Research Themes in the Case-Based Reasoning in Health Sciences Core Literature

Isabelle Bichindaritz

State University of New York
Oswego, New York 13126, USA
ibichind@gmail.com

Abstract. Research in case-based reasoning (CBR) in the health sciences started more than 20 years ago and has been steadily expanding during these years. This paper describes the state of the research through an analysis of its mainstream, or core, literature. The methodology followed involves first the definition of a classification and indexing scheme for this research area using a tiered approach to paper categorization based on application domain, purpose of the research, memory organization, reasoning characteristics, and system design. A research theme can be tied to any of the previous classification elements. The paper further analyzes the evolution of the literature, its characteristics in terms of highest impact, or most cited, papers, and draws conclusions from this analysis. Finally, a comparison with the themes automatically learned through clustering co-citations matrices with the Ensemble Non-negative Matrix Factorization (NMF) algorithm in the CBR conference literature is proposed. This comparison helps better understand the main characteristics of the field and propose future directions.

Keywords: case-based reasoning, classification, biomedical informatics, biometrics, text mining.

1 Introduction

The field of Case-Based Reasoning (CBR) in the Health Sciences (CBR-HS) [1] has seen a tremendous growth in the last decade. An international group of researchers performs its research mainly in this domain, and constitutes the core CBR-HS research community. Seven specialized conference workshops have been held consecutively between 2003 and 2009 focused solely on this topic. In addition, six journal special issues on CBR-HS were published in the journals Artificial Intelligence in Medicine [2][3][4], Computational Intelligence [5][6], and Applied Intelligence [7]. The domain has been the subject of several survey papers as well, mostly qualitative in nature, hence the need to track the evolution in the research in a more systematic and automatic manner. We developed a classification and indexing scheme for CBR research in the Health Sciences to make possible the meta-analysis of this interdisciplinary research area [1] in a semi-automatic manner. This paper details knowledge of CBR-HS gained by building and using this classification scheme and the research

P. Perner (Ed.): ICDM 2012, LNAI 7377, pp. 9–23, 2012.

trends identified in terms of application domains, application purposes, system memory, reasoning, and design, as well as evolution of number of papers, citations, and research themes. In addition, a comparison is proposed with an automatic clustering method called Ensemble Non-negative Matrix Factorization (NMF) [8] to determine how the major themes in the CBR conference literature differ from those in the core CBR-HS literature.

2 Methods

The specific application of CBR to the health sciences has been discussed in several surveys [9, 10, 11, 12, 13, 14, 15]. However recent trend analyzes in CBR as a whole failed to identify CBR-HS as a sub-research area through automatic methods [8]. This may in particular be due to the variety of application domains comprising the health sciences, which prompts for the need to index systems capable in particular of grouping documents related to, for example, oncology, diabetology, phrenology and so forth. Therefore we developed a classification and indexing system capable of drilling down and rolling up in its different components and presented in detail elsewhere [1]. This domain-specific indexing is enabled by the use of one of the most used classification schemes in the health sciences: the Medical Subject Headings (MeSH) [16]. Like most other classifications, it uses a tree like structure where broader categories are narrowed down with each branch and branches are represented by dots.

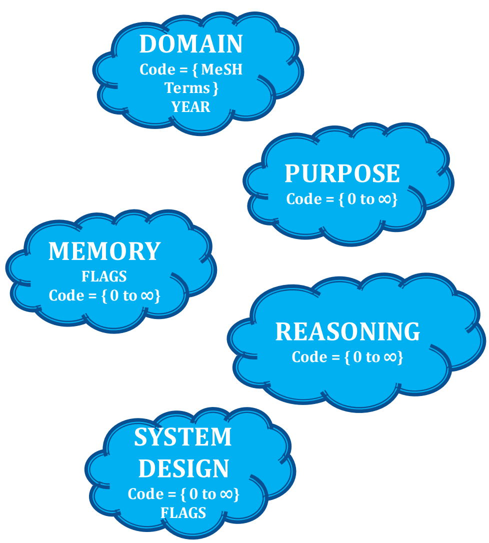

Fig. 1. CBR Health Sciences tiered classification scheme

The papers selected for CBR-HS in this paper cover all the 16 EWCBR (European Workshop on Case Based Reasoning), ECCBR (European Conference on Case Based Reasoning), and ICCBR (International Conference on Case Based Reasoning) conferences from 1993 until 2011, the 7 Workshops on CBR in Health Sciences, the 5 special issues on CBR in the Health Sciences, the 2 DARPA workshops of 1989 and 1991, which preceded the CBR official conferences, and the survey papers on CBR in the Health Sciences. We also added papers preceding the papers published in the official CBR-HS venues, before they existed. These papers were identified by the group of CBR-HS researchers who prepared the 2007 survey. 156 papers were indexed with the CBR-HS classification scheme, presented in the next section. Therefore the terminology learned for the classification has been refined on these 156 papers.

