Retrieval and Reasoning in Distributed Case Bases

M V Nagendra Prasad, Victor R Lesser and Susan Lander
Department of Computer Science
University of Massachusetts
{nagendra,lesser,lander}@cs.umass.edu

CIIR Technical Report IC-5

Abstract

Recent explosion in networked information resources has been attracting much attention to the problem of automated methods for gathering information in response to a query from a user. However, most of this literature deals with locating, gathering and selecting the best response to a query from among a multitude of responses from different sources. In this paper, we deal with a different model of response to a query. No single source of information may contain the complete response to a query and hence may necessitate piecing together mutually related partial responses from disparate and possibly heterogeneous sources. We present a system for cooperative retrieval and composition of a case in which sub-cases are distributed across different agents in a multi-agent system. From a Gestalt perspective, a good overall case may not be the one derived from the summation of best subcases. Each agent’s partial view may result in local cases that are best matches based on the local view. However, these local cases when assembled may not result in the best overall case in terms of global measures. We propose a negotiation-driven case retrieval algorithm as an approach to dynamically resolving inconsistencies between different case pieces during the retrieval process.

1 The work reported here is supported in part by NSF Center for Intelligent Information Retrieval (CIIR), ARPA contract N00014-92-J-1698, Office of Naval Research contract N00014-92-J-1450, the National Science Foundation contract CDA 8922572. The content of the information does not necessarily reflect the position or the policy of the Government and no official endorsement should be inferred.
1 Introduction

Recent explosion in networked information resources has been attracting much attention towards automated methods for gathering information in response to a query from a user [1, 4, 21]. However, most of this literature deals with locating, gathering and selecting the best response to a query from among a multitude of responses from different sources. In this paper, we deal with a different model of response to a query. No single source of information may contain the complete response to a query and hence may necessitate piecing together mutually related partial responses from disparate and possibly heterogeneous sources [21]. For example, imagine a multi-agent parametric mechanical design system where each agent has to contribute a component to assemble a complete design. An agent has to obtain its components from manufacturer-specified catalogue casebases that may be distributed across the Internet. For example, it may treat WWW-accessible distributed component database system like PARTNET [5, 22] as its case-base from which to draw its component. In addition to the constraints arising from the overall design requirements, constraints on a component could also arise from the specifications of other components in the design. A response to a query in the form of a specification of design requirements consists of the integrated set of components forming the design. Central to the presentation in this paper is the view that a complex query is presented to a set of agents, each of which is responsible for retrieving information relevant to a part of the query. The agents negotiate to piece together a mutually acceptable response to the query. More specifically, we deal with agents retrieving pieces of a distributed case to assemble a composite case from them.

How do distributed case bases arise in these systems? A system that performs rote learning by storing good cases, where each agent stores its own local case in its case base, could give rise to such a Distributed Case Base (DCB). However, this may not be the only way. In the case of a set of reusable agents [16], each agent could have acquired its own independent problem-solving experiences by participating in different teams of agents. Or a Case-Knowledge Engineer could design each of the case bases by giving them episodes from his collection of cases. Another scenario one could envisage is the existence of casebases spread across a communication network, like PARTNET on World Wide Web.

A multi-agent system comprises a group of intelligent agents working towards a set of common global goals or separate individual goals that interact [3]. In such a system, each of the agents may not be individually capable of achieving the global goal and/or their goals have interactions and hence a need for coordination among the set of agents arises. Due to its partial view of the problem-solving situation, an agent may have access only to a part of the environment, and communication bandwidth limitations and heterogeneity of representations may limit its view of other agents’ state. Given these uncertainties [18], an agent may have to do evidential reasoning and focused negotiation to resolve them to the extent that it can make positive contributions to the on-going problem solving process. More specifically, in a distributed case-based reasoning system (DCBR), each agent’s partial view may result in best local cases that when assembled may not result in the best overall case in terms of global measures. This gives rise to a need for the agents to cooperatively access their casebases to retrieve the “best sub-cases” which can be combined in a consistent manner to provide a complete case most useful for the present problem-solving situation. This paper is concerned with this kind of cooperative retrieval of compatible distributed cases.
our work on Negotiated Search[17] and DRESUN[6], we proposed certain general mechanisms for performing distributed search in multi-agent systems. We draw upon our experience with these methods to extend them to case-based systems in a distributed environment.

