

# AN INTELLIGENT DIAGNOSIS SYSTEM HANDLING MULTIPLE DISORDERS

Wenqi Shi

*School of Computer Science, The University of Birmingham  
Edgbaston, Birmingham, B15 2TT, United Kingdom  
Tel: 0044-121-414-2884*

W.Shi@cs.bham.ac.uk

John A. Barnden

*School of Computer Science, The University of Birmingham  
Edgbaston, Birmingham, B15 2TT, United Kingdom*

J.A.Barnden@cs.bham.ac.uk

Martin Atzmueller

*Department of Computer Science, University of Wuerzburg  
97074 Wuerzburg, Germany*

atzmueller@informatik.uni-wuerzburg.de

Joachim Baumeister

*Department of Computer Science, University of Wuerzburg  
97074 Wuerzburg, Germany*

baumeister@informatik.uni-wuerzburg.de

**Abstract** Although Case-based Reasoning has been applied successfully in medical domains, case-based diagnosis handling multiple disorders is often not sufficient while multiple disorders is a daily problem in medical diagnosis and treatment. In this paper, we present an approach which integrates two case-based reasoners for diagnosing multiple faults. This multiple case-based reasoning approach has been evaluated on a medical case base taken from real world application and demonstrated to be very promising.

**Keywords:** Intelligent Diagnosis System, Multiple Case-based Reasoning (Multiple CBR), Multiple Disorders

## 1. Introduction

Medical Diagnosis problem has absorbed lots of AI researchers' attention since medical domain is not really well understood in some aspects by human being and Artificial Intelligence has potential to help diagnosis. Case-based Reasoning ( CBR ) employs existing experience to support problem solving without necessarily understanding the underlying principles of application domain. It takes a very different view from normal reasoning process which draws its conclusions by starting from scratch and chaining generalized rules. Case-based Reasoning has been demonstrated to be very suitable for weak theory domains, especially for medical domain.

Multiple disorders are a daily problem in medical diagnosis and treatment. However, case-based diagnosis handling multiple disorders is still a challenging task. Moreover, multiple disorder cases always occur with single disorder cases in real world applications which makes diagnosis more difficult.

In this paper, we present a multiple case-based reasoning approach (Multiple CBR)<sup>1</sup> which integrates naive case-based reasoning and compositional case-based reasoning to handle multiple disorder problem<sup>2</sup>. In the following section, we firstly describe Compositional Case-based Reasoning, then a multiple case-based reasoning system will be presented. This approach is evaluated on a medical case base in the subsequent section. We conclude the paper after a discussion of the presented work and related work, then give pointers to promising work in the future.

## 2. Multiple Disorder Diagnostic Problem

Our context is a medical documentation and consultation system. In our application domain of sonography, the examination considers several partially disjunctive subdomains, eg., liver or kidney, which results in multiple disorders. To support diagnosis, we want to retrieve experiences such as explanations for a query case based on the presented similarity to former cases. We also want to retrieve additional information about therapy, complications, prognosis or the treating physician as contact person for special questions.

We use case-based reasoning to retrieve experiences and help diagnosis. However case-based diagnosis handling multiple disorders differs from handling single disorder. For instance, for a single disorder casebase dealing with 100 disorders, the chance of reusing a case is roughly one to one hundred. However, due to the combinatorial rules, the chance of reusing a case with 3 independent diagnoses from 100 alternatives is just one to one million. In a real world setting, our medical case base contained about 7 disorders per case on average, and have 221 disorders in total, thus the chance of reusing an entire previous case is very small. Moreover, naive case-based reasoning approach which performs good on single disorder casebase shows a poor per-

formance. Naive case-based reasoning means to retrieve the most similar case to the query case and to adopt the solution of that similar case as the solution for the query case. This result indicates that we can not guarantee to retrieve a similar enough case and a desirable solution from a multiple disorder casebase.

In this paper, we argue that we can reuse portions of previous cases to construct disirable solution for a query case in the multiple disorder situation, although a similar enough entire case might be unreachable.