3 Classification System

Figure 1 presents the tiered architecture of the CBR-HS classification scheme. There are five distinct categories (domain, purpose, memory and case management, reasoning, and system design) defined in this section. A research theme can be selected by researchers among any of these categories, to characterize the main research hypothesis and findings of the paper. Codes have been created to represent each classification category. We refer the reader to another article [1] for the coding details.

1. Domain: The range of domains, such as for example oncology or diabetology, in the health sciences fields is vast and, as a result, it was chosen as the first level of classification. However, rather than creating a new set of descriptors, it is proposed to use the MeSH descriptors [16], of which there are over 24,000 that cover just about every aspect of the health sciences. Along with the domain, another primary means of discriminating the relevance of an article is its publication date.
2. Purpose: The purposes, or tasks, of CBR systems have been thoroughly discussed in many articles summarizing the CBR-HS domain. One of the first papers to survey the field in 1998, by Gierl et al., used the purpose as the primary means to subdivide the different systems [9]. In their paper, Gierl et al. specified four main purposes: diagnosis, classification, planning, and tutoring. Later, both Holt et al. 2006 [13] and Nilsson and Sollenborn 2004 [12] used the same four descriptors. In the early years the majority of systems were diagnostic in nature, but in recent years more therapeutic and treatment systems have been developed [14]. We have replaced planning by treatment since most of the time planning refers to treatment planning. However, planning tasks may involve not only treatment but also other aspects such as diagnosis assessment, which often consists in a series of exams and labs orchestrated in a plan. Planning is a classical major task performed by artificial intelligence systems. Therefore planning is listed in our system as a design option and thus can be added to the treatment choice in the purpose dimension. CBR systems generally support either medical clinical work, or research. Therefore we have added these as top level purpose categories (see Table 2). In the clinic, decision support systems support mostly diagnosis, treatment, prognosis, follow-up, and/or classification, such as in image interpretation. More recent articles require to differentiate between the purpose of the system developed, which is generally a clinical purpose, from the

purpose of the research paper, which can be, among others, a survey paper or a classification paper like this one. Some papers focus on formalization, a method, or a concept. Among these, the evaluation of a system can be performed more or less thoroughly. This is an important dimension to note about a research paper: whether the system was tested only at the system level, which is the most frequent, at the pilot testing level, at the clinical trial level, or finally whether the system is in routine clinical use.

3. Memory and case management: This is a very broad category and could easily be subdivided. It encompasses both how the cases are represented and also how they are organized in memory for retrieval purposes and more (see Table 3). As a result, it is made up of more than one code. The first part of the code represents the format of the cases. The primary types being images, signals, mass spectrometry, microarray, time series data and regular attribute/values pairs, which is used by the majority of the systems. Similar to the different formats of data are the flags that represent what kinds of memory structures the CBR system uses to represent the data, such as ground cases (G), prototypical cases (P), clusters (L), or concepts (O). Lastly, when it comes to memory management there are potentially an infinite number of possibilities, some of which may never have been used before. The main types, however, represent how the memory is organized, whether it is flat or hierarchical, what kind of hierarchical structure, such as decision tree, concept lattice, conceptual clustering tree, or others.
4. Reasoning: This category regroups the inferential aspects of the CBR. Classically, retrieve, reuse, revise, and retain have been described. Nevertheless, researchers have often added many more aspects to the inferences, such that it is best to keep this category open to important variations. Each of these parts of the reasoning cycle can be hierarchically refined so that a tree is formed here also.