The issue of distributed casebases, to our knowledge, has not been studied in the multi-agent context. Barletta et al[2] and Redmond et al [23] deal with distributed cases in single agent systems. In these approaches, each case is divided into subcases or snippets and a snippet is indexed using both global goals and the local context of that snippet within the case. This kind of elaborate engineering in the form of indexing the case pieces using both global and local problem solving contexts may not be feasible for multi-agent CBR systems. The agents may only have a partial view of the global problem solving context and the internal context of a case piece. Case bases for individual agents may be built independently, without the knowledge of the kinds of problem solving systems in which they are going to participate. In this paper, we propose an alternative to using elaborate indexing to avoid interactions when case pieces are re-instantiated in a new problem solving context. Case pieces are iteratively retrieved and assembled into a case, dynamically resolving any conflicts that arise during the process through negotiation among the participant agents. There are a few other important issues to note here. It may be that even though the cases for individual agents were derived from past problem-solving experience, there could be combinations of these subcases that may be assembled into a case that the system as a whole has never seen before. As discussed later, this has some important implications in DCBs. In addition, local case integration does not require that the overall case be completely represented at any one node; in certain situations, the distributed case components are integrated only by their mutual consistency.

The rest of the paper is organized as follows. Section 2 discusses the Negotiated Retrieval strategy for distributed cases. Section 3 introduces the CBR-TEAM system as our domain to illustrate the algorithm, and provides a trace of the system run on an example. Section 4 presents some empirical results from runs of an abstract version of CBR-TEAM. Section 5 relates our work to some of the existing literature and Section 6 concludes the paper with a discussion of the implications of the proposed mechanism and future work.

2 Negotiated Case Retrieval

Reasoning with cases drawn from a casebase that is a component of a DCB presents an agent with additional uncertainties versus single agent CBR systems. As discussed previously, each agent has to rely on its possibly incomplete local view of problem-solving to retrieve a local case that best contributes to the overall case. However, a combination of the best local cases for the agent set need not lead to a best overall case because it may happen that the sub-cases cannot be effectively put together or there may be requirements on the solution that cannot be ascertained until the sub-cases are aggregated. Thus, agents need to augment their local views with constraining information from other agents to achieve the retrieval and assembly of a better overall case.

A simple way to overcome the problem of distributed retrieval is to let the set of agents form their DCB by each agent storing its sub-case when the agents come across a good case and labeling the sub-cases under the same unique label. During retrieval, each agent retrieves a set of sub-cases

---

1 In this paper, we view negotiation as a form of goal-directed communication among agents.
which are good matches. Each of the agents exchanges the labels of its locally retrieved cases with other agents and the agents try to assemble a complete case with a unique label. If they fail, they can expand their retrieval set, or if most of the case is assembled under a unique label, those agents with incompatible labels can do another round of retrieval for sub-cases with the label of the overall case\(^2\). However, this strategy may not work due to a variety of reasons. The casebases may not be formed with the same set of agents participating in the task. They may be formed in any of the ways discussed previously. This eliminates the possibility of labeling the corresponding local cases belonging to a unique case by the same label; local cases are not assembled from the cases seen by these agent sets. In addition, the agent sets may themselves be varying - some agents may go off-line and some other new ones may come on-line - during the problem-solving process. Yet another reason for the infeasibility of this strategy is the possible heterogeneity of the different agents. Unique labeling schemes may not be possible in heterogeneous systems with disparate representations.