### 3. Intelligent Diagnostic System Using Multiple Case-based Reasoning

In this paper, we propose an integrated approach which aims at a real world application, where single disorder cases occurred with multiple disorder cases when these cases or data were collected by hospitals.

Before we introduce our method, we define necessary notions concerning our knowledge representation schema as follows: Let  $\Omega_D$  be the set of all diagnoses and  $\Omega_A$  the set of all attributes. To each attribute  $a \in \Omega_A$  a range  $dom(a)$  of values is assigned. Further we assume  $\Omega_F$  to be the (universal) set of findings ( $a = v$ ), where  $a \in \Omega_A$  is an attribute and  $v \in dom(a)$  is an assignable value. Let  $CB$  be the case base containing all available cases that have been solved previously. A case  $c \in CB$  is defined as a tuple

$$c = (\mathcal{F}_c, \mathcal{D}_c, I_c), \quad (1)$$

where  $\mathcal{F}_c \subseteq \Omega_F$  is the set of findings observed in the case  $c$ . In CBR-problems these findings are commonly called *problem description*. The set  $\mathcal{D}_c \subseteq \Omega_D$  is the set of diagnoses describing the *solution* for this case.  $I_c$  contains additional information like therapy advices or prognostic hints.

To compare the similarity of a query case  $c$  with another case  $c'$ , we apply Manhattan distance for continuous or scaled parameters

$$md(x, y) = \frac{1}{k} \sum_{i=1}^k W_i \left| \frac{x - x_{max}}{x_{max} - x_{min}} - \frac{y - x_{min}}{x_{max} - x_{min}} \right| \quad (2)$$

and Value Difference Metric ( VDM ) for discrete parameters (Wilson and Martinez, 1997).

$$vdm_a(x, y) = \frac{1}{|\Omega_D|} \cdot \sum_{D \in \Omega_D} \left| \frac{N(a = x|d)}{N(a = x)} - \frac{N(a = y|d)}{N(a = y)} \right| \quad (3)$$

where  $x$  and  $y$  are values of parameter  $a$  in case  $c$  and  $c'$  respectively.

### 3.1 Compositional CBR on Multiple Disorder Casebase

Compositional adaptation was originally developed for configuration tasks. It decomposes problems into sub-problems and retrieves those sub-problems in the casebase, and then combines different parts of the solutions of similar cases (Wilke and Bergmann, 1998). However, in a multiple disorder situation, we can not perform decomposition because the findings cannot explicitly be divided automatically into different groups which are corresponding to different diagnoses without the help of experts.

We assume that in the multiple disorder situation, not all the diagnoses in the solutions of the  $k$  most similar cases will be suggested as the final diagnosis. Only the diagnoses with a high occurrence among the  $k$  most similar cases have a high probability to appear in the final solution of the query case. The underlying meaning of this assumption is that those findings in the  $k$  similar cases which are similar to the query case will contribute to those desired diagnoses with a high occurrence in the similar case solutions.

At the same time, we assume that the more similar the retrieved case is to the query case, the higher the probability that the diagnoses in this retrieved case will appear in the final solution. Thus, we add weights to the frequency of diagnoses in the set of retrieved cases.

Given a query case  $C_q$ , we retrieve the  $k$  most cases, and calculate the similarity-weighted frequency  $Fqc$  for each diagnosis  $D_j$ .

**Definition 1 (Similarity-Weighted Frequency).** *The similarity-weighted frequency of a diagnosis  $D$  is the weighted frequency of  $D$  within the  $k$  most similar cases.*

$$Fqc(D_j) = \frac{\sum_{i=1}^k W_i \cdot \delta(C_i, D)}{\sum_{i=1}^k W_i}, \quad (4)$$

where  $D \in \Omega_D$  is a diagnosis;  $C_i \in CB$  is the  $i$ th most similar case to the query case;  $\delta(C_i, D)$  is 1, if  $D$  occurs in the  $i$  most similar case  $C_i$ , and 0 otherwise.  $W_i$  represents the associated weight, where we used the squared relative similarity between  $C_i$  and the query case  $C_q$ . Therefore, the weight is proportional to the similarity.