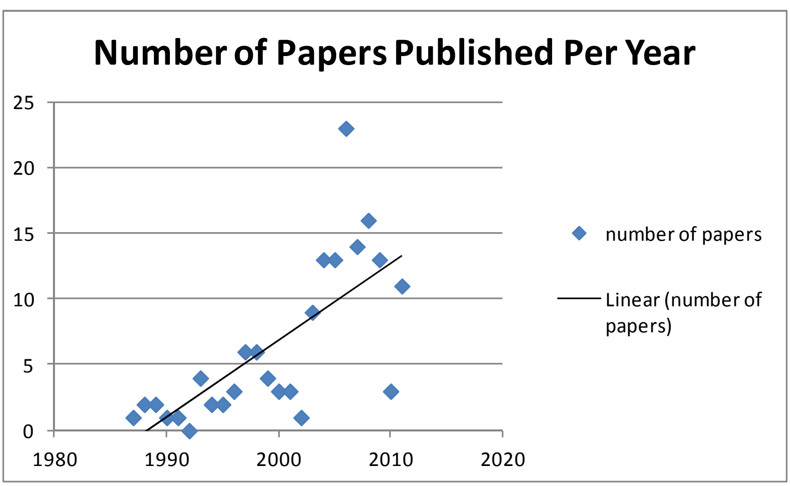

Fig. 2. Evolution of the number of papers published

5. System design: The construction of the CBR system specifies what technologies it uses. This area of classification may not seem intuitive at first, but upon the examination of CBR systems it can be seen that many use a combination of technologies, not just case-based reasoning. The most common technology used in conjunction with CBR is rule-based reasoning; however some systems combine CBR with information retrieval, data mining, or other artificial intelligence methods. See table 4 for an example of different possible construction classifications. If the construction of the system does use additional technologies, a flag should be appended to the end of the code to denote whether the case-based reasoning is executed separately. Also, an additional flag is used to designate CBR's role in the system, whether primary, secondary, or equivalent.

Table 1. A ranked list of the top 10 highest impact papers in the core CBR-HS collection based on total citation count

#	*Paper*	*Year*	*Citations*
1	Concept learning and heuristic classification in weak-theory domains Porter, Bareiss & Holte [17]	1990	290
2	Reasoning about evidence in causal explanations Koton [18]	1988	239
3	Protos: an exemplar-based learning apprentice Bareiss, Perter & Wier [19]	1988	156
4	Case-based reasoning in CARE-PARTNER: gathering evidence for evidence-based medical practice Bichindaritz, Kansu & Sullivan [20]	1998	86
5	Cased-based reasoning for medical knowledge-based systems Schmidt, Montani, Bellazzi, Portinale & Gierl [10]	2001	83
6	Using experience in clinical problem solving: introduction and Framework Kolodner & Kolodner [21]	1987	81
7	A two layer case-based reasoning architecture for medical image understanding Grimnes & Aamodt [22]	1996	76
8	Case-based reasoning in the health sciences: what's next? Bichindaritz & Marling [11]	2006	72
9	An architecture for a CBR image segmentation system Perner [23]	1999	69
10	Advancements and trends in medical case based reasoning: an overview of systems and system development Nilsson & Sollenborn [12]	2004	65

4 Global Picture

The global picture of the core CBR-HS literature shows a total of 156 papers being published between 1987 and 2011 from 179 different authors from all over the world. The average number of papers per author is 2.27, and the range is 1 to 27. We searched these papers in Google Scholar to count their number of citations and calculated a total of 3237 citations.

The evolution of the number of papers is provided on Fig. 1. It shows a regular increase in the research productivity in this domain, which attests of the vitality of the field. This graph demonstrates in particular that the number of papers by year has seen a rapid increase after 2003 – corresponding to the first workshop dedicated to CBR-HS (see Figure 1).

In terms of impact, Table 1 lists the 10 highest impact papers based on their number of citations in Google Scholar, after removing the number of self-citations (only the order of the papers changes if taking into account all citations). It is interesting to note that the pioneering papers in the domain are ranked in positions #1, 2, 3, and 6. These papers preceded the creation of the CBR-HS research field, however have impacted the field tremendously. These papers do not refer to the term of CBR yet, however they have served to define the feasibility and direction of this research. In that sense, they can be regarded as its seed papers. The other papers took about 10 years to emerge from the tracks defined by the seed papers (paper #4 in particular). Paper #5 is the first survey paper in CBR-HS, and papers #8 and 10 are later surveys. Papers #7 and 9 represent the seeds in a group of CBR-HS papers devoted to the research theme of medical image interpretation.