In view of the discussion above, we propose that the agents need to perform a coordinated retrieval of local cases through communication of relevant non-local information. We will now elaborate on a negotiated case retrieval (NCR) strategy which is more general than the unique labeling scheme. The NCR strategy involves each agent asynchronously executing one of the set of possible operations: initiate a seed subcase, extend an existing partial case, merge existing partial subcases or inform others about a new partial case. Initiating a seed subcase involves an agent retrieving a local subcase from its local casebase using the local problem solving state and the relevant portion of the user specification and forming a seed subcase which can be extended by local cases from other agents to obtain a complete case. An agent intending to extend a subcase from another agent obtains the subcase’s relevant feature values that serve as an anchor for the local case retrieval, the result of which is integrated with the corresponding partial case. Merge is similar to the extend operation. An agent intending to merge one of its chosen partial cases with another agent’s partial case obtains the relevant feature values and performs the merge operation. The inform operation involves an agent telling others about the existence of a newly formed partial case that results from the local execution of one of the three previous operators. An extend or merge operation involves checking for any violations of local constraints by the set of feature values from the non-local partial subcase and the local case or partial subcase. Detection of such violations leads to an interaction process among the agents by which they negotiate on conflict resolution alternatives. The negotiation process involves an agent communicating feedback to other agents on the causes and possible resolutions for each of the constraint violations. The receiving agents assimilate this feedback, leading to an enhanced view of the global requirements for future operations. The subsequent initiation or extend or merge is more likely to avoid the same conflicts. Thus, our problem is cast as a distributed constraint optimization problem\(^16\) implying that not all constraints need to be satisfied in a solution. As many constraints are satisfied as is possible. Constraints have differing amounts of flexibility. Some may be hard, meaning that they must be satisfied in any legal solution. Some others may be soft constraints that may be relaxed as and when needed. Figure 1 shows the conceptual view of a two agent DCBR system. Before we formally present our NCR strategy, we need to introduce some notation.

Let the set of agents be denoted by \(\mathcal{T}\). Let \(\Omega_i\) denote the locally known constraint set of agent

\(^2\)Note that even this strategy is not a solution to all problems that can arise in this scenario and in the worst case, the labeled sub-case retrieval is as complex as unlabeled sub-case retrieval.
\( A_i \). Let \( \Omega_i^p, 1 \leq p \leq |\Omega_i| \) represent the \( p \)th constraint at \( A_i \) and let \( \mathcal{F}_i^p \) represent the set of features on which \( \Omega_i^p \) is defined.

Let the user-specified query be denoted by \( \Psi \). The part of the query that is relevant to \( A_i \) is denoted by \( \Psi_i \) which is cast as a set of constraints \( \Omega_{\Psi_i} \subseteq \Omega_i \). Casting both user specifications and the set of required subcase consistency constraints as the set of hard constraints for an agent gives us a uniform way to ensure that the user specifications are satisfied and the subcases are mutually acceptable. Subcase consistency constraints arise from the knowledge an agent has about the requirements of the context in which its local subcase can usefully participate\(^\text{3}\). The output of the system is a case, whose subcases are mutually acceptable to the entire set of agents. As discussed previously, the subcases need not all reside at a single physical site and may be integrated just by their mutual consistency.

Let a locally retrieved case \( t \) by agent \( A_i \) be denoted by \( \sigma_i^t \) (we simply denote it as \( \sigma_i \) when the name of the case is not important). A partial case \( s \), denoted by \( \Sigma_s \), is composed of one or more locally retrieved cases (from different agents):

\[
\Sigma_s = \sigma_{i_1}^{t_1} \sigma_{i_2}^{t_2} \ldots \sigma_{i_z}^{t_z}
\]

where \( 1 \leq z \leq |\mathcal{Y}| \)