After we calculate the similarity-weighted frequency of the diagnoses appearing in the  $k$  most similar cases, we generate a candidate solution defined as follows:

**Definition 2 (Candidate Solution).** *A candidate solution*

$$CS = \{D \in \Omega_D : FQC(D) \geq \epsilon\}, \quad (5)$$

*is the set of diagnoses with a similarity-weighted frequency above a dynamic threshold  $\epsilon = \alpha * \max_{D \in \Omega_D} FQC(D)$ . This threshold has a linear relation-*

ship with the mode of the frequency value of the diagnosis in the  $k$  most similar cases and  $\alpha$  is a coefficient.

Thus, we only include a diagnosis into the candidate solution, if the similarity-weighted frequency of the diagnosis is greater or equal than the threshold defined by  $\epsilon$ . The diagnoses that do not appear in the  $k$  most similar cases are not considered.

We summarize our Compositional CBR approach as follows, and evaluate it in the subsequent evaluation section, i.e. in section 4.

#### CompositionalCBR algorithm

```

{
  Given a query case  $C_q$  and casebase  $MCB$ ,
   $KMostSimiCase = \mathbf{Retrieve\_kMostSimi}(C_q, MCB)$ ;
  for( int j=0; j<  $\mathbf{Disorder\_Number}(MCB)$ ; j++)
  {
     $Fqc(D_j) = 0$ ;
     $Fqc(D_j) = \mathbf{Calculate\_Fqc}(D_j, KMostSimiCase)$ ;
    if ( $Fqc(D_j) \geq \epsilon$ )
    {
       $\mathbf{Add\_Solution}(D_j, solution)$ ;
    }
  }
  return  $solution$ ;
}

```

### 3.2 Multiple Case-based Reasoning System

To deal with the real world situation in which single disorder cases occur together with multiple disorder, we separate the existing casebase into single disorder casebase (SDCB) <sup>3</sup> and multiple disorder casebase (MDCB) <sup>4</sup>, and integrate compositional case-based reasoning and naive case-based reasoning.

Given a query case  $C_q$ , we use a ReasonerSelector component to find out which case-based reasoner should be applied for this case. According to the result from the ReasonerSelector component, naive CBR or compositional CBR will be applied. Compositional case-based reasoning retrieves a group of the most similar cases in multiple disorder casebase, then use compositional adaptation to get portions of the solutions of the most similar cases to get the final solution, while naive case-based reasoning retrieves the most similar case and adopts the solution of that retrieved case. After that, we combine this query case with the candidate solution and restore it either into the single or the multiple casebase according to the number of diagnoses in candidate solution.

**Multiple Case-based Reasoning algorithm**

```

{
  Given a query case  $C_q$ , casebase  $SDCB$ , casebase  $MDCB$ ,
   $selectSingle = \mathbf{ReasonerSelector}(C_q, SDCB, MDCB)$ ;
  if (  $selectSingle$  )
    {
       $solution = \mathbf{TraditionalCBR}(C_q, SDCB)$ ;
       $blSingle = true$ ;
    }
  else
    {  $solution = \mathbf{CompositionalCBR}(C_q, MDCB)$ ; }
   $SolvedCase = \mathbf{Construct}(C_q, solution)$ ;
  if (  $blSingle == true$  )
    {  $\mathbf{Restore}(SolvedCase, SDCB)$ ; }
  else
    {  $\mathbf{Restore}(SolvedCase, MDCB)$ ; }
}

```

In the ReasonerSelector component, we retrieve the most similar case from both SDCB and MDCB respectively, and mark the corresponding similarity value as  $S_s$  and  $S_m$ . If  $S_s$  is greater than  $S_m$ , then we assume that the final solution for this query case has more potential to be a single disorder, thus we use naive CBR to retrieve the most similar case and adopt the solution for that most similar case. If  $S_s$  is smaller than  $S_m$ , then the solution comes up after compositional adaptation is applied. If  $S_s$  is equal with  $S_m$ , we prefer to use the single disorder casebase to get a single disorder solution, according to the Parsimony goal ( also referred as ‘‘Occam’s razor’’, the goal of minimizing the complexity of explanation (Peng and Reggia, 1990) ).

