn van den Branden [a,b], Nirmalie Wiratunga [a,*], Dean Burton [b], Susan Craw [a]

[a] School of Computing, The Robert Gordon University, Aberdeen, AB25 1HG Scotland, UK
[b] AxSys Technology Ltd, Glasgow Airport Business Park, Paisley, PA3 2SJ Scotland, UK

ARTICLE INFO

Article history:
Received 7 January 2009
Received in revised form
20 September 2010
Accepted 7 December 2010

keywords:
Case-based reasoning
Electronic patient record systems
Clinical decision support

ABSTRACT

Electronic patient records (EPRs) contain a wealth of patient-related data and capture clinical problem-solving experiences and decisions. EXCELICARE is such a system which is also a platform for the national generic clinical system in the UK.
Objective: This paper presents, EXCELICARECBR, a case-based reasoning (CBR) system which has been developed to complement EXCELICARE. Objective of this work is to integrate CBR to support clinical decision making by harnessing electronic patient records for clinical experience reuse.
Methods: CBR is a proven problem solving methodology in which past solutions are reused to solve new problems. A key challenge that we address in this paper is how to extract and represent a case from an EPR. Using an example from the lung cancer domain we demonstrate our generic case representation approach where EXCELICARE fields are mapped to case features. Once the case base is populated with cases containing data from the EPRs database a standard weighted k-nearest neighbour algorithm combined with a genetic algorithm based feature weighting mechanism is used for case retrieval and reuse.
Conclusions: We conclude that incorporating case authoring functionality and a generic retrieval mechanism were key to successful integration of EXCELICARECBR. This paper also demonstrates how the application of CBR can enable sharing of lessons learned through the retrieval and reuse of EPRs captured as cases in a case base.

1. Introduction

Increasingly, clinicians are required to process a large quantity of information in addition to meeting in-patient treatment time constraints. A system which presents evidence to reinforce clinical decision making in real-time can assist clinicians with their daily routine enabling informed decisions. Such systems are also crucial for training purposes whereby junior doctors can benefit from repositories that capture previous decision making experiences of senior doctors [1]. Case-based reasoning (CBR) is a proven methodology that is used to reuse previous experiences to solve current problems [2]. This paper describes the development of a CBR decision support module that has been integrated within a commercial generic clinical system to provide decision support.

We introduce EXCELICARECBR, a CBR module that has been created and integrated within a commercial generic clinical system called EXCELICARE developed by AxSys Technology. EXCELICARE is an advanced electronic patient record (EPR) system and is currently in use in several sites across the national health service (NHS) in Scotland and England. These systems have largely replaced paper-based records in medical practices throughout the UK. The primary requirement for the EXCELICARECBR module is the need for generality and flexible configuration to provide support in any (existing) clinical process. In order to meet this demand, EXCELICARECBR incorporates a case authoring tool that allows a domain expert to map data items to a case representation followed by a generic retrieval mechanism that can operate on any clinical domain. A genetic algorithm is included to learn relevant features for the clinical case representation and reuse of previous clinician queries allows sharing of search knowledge in addition to standard experience reuse.

The next two sections describe the context of the current work in generic systems followed by the development of EPRs in the UK. An overview of the EXCELICARE system appears in Section 4. A formal discussion of the case representation and similarity assessment is discussed in Section 5. Architecture and integration of the CBR component within EXCELICARE is presented in Section 6. Retrieval and reuse functionality is discussed with radiotherapy treatment examples from lung cancer in Section 8 followed by conclusions in Section 9.

2. Related work in generic health care systems

In recent years the development of CBR applications in health care and medicine has progressed tremendously with successful

* Corresponding author. Tel.: +44 1224 26 2573; fax: +44 1224 26 2727.
E-mail addresses: martijn.vdbranden@axsys.co.uk (M. van den Branden), n.wiratunga@rgu.ac.uk (N. Wiratunga), dean@axsys.co.uk (D. Burton), smc@comp.rgu.ac.uk (S. Craw).

0933-3657/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2010.12.004

systems reported in the literature [3]. CBR methodology has been used as a problem solving strategy in a variety of health care domains ranging from planning of care in geriatric residences [4], hypothyroidism [5], hemodialysis [6], diabetes [7] and diagnosis of stress from bio-measurements [8] to explanation-driven support for bronchiolitis [9]. Although a few systems have been deployed, bridging the gap between the prototype stage and large scale commercialisation remains to some extent a challenge. Moreover, although CBR methodology is applicable to any (medical) problem, provided that it can be expressed using the CBR cycle, most systems seem to be more or less tied to specific clinical disciplines.