\(^3\)For example in TEAM[16], from which our DCBR system draws its inspiration, an agent’s sub-design possesses shared features with another agent’s sub-design. Pump and Motor share a parameter \textit{required-power}. Both agents have to arrive at the same value for the \textit{required-power} parameter. This is equivalent to having a constraint \( \text{Pump-required-power} = \text{Motor-required-power} \).
We will introduce a simple example from the ABSTRACT-TEAM system to be discussed in the later sections. Let the agent set be \( \{ A_1, A_3 \} \). In ABSTRACT-TEAM, a query is denoted by specifying the ranges of the features \( f_2 \) and \( f_7 \). For example, \( 0.03 \leq f_2 \leq 2.0 \) & \( 10 \leq f_7 \leq 15 \). \( \Psi_1 \) is \( 0.03 \leq f_2 \leq 2.0 \) and \( \Psi_2 \) is \( 10 \leq f_7 \leq 15 \). In our example, \( A_1 \) retrieves a local case which is simply a set of three features \( \{ f_1, f_3, f_8 \} \) and \( A_2 \) retrieves a local case \( \{ f_4, f_9, f_6, f_7 \} \). \( \Omega_1 \) contains four constraints \{C1: \( f_2 \geq 0.03 \), C2: \( f_2 \leq 2.0 \), C3: \( f_4 \geq 0.55 \), C4: \( f_4 \leq 0.75 \)\} and \( \Omega_2 \) has three constraints \{C5: \( f_7 \geq 10.0 \), C6: \( f_7 \leq 15.0 \), C7: \( f_2 \leq 0.1 \)\}.

A case has a number associated attributes that involve measures of certain characteristics of a case and are functions of the feature values of a case. Examples of attributes include reliability, quality, uncertainty and cost. In addition to being acceptable to all agents, it is desirable that a case be optimized along the attribute set. These requirements lead to organization of an agent’s constraints as soft constraints and hard constraints where the former set represents solution preferences and the later set represents those constraints that are relaxed only as a result of explicit recognition by the agents that the set is too constrained to lead to a mutually acceptable solution[19]. Relaxing a soft constraint may only involve penalties in terms of loss of optimality in the desirable attributes. For example, relaxing a soft constraint in a multi-agent design system like CBR-TEAM (to be discussed below) leads to less reliable components. Softness of a constraint represents its degree of flexibility with some constraints being softer than others. On the other hand, hard constraints are generally not relaxed except with an explicit understanding that the resulting responses satisfy the query specifications only partially or are consistent with only a subset of the agents.

More formally, let a local solution space \( S \) be defined as a region in an \( n \)-dimensional feature-space \( R^n \). Each point in this space represents a local case of \( n \) features. The set of hard-constraints in an agent defines \( S \). We can then define subspaces of \( S \) that represent satisfaction of various sets of soft-constraints in addition to the set of hard constraints. Given \( S_c \), a subspace of \( S \) satisfying a set of soft-constraints \( c \) and the set of hard constraints, relaxing a constraint \( c_i \in c \) leads to a subspace \( S_{c-c_i} \) such that

\[
S_c \subset S_{c-c_i} \subset S
\]

An agent iteratively identifies those regions in its local solution space \( S \) that satisfy as many of the soft constraints as possible and the relevant non-local requirements of the other agents.

Adding a locally retrieved subcase \( \sigma_{lm} \) to \( \Sigma_s \) to form \( \Sigma_s \) is represented as:

\[
\Sigma_s = \Sigma_s \sigma_{lm}
\]

Let the set of features for \( \Sigma_s \) be represent by \( F_{\Sigma_s} \). Note that the feature set \( F_k \) of a constraint \( \Omega_k \) may involve features that are not members of the feature set of the local subcase. Such features are called non-local features (or non-local parameters). For example, \( f_4 \) is a non-local parameter for \( A_1 \).