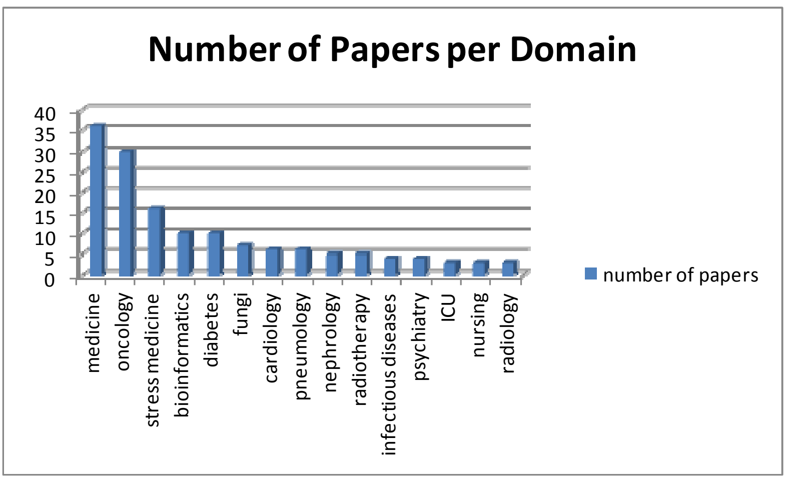

Fig. 3. Domains the most studied by CBR-HS papers

5 Analysis of Classes

Based on the classification we have defined, interesting research themes appear in terms of domain, purpose, memory, reasoning, and design. Since most papers focus both on a domain and another dimension such as design for example, the papers contribute to several classes. In addition, in each class, they also very often contribute to several categories, such as treatment and diagnosis for example.

Domains

The 156 papers cover 38 domains all together. Although the domains of application all belong to the Health Sciences, some domains are more represented than the other ones. The most represented domain is medicine with 36 papers as a whole, which correspond to either survey papers, editorials, or general frameworks and concepts applicable to any health sciences domain. Close second comes oncology (30 papers), then further come stress medicine (16 papers), diabetes (10 papers), fungi detection (7 papers) – which could have been added to the infectious diseases papers, cardiology and pulmonology (6 papers each), nephrology and radiotherapy (5 papers each), infectious diseases and psychiatry (4 papers each), and intensive care (ICU), nursing, and radiology (3 papers each). All the other domains count less than 3 papers. It is interesting to note in particular that cancer, being a very prominent disease, is studied by several CBR-HS teams in the world.

Table 2. Main Purpose Themes and the corresponding number of papers

Purpose	#	Purpose	#
Medical Purpose	156	CBR-HS Research Purpose	32
Decision Support	136	Survey	17
Treatment/therapy	46	Evaluation / testing	8
Diagnosis	36	Role of CBR	4
Classification	27	Concept	2
Interpretation	13	Formalization	1
Prognosis / prediction	7		
Follow-up	5		
Assessment	2		
Medical research support	8		
Quality control / monitoring	3		
Information retrieval / navigation	3		
Tutoring	2		
Parameter configuration	2		
Drug design	1		
Explanation	1		

Purpose

Among the 24 purposes listed in these papers, we can distinguish between medical purpose, tied to the application domain, and research purpose, tied to the CBR-HS domain.

In terms of medical purpose, 46 papers propose treatments / therapies (among which two propose prescriptions), 36 propose diagnosis recommendations, 27 classifications, 17 papers refer globally to decision support (to which we can add the sub-types of decision-support tasks – see Table 2), and there are additional decision-support tasks such as interpretation (mostly for image interpretation). Other papers propose to help medical research (8), quality control and monitoring (3), and information retrieval or case-base navigation (3), among several other medical purposes.

In terms of research or methodological purpose, 17 papers are survey, editorial, or systematization papers, 8 papers focus on evaluation methods, 4 papers investigate the role CBR can play in medical domains, and a few papers focus on formalization, concepts, and methods.

It is notable that 28 papers describe several purposes, for example they tackle both diagnosis and treatment decision-support, although each of these tasks alone, given its complexity, could be the topic of an entire paper. Another important characteristic is that several systems focus on differential diagnosis, which involves the value of diversity in the diagnostic recommendation.