The work presented here enables the application of CBR methodology more generally, that is to say, without targeting a specific clinical problem. Although there are a few examples of generic CBR frameworks being developed (e.g. JColibri [10], myCBR [11] and BioVision [12]) these are predominantly targeted towards classification, recommender, textual CBR systems and medical image processing. In separate work CBR has also been recognised as a successful strategy to manage clinical guidelines, with case retrieval to identify deviations from standards and to form the basis for the revision of initial standards over time [13]. Although we do not concern ourselves with revision as such, we acknowledge the need to develop systems that adhere to generic guidelines. In particular we demonstrate the benefits of modelling the sequencing of activities within clinical pathways as a generic clinical process. The design of the case representation to capture this process is presented forming the focus of our contribution in this paper. However it is not hard to imagine that revisions at the level of guidelines could lead to revisions in activity sequencing thereby impacting on our generic clinical process.

3. Electronic patient records in the UK

The potential benefits of EPRs to health care have been widely recognised, in particular their ability to aid in the prevention of medical errors and to allow real-time access to patient data, integrated across several disciplines. This has lead to a gradual replacement of paper-based records in most clinical practices in the UK. The transition to EPRs is governed by different NHS bodies and carried out under programmes such as the national programme for information technology (NPfIT) in England and eHealth in Scotland. As part of the department of health information strategy to modernise the national health services, a publication in 2005 had identified six levels of maturity for EPR implementation:

1. Clinical administrative data.
2. Integrated clinical diagnosis and treatment support.
3. Clinical activity support.
4. Clinical knowledge and decision support.
5. Speciality specific support.
6. Advanced multi-media and telematics.

The first three levels broadly aim to define a unified view of patient data enabling successful integration across clinical disciplines, with support for complex care pathways. The main aim of the NPfIT initiative was to ensure a minimum level-three implementation of EPR within all acute trusts. However, due to the complexity of the programme, projects have suffered delays and difficulties and ongoing efforts are directed towards accomplishing this goal. The EPR level model has in some projects been thought of as too restrictive and has therefore been abandoned in favour of a more flexible approach that allows systems and implementation to be tailored to the needs of individual practices. Of course such ad hoc approaches make integration at national level harder.

AxSys Technology Ltd. is an information technology service provider who has successfully implemented clinical systems that operate across all six EPR levels to meet needs of individual clinical practices. Of particular interest to this paper are its efforts to create the generic EXCELICARE system to address level four requirements related to clinical knowledge and decision support in collaboration with the Robert Gordon University. CBR has been identified as the foundational methodology of the framework, because of its inherent ability to exploit previous experiences already captured in EPRs and allow these to be used as the basis for future clinical decisions.

4. EXCELICARE: a generic clinical system

EXCELICARE implements different aspects of EPR levels mentioned in the previous section in order to support clinical processes. In health care, a methodology often used to manage clinical processes is the clinical pathway (or Integrated Care Pathway). EXCELICARE enables clinicians to model patient care aligned with this approach. Within the clinical pathway, the sequencing of activities is modelled by the generic clinical process outlined in Fig. 1.

Typically, the patient enters the clinical process when a referral letter is issued. Subsequently, patient data is captured, appointments are made, letters are sent, lab tests are ordered and results are received, digital images are stored, a video conference is held between physicians to plan a patient's treatment etc. These are all examples of subtasks facilitated within EXCELICARE. If the patient's condition is chronic or if there is a need for follow-up after treatment, the process enters a Manage stage or cycle. The number of times this cycle is traversed before discharge is generally not known, and may be different for each patient.

EXCELICARE has been successfully deployed to support several clinical processes in the management of cancer, orthopaedics, infection control, sexual health and many others. Each implementation of EXCELICARE generally captures a large volume of data related to each patient and clinical decisions made during the traversal of a clinical process. This is a rich source of data that can be mined in order to provide real-time support to a clinician. The EXCELICARECBR module needs to be able to support any clinical decision that is part of a generic clinical process. In order to successfully integrate the EXCELICARECBR module within EXCELICARE, existing data structures in EXCELICARE need to be mapped to a case representation.