A projection of a partial case \( \Sigma_s \) onto \( \Omega_k \), the constraint set of agent \( A_k \), represents the set of features from \( \Sigma_s \) that participate in some constraint of \( \Omega_k \).

\[
\Sigma_k^s = \{ f \mid f \in F_{\Sigma_s} \land \exists 1 \leq p \leq |\Omega_k| f \in F_k \}
\]

\(^4\)In a specific problem-solving situation, an agent can substitute for its local features from the problem-specification in its local constraint to get new constraints that are described only on non-local features. Though the example here deals with simple numerical constraints, note that the mechanisms presented in the paper are general enough to deal with constraints of any form.
An agent $A_i$ wanting to obtain the projection of a partial case $\Sigma_s = \sigma^{i_1}_s \sigma^{i_2}_s \ldots \sigma^{i_t}_s$ communicates with agents $i_1, i_2, \ldots, i_t$ to obtain the those relevant feature values that were not already communicated to the agent previously. An agent $A_{ik}$, upon a request for the value of a particular feature in a local case $\sigma^{i_k}_s$, responds with transmission of the value to the requesting agent. Each agent’s organizational knowledge\(^5\) about the distribution of relevant features among the set of agents can aid it in deciding which agents to ask for what features. Projection is the minimum information an agent needs to check the consistency of a partial case against its local constraint set. However, an agent which anticipates and intends to serve the needs of other agents may obtain more information than the projection and pass it on to these other agents, saving them the need to communicate with a number of agents to obtain a projection. We have here a trade-off between the number of communication events versus the amount of information communicated in these events. An agent obtaining more information than a projection increases the number of bytes of communication but can reduce the necessity for some other agents to communicate with a number of agents to obtain their projections and consequently reduces the communication events.

For example, in order to check a partial case against $\Omega_1$, $A_1$ has to obtain its projection by requesting $A_2$ for the value of feature $f_4$ of the corresponding local case, as $f_4$ is the only feature participating in some constraint in $\Omega_1$ and is not available locally. Similarly, $A_2$ requests $A_1$ for the value of feature $f_2$ of the corresponding local case to validate a partial case against $\Omega_2$.

Each agent can choose to execute one of the following operators at a particular instance of time.

- **Initiate a seed subcase**:  
  Input: Agent $A_i$ that chooses to initiate a seed subcase uses the problem specification $\Psi_i$, and locally known problem-solving state. 
  Output: A partial case $\Sigma_s = \sigma^i_s$.
  
  $A_i$ retrieves a subcase from its local casebase and forms a partial subcase with just one local subcase. It uses the problem specification $\Psi_i$ and the presently available information on the problem-solving state (including previously tried solutions, conflicts they caused and feedback in the form of violated constraints from other agents) to achieve this task\(^6\).

- **Extend an existing partial case**:  
  Input: $\Sigma^i_s$ to $A_i$ 
  Output: $\Sigma_s = \Sigma_s \sigma^i_s$ or a failure to extend. 
  
  $A_i$ decides to extend a partial subcase $\Sigma_s$ residing at $A_k$ and obtains the projection with respect to the local constraints $\Omega_i$. It tries to retrieve a local case based on the information available in $\Sigma^i_s$ and the locally known problem solving state. $A_i$ can either succeed or fail to return a subcase or return an infeasible subcase that violates some constraints in the set $\Omega_i$. If any violations are detected due to poor or infeasible values for features then feedback is provided to the relevant agents (the feedback process is discussed in detail below). Note that the extension operation need not assemble all the local cases in $\Sigma_s$ at any one physical site.

\(^5\)Organizational knowledge consists of a specification of general node interaction patterns\(^9\) or static meta-level information about knowledge/case organization in the local databases of the agents.

\(^6\)In general, a locally retrieved subcase is re-instantiated in the present context during this operation. Adaptation of the retrieved subcase to the new context could also be performed in some systems.
It just needs to record the labels of all the subcases that can be assembled without any local constraint violations with regard to the relevant agents.