Table 3. Sample Memory and Case Management Themes

Generalized Memory Structures	#	Data Types	#
Prototypes	27	Time Series / signals / sensor data	24
Clusters	4	Images	17
Categories	3	Microarray data / genetic sequences	10
Generalized cases	3	Text	7
Inverted indexes	2	Scenarios	2
Schemas	2	Graphs	1
Scenarios	2	Networks	1
Concepts	1	Plans	1
Trends	1	Visio-spatial data	1

Memory and Case Management

Memory structures and organization refers to at least 24 different concepts, which encompass generalized memory structures and a variety of ground cases which can be identified by their case data types (see Table 3). In addition to traditional ground cases or exemplars, which appear in almost all papers, the most represented memory structures are prototypes (27 papers), closely followed by time series ground cases (24 papers, most of them being from signals). Further come image ground cases (17 papers), microarray data ground cases (10 papers), text ground cases (7 papers), clusters (4 papers), categories (3 papers), generalized cases (3), inverted indexes (2 papers), scenarios (2 papers), and schemas (2 papers). The other listed memory structures contain, among others, networks, graphs, multimedia data, plans, structured cases, and visio-spatial cases.

In terms of memory organization, the various types are exemplified in these systems, ranging from flat memories, to decision trees and concept hierarchies. Hierarchical organizations are very prominent in the systems using the generalized memory structures (there are 45 of these papers).

Table 4. System Design Classification

Artificial intelligence methodology / component	#	Biomedical methodology / component	#
Machine learning & data mining	34	Clinical guidelines	6
Prototype learning / generalized case learning	11	Electronic medical records	3
Feature mining / key sequence learning	6		
kNN	5		
Statistical learning	4		
Conceptual clustering	3		
Text mining / case mining	3		
Genetic algorithms	2		
Feature selection / dimensionality reduction	7		
Rule based reasoning	16		
Temporal abstraction	14		
Fuzzy logic	9		
Information retrieval	9		
Knowledge discovery	7		
Knowledge-based systems / semantic Web	6		
Planning	6		
Knowledge acquisition	5		
Temporal reasoning	3		

Reasoning

In CBR-HS the vast majority of systems refer to retrieval and similarity assessment (92 papers) as well as another form of reasoning. Next, maintenance is also well represented (16 papers), as well as adaptation and reuse (15 papers). Further are represented: retain step (8 papers), indexing (5 papers), and revision (5 papers). The retain step could be combined with case base maintenance, even though authors using one term sometimes do not use the other term. Most systems perform several reasoning steps, even though the papers studied did not detail these steps, focusing on retrieval aspects instead. Many papers deal with several reasoning steps in the same paper.

System Design

Main characteristics of developed systems describe the types of components involved in building CBR systems in the health sciences. Although there are many “pure” CBR systems, most systems describe a combination of components to manage to solve a problem in the application domain, thus making them hybrid systems. The role of CBR in the hybrid system is most of the time the primary methodology, although many systems report methodologies of equivalent role. Few describe CBR as a secondary methodology. There are mostly two types of hybrid methodologies: those coming from artificial intelligence and more broadly computer science (such as ambient systems), and those coming from the field of biomedical informatics.

We have listed 52 different methodologies added to CBR (see Table 4). The main methodologies are: machine learning and data mining (34 papers), with different methods such as prototype and generalized case learning, feature mining, kNN, conceptual clustering, statistical learning, text mining, case mining, and genetic algorithms. In second place and beyond come rule-based reasoning (16 papers), temporal abstraction (14 papers), fuzzy logic (9 papers), information retrieval (9 papers), knowledge discovery (7), knowledge-based systems (6), and planning (6) combinations.

However two categories are specific to medical domains: clinical guidelines integration, and electronic medical records integration.

Table 5. Major research themes in the core CBR-HS literature based on the number of papers addressing them

#	CBR-HS core literature	Number of papers
1	Reasoning: retrieval & similarity assessment	92
2	Purpose: treatment or therapy decision-support	46
3	Purpose: diagnosis decision-support	36
4	Design: machine learning / data mining combination	34
5	Domain: oncology	30
6	Memory: prototypes	27
7	Purpose: classification	27
8	Memory: time series / signals / sensor data	24
9	Memory: images	17
10	Design: temporal abstraction & reasoning	17
11	Design: rule based reasoning combination	16
12	Reasoning: case base maintenance	16
13	Reasoning: adaptation	15
14	Purpose: interpretation decision-support	13
15	Memory: microarray data / genetic sequences	10
16	Design: prototype learning / generalized case learning	11
17	Design: fuzzy logic combination	9
18	Design: information retrieval combination	9
19	Reasoning: retain	8
20	Purpose: evaluation & testing	7

Top Twenty Research Themes

Combining results from the previous sub-sections, we get a clear picture of the major themes in the core CBR-HS literature (see Table 5). We can also note that some very important themes are not in the top twenty research themes, however they are promising and very important for the future development of the field (clinical guidelines integration and electronic medical records integration are some examples) [11].