In a patient-centred EPR such as EXCELICARE, all captured data is ultimately related to a patient. This is illustrated by the root node in the simple hierarchy in Fig. 2. The EXCELICARE EPR uses *special forms* to capture data items related to a specific clinical concept or context. This allows Clinicians to design their own special forms to capture data at different stages in a clinical process. An example of the data items captured by 2 simplified radiotherapy forms is illustrated in the second level of the hierarchy in Fig. 2. The lowest level of the hierarchy consists of actual *instances* of captured data. The bottom right instance displays patient demographic data that is represented by a single instance of a form. Alternatively a form can have multiple instances associated with it. In Fig. 2, we illustrate multiple instances with the radiotherapy treatment where 2 instances for different radiotherapy courses have been administered to the same patient. Naturally instances are ordered and thereby capture trend information. For example, the site 1 total dose has increased from 8 to 17.

5. Case representation and similarity assessment

A case needs to capture both the multiple ordered instances of special forms from different stages in the clinical process, as well as single instance data that is only captured once for every patient.

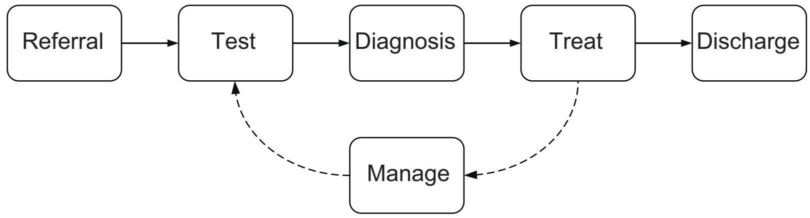

Figure 1. Different stages of a clinical process.

We achieve this in ExcelicareCBR by a case C formed by a set of groups G. Each group represents a *special form* as described in the previous section and is an ordered set of instances $I = (I_i)_{i \in \mathbb{N}}$ where $I_1 < I_2 < \ldots < I_i < I_{i+1}$. This ordering is influenced by time and as such i can be viewed as a rank assigned after sorting by time. Similarity between two cases C and C' is an aggregation over the constituent groups.

$$CaseSim(C, C') = \sum_j GroupSim(G_j, G'_j) \tag{1}$$

Similarity between singleton groups (containing a single instance) is easily established however groups with multiple instances require more sophisticated aggregation methods. One approach would be to create a single representative prototype instance for each group:

$$GroupSim(G, G') = Sim(\bar{I}, \bar{I}') \tag{2}$$

For example, a straightforward way to represent a numeric feature is to compute its average from the different instance values whereas a nominal feature could be represented by its majority value. Furthermore, this approach could be extended to capture *trends* within different groups. In a particular clinical context such as cancer treatment this may be valuable to identify chemotherapy treatments that show a trend with regards to tumour size. The implementation specifics are likely to be different depending on the requirements in the clinical context. However one difficulty with using prototypes is that skewed distributions within groups can result in lower retrieval precision.

An alternative approach is to compare and aggregate each ordered instance pair:

$$GroupSim(G, G') = \sum_i^n w_i \times Sim(I_i, I'_i) \tag{3}$$

Here n specifies the number of instances to be used when comparing the ordered sets. Similarity is computed by comparing the different instances of the clinical process stages. As such when the number of stages differ between two cases the comparison is constrained by the minimum of the two ordered set sizes. Alternatively a threshold could be set on how far back the system should go when comparing instances from these ordered sets. The weight value w_i ensures that similarity due to recent instances contribute more towards the final similarity aggregation.

6. ExcelicareCBR system architecture

The CBR component has been designed and developed from scratch and is integrated using a multi-tiered architecture (see Fig. 3). Existing Excelicare modules and components appear in grey whilst the rest are ExcelicareCBR components. The bottom layer in the figure shows the interface to different databases. The Excelicare database stores all patient-related data. The ExcelicareCBR case base (database) is hosted on a separate server thereby ensuring that performance of the primary EPR functions are not compromised. The case base will potentially hold a large number of cases gathered from all the hospitals within a trust or will even be used on a national level. New cases will be transferred from the Excelicare database to this case base and made available for retrieval.

The application architecture as presented in Fig. 3 follows the multi-tier model from software engineering, in which logically separate processes are decoupled. The following example illustrates the role of the different application tiers:

1. The user interacts with the system and requests a list of all cases for a particular case base (or domain). The interaction takes place in the user interface (UI) layer.
2. This request is passed on to the Web service layer, which acts as a mediator between UI and underlying components.