- **Merge existing partial subcases:**
  Input: $\Sigma_i^i$ and $\Sigma_j^j$ to $A_i$
  Output: $\Sigma_u = \Sigma_i \Sigma_j$
  An agent $A_i$ tries to merge two partially assembled composite cases $\Sigma_{s_j}$ and $\Sigma_{s_k}$ from agent $A_j$ and $A_k$ respectively. Merging involves obtaining projections $\Sigma_i^{s_j}$ and $\Sigma_i^{s_k}$ and testing for violations on the set $\Omega_i$. If any violations are detected due to poor or infeasible values for parameters then feedback is provided to the relevant agents\(^7\).

- **Inform other agents about a partial subcase:**
  Input: A set of partial cases that are the result of initiate or extend or merge operations at an agent $A_i$.
  Output: A message sent to other agents.
  An agent $A_i$ forms a new partial case $\Sigma_u$ due the successful execution of an operation like initiate or extend or merge. Other agents have to be informed of the existence of this new partial case so that they can make their decisions on the next operator to execute taking into consideration this new partial case. An agent need not immediately inform all the others of the existence of every new partial case it forms. It can decide to inform the existence of only the “best” ones or to inform once every few time units, all the new partial cases that emerged during the last time window. An agent may also decide to inform others about more than just the existence of the cases. It may inform others about certain features or attributes of the partial cases to aid them in their choice and execution of their operators.

Agents perform an asynchronous parallel distributed search to obtain a good overall case from case pieces. At a given point in time, there may be more than one developing partial case.

Negotiation process involves an agent delivering feedback to another agent that in turn may decide to accept it. The details of this process are discussed below\(^8\).

A partial case $\Sigma_s$ is consistent with respect to $\Omega_k$ if the feature values of the partial case do not violate any constraint in $\Omega_k$. We denote this by consistent$\_{\Omega_k}(F_{\Sigma_s})$. Two subcases $\Sigma_s$ and $\Sigma_t$ are said to be consistent with respect to $A_k$ if the projections of $\Sigma_s$ and $\Sigma_t$ with respect to $\Omega_k$, $\Sigma^k_s$ and $\Sigma^k_t$ do not violate any constraints in $\Omega_k$, i.e. consistent$\_{\Omega_k}(F_{\Sigma^k_s} \cup F_{\Sigma^k_t})$.

Agent $A_k$ has a set of predicates $I_k = \{\tau_{1k}, \tau_{2k}, ..., \tau_{nk}\}$ which are applied to any detected inconsistent partial case projections $\{\Sigma_s^k, \Sigma_t^k\}$. Associated with each $\tau_{ik}$ is a set of advice that can act as a feedback to the agents corresponding to the subcases involved in the detected conflicts.

$$\Pi_{\tau_{ik}} = \{\pi_{\tau_{1ik}}, \pi_{\tau_{2ik}}, ..., \pi_{\tau_{nik}}\}$$

\(^7\)Though we talk of applying the entire set of constraints at an agent at once, this process could be distributed in time. An agent could apply different subsets of its local constraints at different times and perform evidential reasoning on partially satisfied sets of solutions to decide on further application of constraints like in DRESUN[6].

\(^8\)A similar method for generating feedback has been presented in [7].
FEEDBACK = \( \bigcup \{ \Pi_{\tau_{rk}}^{i} / \tau_{rk}(\Sigma_{s}^{h}, \Sigma_{t}^{h}) \text{ is true} \} \)

The set of advice could range from domain independent strategies to highly domain specific ones. For example

1. **Broaden the Retrieval:** If the retrieval is similarity based i.e. based on numerical measures of “closeness” of the retrieval feature vector to the corresponding feature vector of a retrieved case, then broaden the search by obtaining cases with poorer similarity values.

2. Some CBR systems retrieve a case and use an adaptation strategy to massage the retrieved case to fit the new situation. An agent could advise another agent to modify the retrieved case in a different way - use a different adaptation strategy.