A research theme can pertain to any of the classification tiers previously presented. Table 5 ranks the top twenty research themes in term of number of papers dealing with it in some way. Since the papers often cover several of these research themes, the sum of these figures does not have to be equal to the number of papers.

With this simplification, the major research themes are the ones ranked #1 through 8 since there is a clear break in number of papers between 24 and 17, in terms of number of papers. These major research themes are therefore:

- In terms of the reasoning dimension, retrieval and similarity assessment (#1);
- In terms of the purpose dimension, treatment / therapy (#2), diagnosis (#3), and classification (#7) decision-support;
- In terms of the design dimension, machine learning / data mining combination (#4);
- In terms of the domain dimension, oncology (#5);
- In terms of the memory dimension, prototypes (#6) and time series / signals / sensor data (#8).

6 Comparison with CBR Conference Research Themes

Greene et al. have identified through an automatic method, called NMF, a number of major themes in the CBR conference literature [8]. It is interesting to compare the major themes identified above with those they have identified (see Table 6).

Table 6 shows the themes in correspondence, namely Case base maintenance, Case retrieval & similarity assessment, Adaptation, Image analysis, Textual CBR, Creativity & knowledge-intensive CBR, CBR on temporal problems, and Structural cases. A 'Yes' in the 3rd column indicates that this theme from the CBR conference literature [8] is also present along the highest ranked themes in the CBR_HS literature as identified in the present paper. The '#' symbol refers to the ranking in either the CBR conference literature [8] (2nd column) or the CBR-HS core literature (4th column).

As for CBR conference literature main research themes not represented on Table 6, they are of interest for suggesting future research themes in CBR-HS:

- Recommender systems & diversity: diversity is an important aspect for differential diagnosis. Even though a few CBR-HS systems show some interest in this direction, it is a promising topic to focus on for the future. The spread of health-related online communities and social networks may very well join the research efforts in recommender systems. In addition, the team-based work in the clinic could also take example on this core CBR research for CBR-HS systems.
- Learning similarity measures: even though CBR-HS systems have not yet applied this to their systems yet, it is probably a potential improvement to test.
- Conversational CBR: very few CBR-HS systems actually interact so closely with healthcare professionals, however this could become very

important for patient-centered CBR-HS systems (another potentially very important research direction).

- Feature weighting and similarity: CBR-HS systems in bioinformatics have started researching in this direction, which is connected with the feature mining themes and the feature selection / dimensionality reduction theme. However it is not clear from the Greene et al. paper whether they encompass these in this category [8] – it would make sense to connect them.
- Games & chess: the field of serious games will provide in the future opportunities for common projects with CBR-HS, for example for pain management and phobia treatment.
- Scheduling & agents: there are potential common research projects in health care management and in public health.

Table 6. Comparison of research themes between the CBR conference literature and the CBR-HS core literature (the '#' columns represent the ranking in the articles of reference)

CBR conference literature	#	CBR-HS core literature	#
Recommender systems & diversity	1	Not a major theme currently	N/A
Case base maintenance	**2**	**Yes**	**12**
Case retrieval & similarity assessment	**3**	**Yes**	**1**
Learning similarity measures	4	Not a major theme currently	N/A
Adaptation	**5**	**Yes**	**13**
Image analysis	**6**	**Yes**	**9**
Textual CBR	**7**	**Yes**	**18**
Conversational CBR	8	Not a major theme currently	N/A
Feature weighting & similarity	9	Not a major theme currently	N/A
Creativity & knowledge-intensive CBR	**10**	**Yes – rule-based combination**	**11**
CBR on temporal problems	**11**	**Yes**	**8**
Games & chess	12	Not a major theme currently	N/A
Scheduling & agents	13	Not a major theme currently	N/A
Structural cases	**14**	**Yes – prototypes & prototype learning / generalized case learning**	**6**

In terms of research themes little represented in the CBR conference literature, we can list those with a medical purpose, in particular treatment / therapy decision-support (#2), diagnosis decision-support (#3), and classification decision-support (#7). Of course the oncology domain (#5) is not a major research them in the CBR conference literature. We can also list complex structured cases and complex case data types in memory, as they exist in biomedical domains, such as prototypes (#6) and time series / signals / sensor data (#8). Hybrid systems were not identified either as a major theme in the CBR conference literature, hence the non-existence of machine learning / data mining combination (#4) for example.