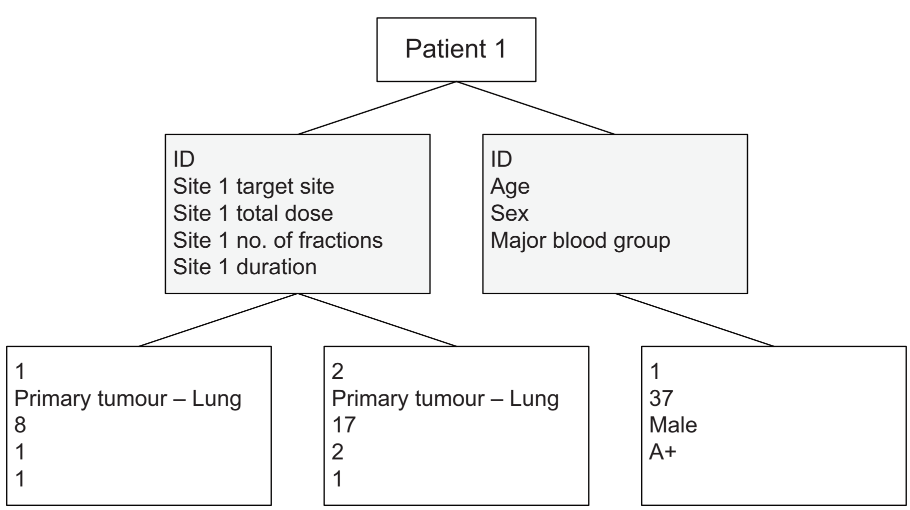

Figure 2. Simplified data structure in Excelicare.

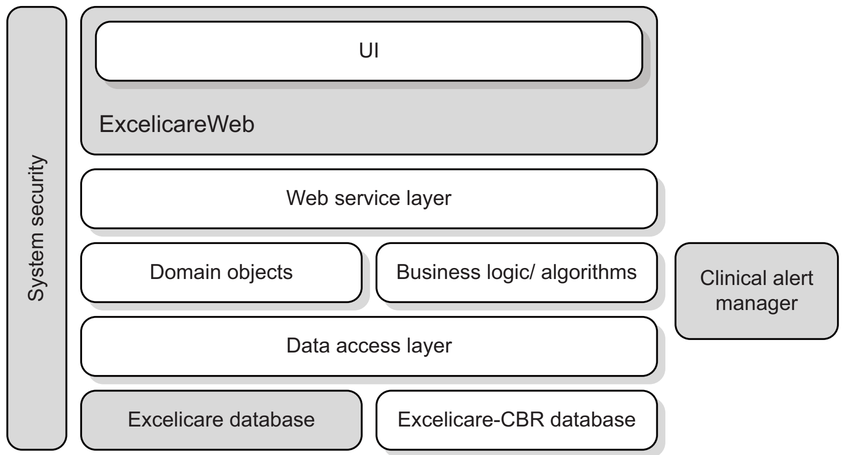

Figure 3. Multi-tiered system architecture. Grey blocks represent existing EXCELICARE components.

3. Any business logic is executed before calls are made into the data layer to retrieve cases, e.g. a user may not be authorised to view details of certain cases.
4. The data access layer retrieves cases from the case base and returns a list of cases to the business logic layer. The web service responds to the UI request and the UI presents an updated list of cases to the user.

The clinical alert manager (see right of Fig. 3) is an existing EXCELICARE component which can be used to define rules that trigger actions and notifications related to individual data items. For example, it can be used to notify clinicians when a patient's blood pressure level exceeds a value according to a predefined rule. The clinical alert manager can be referenced from the EXCELICARECBR module in order to send notifications when certain solutions are found by the system.

Application security is crucial to protect patient privacy and clinical data, and thus it is included in all layers of the architecture presented above. In practice, this will have different implications depending on the security requirements of each of the clinical environments in which the system operates. For example, a clinician may not have access to specific patient details related to a previous case that has been retrieved from the case base. However, an anonymised case will still be useful in terms of evidential support.

7. Case authoring and creation of the initial case base

A case authoring tool is part of the key components that have been developed to enable integration of EXCELICARECBR with the existing EXCELICARE system. This component allows the domain expert to define the indexing vocabulary and design the case representation by selecting individual data fields from EXCELICARE that map onto case features. In this section, the case authoring process will be illustrated by an example taken from the EXCELICARE lung cancer system.

Fig. 4 shows part of a radiotherapy form in the EXCELICARE Lung Cancer system. The form is used by clinicians to capture data of a radiotherapy course required for a specific patient. Multiple instances for each patient can exist, as patients can undergo multiple radiotherapy courses as part of their treatment.