3. Each agent is expected to have some knowledge of the importance of a particular parameter’s values and constraints, based on which it can advise another agent to relax a soft constraint involving certain parameters.

4. An agent can advise another agent to change the values or ranges of certain parameters in order to obtain better local solutions.

5. **More Generic Retrieval:** When an agent detects lack of progress either locally or at other agents (based on the projections it receives from those agents) it could advise some of them to relax their hard constraints. This is expected to take the retrieval process to qualitatively different regions of the case base. Just as with soft constraints, the choice of which constraint to relax is based on system-wide knowledge of some sort or on generic strategies each agent possesses.

6. An agent recognizes particular features of the solution space and decides to use a more efficient customized search strategy. It can advise the other agent to play a particular role in this customized search. Lander[16] presents a good example of a customized search called *linear compromise* where agents, upon recognizing the linear nature of their solution space, decide to exchange end points and extrapolate between them to find the intersection point as a mutual compromise solution.

For example, in ABSTRACT-TEAM, the feedback is very simple. When an extend or merge violates a constraint, the constraint is communicated as feedback to agents whose local case features include the feature on which the constraint is defined. For example, agent \( A_2 \) tries to merge its locally retrieved case \((f_7 = 12) (f_6 = 0.0058) (f_5 = 0.00154) (f_4 = 1.55)\) with \( A_1 \)’s local case \((f_3 = 0.69) (f_2 = 0.48) (f_1 = 5)\) and finds that the constraint \( C7: f_2 \leq 0.1 \) is violated. For each agent \( A_i \), the \( \tau_{rk} \) is simply the violated \( \Omega_{i}^{k} \). Each \( \Pi_{\text{violated}}^{i} \Omega_{i}^{k} \) is \( \{ \Omega_{i}^{k} \} \). In our example, \( A_2 \) communicates \( C7 \) to \( A_1 \).

Thus, in addition to the three operators we discussed previously for creating and extending partial subcases, agents also have the following two operators for communicating and assimilating feedback to/from other agents.
Send FEEDBACK on conflicts:
Agent $A_i$, upon detection of conflicts during a merge or an extend, creates the FEEDBACK set and communicates a subset of it to the relevant agents.

Assimilate FEEDBACK from other agents:
The process of assimilation enhances an agent’s view of the global requirements. Agent $A_i$, upon receiving a FEEDBACK set, assimilates it into its local constraint set, creating a new and updated $\Omega_i$. Assimilation may involve processes like relaxing a constraint $\Omega_i^k$, or adding a new constraint to $\Omega_i$. Note that the assimilation of FEEDBACK need not be instantaneous and different advice may be assimilated at different times. For example, advice may be assimilated only after it is repeatedly received for at least ‘$x$’ times. In addition, assimilation may be specialized based on context i.e. the feedback assimilated is applicable only in specific contexts at the local agent. For example, a particular constraint may be applicable only if certain features are within certain ranges of values[8]. In addition, assimilation process may also involve transformations where an agent uses the feedback from other agents to generate its own local constraints rather than directly incorporating the feedback. Such a transformation may follow a process similar to that of generating feedback, where a series of transformation operators may be applied to each piece of advice to generate local constraints.

In both the CBR-TEAM and ABSTRACT-TEAM systems to be discussed later, the constraints are numerical and apply to a single parameter. The feedback generated from the violation of such a constraint simply consists of that constraint. Thus agent $A_i$ receives the constraint $f_2 \leq 0.1$ as feedback and the next local case retrieval avoids any case with feature $f_2 \leq 0.1$. However, note that the feedback strategies enumerated above are more general and apply to situations involving more than just single parameter constraint violations. For example, say agent $A_i$ detects a violation of a constraint $I^2 R + I^2 / C \leq 1$, where the capacitor value C is a non-local feature. Based on the domain knowledge, $A_i$ may advise agent $A_j$ whose local case contains C as its feature, asking it to initiate a case with capacitance in the next range of Farads. When this case is merged at a later time with a local case of $A_i$, there is a lesser chance that the same constraint is violated again.