**4. Evaluation**

For our evaluation we applied cases from the knowledge-based documentation and consultation system for sonography SONOCONSULT, an advanced and isolated part of HEPATOCONSULT (Huettig et al., 2004). Our evaluation case base consisted of 744 cases, among which there are 65 single disorder cases and 679 multiple disorder cases. The case base contains an overall number of 221 diagnoses and 556 symptoms, with a mean  $M_D = 6.71 \pm 04.4$  of diagnoses per case and a mean  $M_F = 48.93 \pm 17.9$  of relevant findings per case.

In the usual task of assigning an example to a single category, the accuracy is just the percentage of cases which are correctly classified. But to quantitatively measure the accuracy of multiple disorder diagnosis, the simple accuracy measurement doesn’t fit. Standard accuracy has also be demonstrated to be not

very suitable for multiple disorder problem (Cynthia A and Mooney, 1994), which is defined as  $(T^+ + T^-)/N$ , where  $T^+$  (True Positives) is the number of disorders in the correct diagnosis that are also in the system diagnosis,  $T^-$  (True Negatives) is the number of disorders not in the correct diagnosis and not in the system diagnosis, and  $N$  is the total number of disorders.

We adopted the Intersection Accuracy (Cynthia A and Mooney, 1994), as a measure for multiple disorder problems. Intersection accuracy is derived by the two standard measures: *sensitivity* and *precision*.

**Definition 3 (Intersection Accuracy).** *The Intersection Accuracy  $\mathcal{IA}(c, c')$  is defined as*

$$\mathcal{IA}(c, c') = \frac{1}{2} \cdot \left( \frac{|\mathcal{D}_c \cap \mathcal{D}_{c'}|}{|\mathcal{D}_c|} + \frac{|\mathcal{D}_c \cap \mathcal{D}_{c'}|}{|\mathcal{D}_{c'}|} \right) \quad (6)$$

where  $c$  and  $c'$  are two cases,  $\mathcal{D}_c \subseteq \Omega_D$  is the set of diagnoses of case  $c$ , and  $\mathcal{D}_{c'} \subseteq \Omega_D$  is the set of diagnoses contained in case  $c'$  likewise.

We used leave-one-out cross-validation which is a variation of k-fold cross validation, where each fold consists only of exactly one case.

We compare our method with Naive CBR, Set-Covering method (Baumeister et al., 2002) and Partition Class method (Atzmueller et al., 2003). These four methods were implemented respectively and evaluated using the same casebase. the set-covering approach combined case-based reasoning and set-covering models for diagnosis. The partition class method uses partitioning knowledge provided by the expert to split cases into several parts. Decomposed cases are retrieved and combined to get the candidate solution. The evaluation results are shown in the following table.

744 Cases from the SonoConsult Case Base		
<i>Approach</i>	<i>solved cases (percentage)</i>	<i>mean acc</i>
<i>Naive CBR</i>	20 (3%)	0.66
<i>Set-Covering</i>	502 (67%)	0.70
<i>Multiple CBR</i>	536 (72%)	0.70
<i>Partition Class</i>	624 (84%)	0.73

Table 1. Comparison of the approaches, using 744 cases

The results in the first line show, that the Naive CBR method performs poor with cases having multiple disorders. Naive CBR utilizing no adaptation and no additional background knowledge can only solve 3% of the cases in the case base, which is obviously insufficient. The Multiple CBR method solves 536, i.e., 72% of the cases in the case base, with a mean accuracy of 0.70, which

performs significantly better than naive CBR. This demonstrates the relevance of this method in the multiple disorder situation.