We would also like to comment on the highly cited papers in CBR-HS, namely papers #1 (PROTOS [17], 290 citations), #2 (CASEY [18], 239 citations), #3 (PROTOS [19], 156 citations), and #6 (SHRINK [21], 81 citations). These papers clearly demonstrate that CBR-HS papers can have an impact as high and even higher as those in the CBR conference literature, where the top ranked papers receive 137, 117, 92, and 82 citations on Google Scholar. None of these CBR-HS papers clearly label themselves as CBR papers, and we may wonder whether this is not in part an explanation for their success. They present their concepts and ideas more in cognitive terms understandable to any researcher in biomedical or artificial intelligence domains, which probably contributes to making them attractive to a broader audience.

It is also interesting to note that the highest ranked CBR-HS paper from the CBR conference papers [20] counts 86 citations after removing self-citations (98 otherwise), which positions it at the top 4th position both in the CBR-HS classification (see Table 1) and potentially in the Greene et al. classification [8]. Therefore CBR-HS papers published at CBR conferences can reach a citation count comparable to that of highly cited non applied papers. This is encouraging for authors publishing mostly in the CBR conferences.

7 Discussion and Future Plans

The CBR-HS classification system is being incrementally built. The different categories and each category's list of descriptors are by no means exhaustive. However it proved useful for indexing and tracking CBR-HS research literature. With its system of tiers, some of which may be omitted, this system is very flexible and can index either fielded applications, frameworks, or survey papers. This study has identified interesting research themes characteristic of applied domains such as health sciences domains. The classification system allows for an easy tracking of these trends over time.

In comparison with previous survey papers, which are more qualitative in nature, the results presented in this paper share many important facts. For example, in the most recent survey, Begum et al. classify the CBR-HS papers between those that are purpose-oriented and those that are construction-oriented [15]. We also classify them in terms of their purpose dimension and in terms of their design dimension. The major themes they list correspond to a large extent to the ones we have identified; however we quantify the weight of each group of papers. In addition we have conducted a more exhaustive study on a larger pool of papers (156) and along more dimensions, made possible by the indexing simplification provided by the classification system. We intend to continue tracking progress in CBR-HS through this indexing mechanism and to make the papers and their indexing available from a Web-site to better showcase accomplishments in CBR-HS.

Our next goal is to attempt an automatic classification with NMF algorithm as described by Green et al. [8]. Although we do not expect very interesting results from this additional study, since these authors report that they could not identify a cluster for the CBR-HS domain, it is possible that some sub-clusters could overlap with the ones we found with a semi-automatic indexing of the papers, as presented in this document. These automatically found clusters may also suggest some indexing terms and concepts we may have overlooked in the current study. In addition, the co-citation analysis will provide a different view of the most influential literature, however, as shown in Greene et al., the overlap with the number of citations is expected to be very important [8].

Another planned activity is to provide an automatic or semi-automatic indexing of the articles. Right now, the indexing is humanly made, however we are in the process of attempting to index the papers largely automatically – under the supervision of an expert, which is how current literature indexing is accomplished on a large scale. A completely automatic indexing system remains as a research goal for the long-term.

8 Conclusion

The CBR-HS classification system is being incrementally built, and it will continue to be refined as we add papers. The different categories proved useful for indexing and tracking CBR-HS core research literature. With its system of tiers, some of which may be omitted, this system is very flexible and can index either fielded applications, frameworks, or survey papers. This study has identified interesting and major research themes and trends characteristic of applied domains such as health sciences domains. I has also compared these themes with those in the CBR conference literature, and found both common elements and differences. This analysis of CBR-HS literature also permits to identify potential future research directions. Future directions include visualization and evolution tracking of CBR-HS literature, comparison with automatic classification, as well automatizing the indexing system as much as feasible.