To demonstrate EXCELICARECBR and the case authoring process, a simple hypothetical use case can be defined as follows: a clinician needs to decide the appropriate radiotherapy course for a patient. Details of patients which have been treated successfully and are similar to the current patient can be retrieved to support the decision. The course details are captured in the bottom of Fig. 4, and are mapped to the case representation. In order to simplify the example, only the details of the radiotherapy administered to the primary tumour site (Site 1 in Fig. 4) will be included.

A domain expert interacts with the User Interface in Fig. 5 to define a new domain. The grid in the top of the figure contains previously defined domains, which can be selected to perform editing operations. The details of the lung cancer domain (Domain name and Description) appear hidden in the screenshot but are listed below the *Domain* header and are revealed when the user clicks on the header. The relevant details for the features are visible and are shown in the *Features* panel. The dialog window on the right is the EXCELICARE *Field Selector* and presents a view of all existing data fields grouped by *special form* (see Section 4). For example, an expanded view of the *Radiotherapy* form in the bottom right of the figure shows the relevant fields with fields such as *Treatment intent* and *Site 1 target site*. The domain expert uses this EXCELICARE *Field Selector* to map existing EXCELICARE fields to EXCELICARECBR case features. This mapping is particularly important when different labels are used to describe case features from EXCELICARE field names. Essentially such a manual mapping mechanism enables alignment of clinical terminology with different health related ontologies.

Currently, the expert choose from three different feature types: Numeric, Nominal, and Goal. The component checks the underlying EXCELICARE field and suggests a feature type to the user. Fig. 5 illustrates how a Nominal feature type is selected and set for the *Site 1 Target Site* feature from the drop down list (or combo box). Similarly feature *Site 1 Duration* (from treatment course details in Fig. 5) maps to a field that contains numeric values; this results in a pre-selected Numeric feature type. The Goal feature type represents the predictor or solution for each case and its underlying type is Nominal. Functionality to allow the setting of one or more goal features is particularly important for generic reasoning systems. This is because health care domains may have one or more solution features and these can change depending on requirements of the expert.

Once case representation is completed, the domain definition is submitted to the Service layer (see Fig. 3) and a database structure is created to hold cases for the new case base. The domain expert can then import cases from the EXCELICARE database into the EXCELICARECBR database to form the case base for retrieval.

8. Case retrieval and reuse

Retrieval of similar cases and therefore similar patients, is performed using weighted k-nearest neighbour [14]. Currently the EXCELICARECBR module performs a simple exhaustive search, in