Multiple CBR is slightly better than the set-covering approach. This is probably due to two issues: The set-covering approach returns candidate cases in terms of cases with all their solutions and no sophisticated adaptation step is applied. The knowledge-intensive method using partition class knowledge performs best. However the multiple CBR method and the set-covering approach do not need background knowledge, and so can be applied in arbitrary situations when the partitioning knowledge is not available, while the partition class strategy needs additional background knowledge.

Furthermore, we apply naive case-based reasoning and compositional case-based reasoning respectively on both the single disorder casebase (SDCB) and multiple disorder casebase (MDCB). This demonstrates the aptness of naive CBR on SDCB, and compositional CBR on MDCB, as shown in table 2.

65 cases in SDCB, 679 cases in MDCB		
<i>Approach</i>	<i>solved cases (percentage)</i>	<i>mean acc</i>
<i>Naive CBR on SDCB</i>	57 (88%)	0.68
<i>Naive CBR on MDCB</i>	125 (18%)	0.71
<i>Compositional CBR on SDCB</i>	19 (29%)	0.63
<i>Compositional CBR on MDCB</i>	561 (83%)	0.69

Table 2. Aptness of Naive CBR on SDCB and Compositional CBR on MDCB

The results in the first two lines show, that the naive CBR method performs good for cases with single disorder while it performs relatively poor for cases with Multiple disorder. Compositional CBR shows more competence on handling multiple disorder cases, which solved 83% multiple disorder cases.

## 5. Discussion

There are several points worth noting about our approach. Firstly, the case-based reasoning method itself corresponds to the diagnosing process that physicians always use. They recall former similar diagnostic case and compare the symptoms with those current patient have, make some adjustment based on the previous solution to adapt the current symptoms the patients have. The cognitive similarity between Case-based Reasoning and diagnosis makes it easy to get user acceptance.

Secondly our method involves two case-based reasoners working together for handling different casebases. This is different from most case-based reasoning systems using simplex reasoners. It evokes an idea of using multiple case-based reasoner, each of which may be suitable for different casebase or dataset.



Thirdly, our system deals with the problem of multiple disorder which hasn't been identified by most knowledge-based diagnostic systems (Gierl et al., 1998). This is due to the single disorder assumption, which assumes to use only one disorder to explain all the findings presented (Peng and Reggia, 1990).

Forthly, our approach uses flexible knowledge, and allows the automatic generation of the knowledge base from an existing database, which not only makes the CBR system easy to integrate into existing clinical information systems, but also, to some extent, avoids knowledge acquisition problem.

## 6. Related Work on Multiple Disorders

INTERNIST matches symptoms and diseases in general internal medicine based on forward and backward conditional probabilities (Miller et al., 1982). But it does not deal with interacting disorders properly because if the findings can be explained by a disorder, then these findings will be deleted immediately no matter how these findings could also lead to diagnosis of another disorder.

HYDI decomposes knowledge from the causal models into diagnosis units to prevent re-computation for similar problem to improve efficiency (Jang, 1993). But the diagnosis units in HYDI largely rely on the causal models which have been built in Heart Failure Program (HF) on heart disease. Only when all the causal models for other disorders are available, could HYDI's method be applied to diagnose other disorders.

HEPARII (Onisko et al., 2000) extended the structure of Bayesian network and (Gaag and Wessels, 1995) use belief networks to diagnose multiple disorder, but they are both based on the medical literature and conversations with medical domain experts, which highlights knowledge acquisition problem.

## 7. Conclusions and Outlook

In this paper, we introduce a multiple case-based reasoning approach to deal with multiple disorder problems. We apply compositional case-based reasoning to construct diagnostic solution from multiple disorder casebase and employ naive case-based reasoning to access single disorder casebase. Using real medical data, this method has been demonstrated to be promising.