References

1. Bichindaritz, I., Reed, J.: Methodology for Classifying and Indexing Case-Based Reasoning Systems in the Health Sciences. In: Florida Artificial Intelligence Research Society Conference, North America (March 2009), `http://www.aaai.org/ocs/index.php/FLAIRS/2009/paper/view/144`
2. Macura, R.T., Macura, K.J.: Case-based reasoning: opportunities and applications in health care. Artificial Intelligence in Medicine 9(1), 1–4 (1997)
3. Bichindaritz, I.: Case-based reasoning in the health sciences. Artificial Intelligence in Medicine 36(2), 121–125 (2006)
4. Bichindaritz, I., Montani, S.: Advances in case-based reasoning in the health sciences. Artificial Intelligence in Medicine 51(2), 75–79 (2011)
5. Bichindaritz, I., Marling, C.: Introduction to the special issue on case-based reasoning in the health sciences. Computational Intelligence 22(3–4), 143–147 (2006)
6. Bichindaritz, I., Montani, S.: Introduction to the special issue on case-based reasoning in the health sciences. Computational Intelligence 25(3), 161–164 (2009)
7. Bichindaritz, I., Montani, S., Portinale, L.: Special issue on case-based reasoning in the health sciences. Applied Intelligence, 1–3 (2008)

8. Greene, D., Freyne, J., Smyth, B., Cunningham, P.: An Analysis of Research Themes in the CBR Conference Literature. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 18–43. Springer, Heidelberg (2008)
9. Gierl, L., Bull, M., Schmidt, R.: CBR in Medicine. In: Lenz, M., Bartsch-Spörl, B., Burkhard, H.-D., Wess, S. (eds.) Case-Based Reasoning Technology. LNCS (LNAI), vol. 1400, pp. 273–297. Springer, Heidelberg (1998)
10. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., Gierl, L.: Case-based reasoning for medical knowledge-based systems. International Journal of Medical Informatics 64(2-3), 355–367 (2001)
11. Bichindaritz, I., Marling, C.: Case-based reasoning in the health sciences: what's next? Artificial Intelligence in Medicine 36(2), 127–135 (2006)
12. Nilsson, M., Sollenborn, M.: Advancements and trends in medical case-based reasoning: An overview of systems and system development. In: Proceedings of the 17th International FLAIRS Conference, Special Track on Case-Based Reasoning, pp. 178–183. AAAI (2004)
13. Holt, A., Bichindaritz, I., Schmidt, R., Perner, P.: Medical applications in case based reasoning. The Knowledge Engineering Review 20(3), 289–292 (2007)
14. Greene, D., Freyne, J., Smyth, B., Cunningham, P.: An Analysis of Research Themes in the CBR Conference Literature. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 18–43. Springer, Heidelberg (2008)
15. Begum, S., Ahmed, M.U., Funk, P., Xiong, N., Folke, M.: Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments. IEEE Transactions on Systems, Man, and Cybernetics 41(4), 421–434 (2011)
16. Lipscomb, C.E.: Medical subject headings (mesh). Bulletin of the Medical Library Association 88(3), 265–266 (2000)
17. Porter, B.W., Bareiss, E.R., Holte, R.C.: Concept learning and heuristic classification in weak-theory domains. Artificial Intelligence 45(1-2), 229–263 (1990)
18. Koton, P.: Reasoning about evidence in causal explanations. In: Proceedings of AAAI 1988. Seventh National Conference on Artificial Intelligence, pp. 21–26 (1988)
19. Bareiss, E.R., Porter, B.W., Wier, C.C.: Protos: an exemplar-based learning apprentice. In: Proceedings of the Fourth International Workshop on Machine Learning, pp. 12–23. Morgan Kaufmann, Los Altos (1987)
20. Bichindaritz, I., Kansu, E., Sullivan, K.M.: Case-Based Reasoning in CARE-PARTNER: Gathering Evidence for Evidence-Based Medical Practice. In: Smyth, B., Cunningham, P. (eds.) EWCBR 1998. LNCS (LNAI), vol. 1488, pp. 334–345. Springer, Heidelberg (1998)
21. Kolodner, J.L., Kolodner, R.M.: Using experience in clinical problem solving: introduction and framework. IEEE Transactions on Systems, Man, and Cybernetics 17(3), 420–431 (1987)
22. Grimnes, M., Aamodt, A.: A Two Layer Case-Based Reasoning Architecture for Medical Image Understanding. In: Smith, I., Faltings, B.V. (eds.) EWCBR 1996. LNCS, vol. 1168, pp. 164–178. Springer, Heidelberg (1996)
23. Perner, P.: An architecture for a CBR image segmentation system. Journal on Engineering Application in Artificial Intelligence 12, 749–759 (1999)