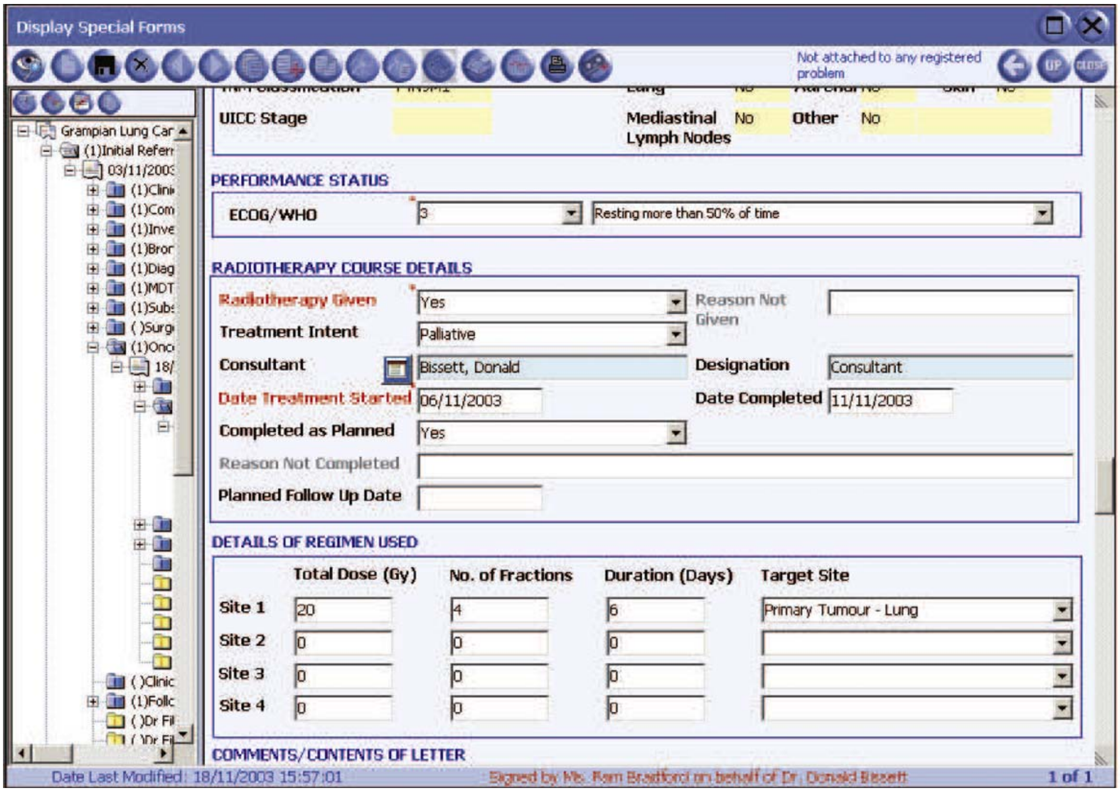

Figure 4. Detailed radiotherapy form in the EXCELICARE lung cancer system.

which each case is compared to the query, and the k most similar cases are returned to the user. Fig. 6 displays part of the interface for case retrieval and reuse that is used to test the current module. The case base can be selected from a drop down box. Furthermore, individual anonymous patients can be selected and are identified by a number. In the figure, patient with ID 109 has been selected and the *Patient summary* panel in the top of the figure presents the details of the proposed treatment for this patient, which will be used to query the system. The query can represent a proposed treatment or used to perform exploratory, what-if, treatment analysis. This can be particularly useful as a search tool for new clinicians.

The 5 similar cases that have been retrieved are presented in the same figure in the bottom *Search results* panel. The similarity between query and retrieved cases is visually presented using a coloured bar graph. With increased similarity resulting in a filled bar with more green colouring. No Goal feature has been used in this example. If however the domain expert defines one or more Goal features, then a majority weighted voting (or averaging) mecha-

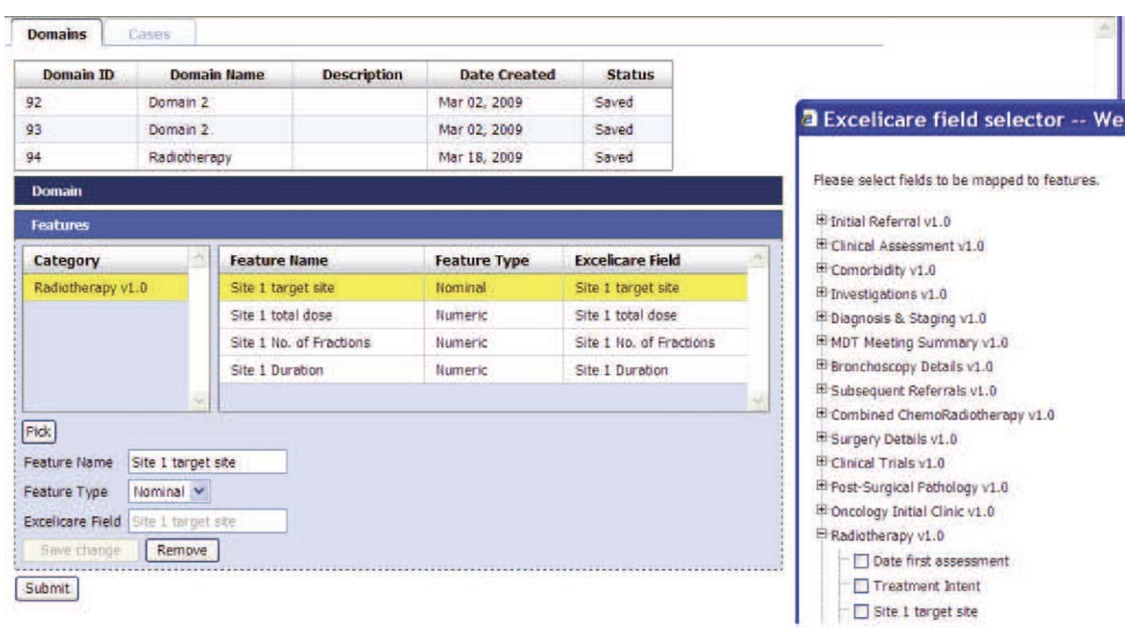

Figure 5. Screenshot of user interface used for testing case authoring in the EXCELICARECBR module.