There are many opportunities for future work. Firstly, our method has a potential source of error, the decision of which case-based reasoner should be chosen. We will investigate it in more detail in the future. Secondly, we believe that employing learning methodology to explore interactions between disorders will help to filter the candidate disorder or to add potential disorder during case adaption. Thirdly, experiments in other domains are desirable. Our work has the potential to be used to diagnose multiple faults in other diagnostic problem areas, such as diagnosis problems concerning machine faults.

## Notes

1. Our Multiple Case-based Reasoning method is different from Multiple-Case-base reasoning (Leake and Sooriamurthi, 2003) which emphasizes case-based reasoning on multiple casebases, while our multiple case-based reasoning focuses on integrating multiple case-based reasoners together.
2. Here Multiple Disorders is in a broad sense, which includes single disorders and multiple disorders.
3. In a single disorder casebase, one disorder is ultimately used to explain all the findings for each case, although the single disorder casebase could have several disorders.
4. This multiple disorder is in a narrow sense, the multiple disorder casebase is a casebase with multiple diagnoses, we also call this multiple disorder only casebase.

## References

- Atzmueller, M., Baumeister, J., and Puppe, F. (2003). Evaluation of two strategies for case-based diagnosis handling multiple faults. In *Proceedings of the 2nd German Workshop on Experience Management (GWEM 2003)*, Luzern, Switzerland.
- Baumeister, J., Atzmueller, M., and Puppe, F. (2002). Inductive learning for case-based diagnosis with multiple faults. In S. Craw and A. Preece, editors, *Advances in Case-based Reasoning (ECCBR2002)*, pages 28–42. Springer Verlag. Proceedings of the 6th European Conference on Case-based Reasoning.
- Cynthia A, T. and Mooney, R. J. (1994). Inductive learning for abductive diagnosis. In *Proc. of the AAAI-94*, volume 1, pages 664–669. [citeseer.nj.nec.com/thompson94inductive.html](http://citeseer.nj.nec.com/thompson94inductive.html).
- Gaag, L. and Wessels, M. (1995). Efficient multiple-disorder diagnosis by strategic focusing. In Gammerman, A., editor, *Probabilistic Reasoning and Bayesian Belief Networks*, pages 187–204, London. UCL Press.
- Gierl, L., Bull, M., and Schmidt, R. (1998). Cbr in medicine. In Lenz, M., Bartsch-Sporl, B., Burkhard, D.-D., and Wess, S., editors, *Case-based Reasoning Technology: From Foundations to Applications*, pages 273–297. Springer-Verlag. ISBN 3-540-64572-1.
- Huettig, M., Buscher, G., Menzel, T., Scheppach, W., Puppe, F., and Buscher, H.-P. (2004). A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. *Medizinische Klinik*, 99(3):117–122.
- Jang, Y. (1993). *HYDI: A Hybrid System with Feedback for Diagnosing Multiple Disorders*. PhD thesis, Massachusetts Institute of Technology.
- Leake, D. B. and Sooriamurthi, R. (2003). Dispatching cases versus merging case-base: When mcbr matters. In *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS-2003)*, pages 129–133. AAAI Press.
- Miller, R. A., Pople, H. E., and Myers, J. D. (1982). Internist-1: an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 8(307):468–476.
- Onisko, A., Druzdzel, M. J., and Wasyluk, H. (2000). Extension of the heparii model to multiple-disorder diagnosis. In Kłopotek, M., Michalewicz, M., and Wierzbach, S. T., editors, *Intelligent Information Systems*, pages 303–313. Physica-Verlag.
- Peng, Y. and Reggia, J. A. (1990). *Abductive Inference Models for Diagnostic Problem-Solving*. Springer-Verlag.
- Wilke, W. and Bergmann, R. (1998). Techniques and knowledge used for adaptation during case-based problem solving. In *IEA/AIE (Vol. 2)*, pages 497–506.
- Wilson, D. R. and Martinez, T. R. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*.