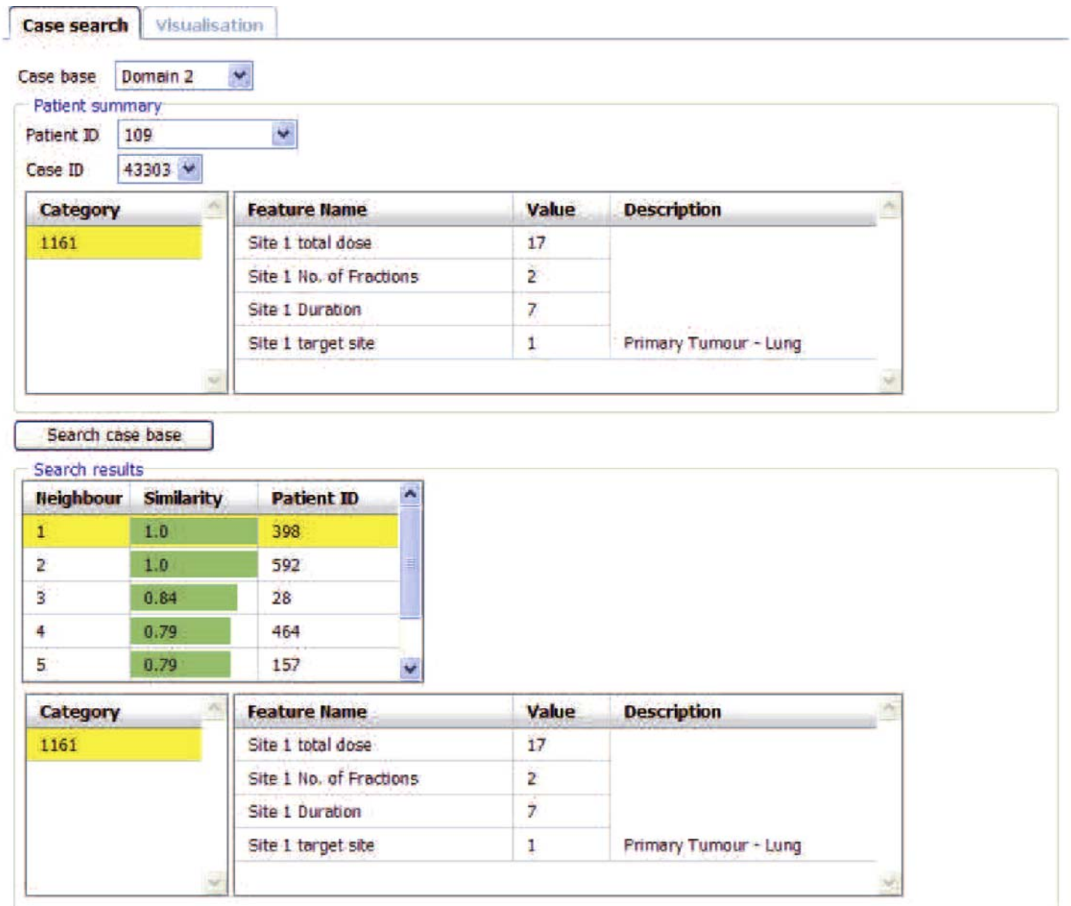

Figure 6. Preliminary retrieval interface displaying retrieval results.

nism is used to predict the value for Goal features. ExcelicareCBR also includes a genetic algorithm to perform feature weighting. This method has been used extensively to increase the accuracy of classifiers, including *k*-nearest neighbour (see, for example [15,16]) and explain the relevance of individual features to the user. If the domain expert has identified a nominal goal feature, then the genetic algorithm will run offline to discover feature weights using cross validation. When multiple nominal goals are specified sets of feature weights can be learned by applying the genetic algorithm to each individual goal feature. The User Interface is currently being extended to manage multiple goal feature specifications and individual feature weights.

9. Conclusions and future work

Addressing the need for generic solutions is key for successful adoption of advanced EPR technology within the UK Clinical practices. We have presented our experiences with ongoing work in integrating ExcelicareCBR with the commercial Excelicare system that is being deployed in several hospitals in the UK. Crucial to the integration of CBR with Excelicare is the ability to author cases from existing EPR databases. Mapping of EPR field data to ExcelicareCBR's case base features is demonstrated using an example taken from radiotherapy treatment data of lung cancer patients. The generic design allows this mapping to work with any Clinical domain on which Excelicare is required to operate on. ExcelicareCBR's approach to retrieval and treatment reuse is presented with emphasise on intuitive user interfaces.

Although the current Beta version of the ExcelicareCBR allows basic integration with Excelicare, several practical system challenges need to be addressed before ExcelicareCBR can be packaged commercially as a value-added component for the Excelicare system. For instance an indexing mechanisms needs to be in place to improve scalability and address retrieval performance needs. Currently ExcelicareCBR only supports the similarity aggregation as established by Eq. 3, and we are working to include other, configurable similarity assessments. In the future we will also develop a system management interface to allow configuration of parameters, experimentation and case maintenance. Furthermore, case representation needs to support complex, time-dependent data structures in Excelicare. A generic approach to this exploiting sequence mining algorithms is currently being explored. Initial responses from clinicians involved in the project were positive. They recognised the potential of CBR to facilitate evidence based practice. One concern that needs to be addressed more fully is patient data protection within CBR's retrieval mechanism. Finally user trials to establish delivery of current functionality and further extensions are planned prior to commercial release in 2010.

Acknowledgment

This research work is funded by the UK Knowledge Transfer Partnership Fund.

References

[1] Aylin P, Bottle A, Jarman B. Standardised mortality ratios, monitoring mortality. British Medical Journal 2009;338:b1745.
[2] Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications 1994;7(1):39–59.

[3] Bichindaritz I, Marling C. Case-based reasoning in the health sciences: What's next? Artificial Intelligence in Medicine 2006;36:127–35.
[4] Corchado JM, Bajo J, Abraham A. Gerami: improving health care delivery in geriatric residences. IEEE Intelligent Systems 2008;23(2):19–25.
[5] Schmidt R, Vorobieva O, Gierl L. Case-based adaptation in medicine focusing on hypothyroidism. In: Reimer U, Abecker A, Staab S, Stumme G, editors. Professionelles wissensmanagement. Kollen Verlag; 2003. p. 267–74.
[6] Montani S, Portinale L, Bellazzi R, Leonardi G. Rhene: a case retrieval system for hemodialysis cases with dynamically monitored parameters. In: Funk P, González-Calero PA, editors. Advances in case-based reasoning, 7th European conference on case-based reasoning. Heidelberg: Springer; 2004. p. 659–72.
[7] Marling C, Shubrook J, Schwartz F. Case-based decision support for patients with type 1 diabetes on insulin pump therapy. In: Althoff KD, Bergmann R, Minor M, Hanft A, editors. Advances in case-based reasoning, 9th European conference on case-based reasoning. Heidelberg: Springer; 2008. p. 325–39.
[8] Begum S, Ahmed MU, Funk P, Xiong N, von Schéele B. A case-based decision support system for individual stress diagnosis using fuzzy similarity matching. Computational Intelligence 2009;25(3):180–95.
[9] Doyle D, Cunningham P, Walsh P. An evaluation of the usefulness of explanation in a CBR system for decision-support in bronchiolitis treatment. Computational Intelligence 2006;22(3–4):269–81.
[10] Diaz-Agudo B, Gonzalez-Calero PA, Recio-Garcia JA, Sanchez A. Building CBR systems with JColibri, special issue on experimental software and toolkits. Science of Computer Programming 2007;69(1–3):68–75.
[11] Stahl A, Roth-Berghofer T. Rapid prototyping of CBR applications with the open source tool myCBR. In: Althoff KD, Bergmann R, Minor M, Hanft A, editors. Advances in case-based reasoning, 9th European conference on case-based reasoning. Heidelberg: Springer; 2008. p. 615–29.
[12] Perner P, Perner H, Jänichen S. Recognition of airborne fungi spores in digital microscopic images. Artificial Intelligence in Medicine 2006;36: 137–57.
[13] Montani S. Case-based reasoning for managing noncompliance with clinical guidelines. Computational Intelligence 2009;25(3):196–213.
[14] Dietrich Wettschereck TM, Aha, D.W. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Intelligence Review 1997;11(1–5):273–314.
[15] Jarmulak J, Craw S, Rowe R. Genetic algorithms to optimise CBR retrieval. In: Blanzieri E, Portinale L, editors. Advances in case-based reasoning, 5th European conference on case-based reasoning. Heidelberg: Springer; 2000. p. 136–47.
[16] Craw S, Jarmulak J, Rowe R. Maintaining retrieval knowledge in a case-based reasoning system. Computational Intelligence 2001;17(2):346–63.