How do the agents decide on which operator to instantiate next and what partial cases on which to apply the chosen operator? This decision is intricately tied to the domain. Domain imperatives dictate the preconditions for each of these operators. These preconditions can either be coded in by the DCBR designer or can be learned\(^9\). The specifics of domain heuristics for the choice of operators and partial cases to work on at a given time are beyond the scope of this paper.

Termination criteria depend on problem solving requirements. Various termination criteria ranging from simple heuristics to complicated decision theoretic methods have been proposed for multi-agent systems[8]. One simple criteria used in our DCBR systems to be discussed later is as follows: Any agent that detects a case that has subcase contributions by all the relevant agents and has been validated against their local constraints issues a termination message to all the agents. More formally, a complete case $\Sigma_u = \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_s}$ is defined as:

$$\forall \sigma_{i_m}, \sigma_{i_n} \in \Sigma_u, i_m \neq i_n \land \forall A_\alpha \in \Upsilon, \exists \sigma_{i_m} \in \Sigma_u, i_m = a \land consistent_{\Omega_\alpha}(F_{\Sigma_u})$$

\(^9\)NagendraPrasad, Lander and Lesser[20] discuss in detail the issues in learning the order of instantiation of such operators.
Algorithms similar to those of the negotiated retrieval already exist in the literature on multi-agent systems. DENEGOT system[19] presents a distributed negotiation algorithm for distributed planning and TEAM presents a negotiated search algorithm for distributed search among a set of heterogeneous agents[16]. Negotiated retrieval is specifically tailored to distributed casebases with strategies for broadening and deepening a search in a casebase not necessarily applicable to planning domains dealt with in DENEGOT. In addition, negotiating over conflicts in subcases can be qualitatively different from negotiations over plans in DENEGOT. For example, detection of interactions in DENEGOT involves elaborate reasoning about the local plans. In Case Based Systems, elaborate reasoning about cases may be infeasible because cases often implicitly encompass a lot of knowledge about subgoal interactions. Thus we resort to weak declarative constraint based knowledge to reason about possible interactions between subcases. The strategies in negotiated retrieval include the strategies in negotiated search as a subset. In this sense, negotiated retrieval is more general than negotiated search. In addition, negotiated search as discussed in Lander[16] imposes a sequentiality in search by giving an agent only two options — either extend an existing solution or initiate a new solution which other agents sequentially extend.

The following section presents a multi-agent system that incorporates a restricted form of negotiated retrieval (i.e. negotiated search as discussed in Lander[16]) and then shows a brief trace from the system. In the section on experiments, we discuss an abstract DCBR system and show some early empirical results on the negotiated retrieval mechanism.

3 CBR-TEAM: A Multi-Agent Design System

We now present a multi-agent system called CBR-TEAM whose core is derived from TEAM[16]. TEAM is a parametric design system that uses a cooperative heterogeneous set of reusable agents, each of which has the capability to produce a component of the overall design which is stored in a centralized blackboard. The TEAM system has six domain agents for the design of six components in a steam condenser and a critic agent that checks for certain features in the assembled design. It is a multi-strategy inferencing system[16] in which different strategies are seamlessly integrated into a design framework. The interactions between the components generate constraints for these strategies leading agents to iteratively negotiate on their results to find an acceptable design. When the components of the individual agents are being assembled, violation of constraints due to mismatches on mutually known parameters leads to information exchange followed by redesign. Agents whose components do not “match” are said to be in conflict. The conflicting set of agents negotiate a resolution which involves a search guided by numerical-valued constraints on the mutually known parameters, or other strategies like linear compromise[16]. CBR-TEAM is a modification of the TEAM system and comprises a subset of the agents that retrieve and use suitable members from catalogues of manufacturer-specified models for designing their components. These agents use the negotiated retrieval strategy where a new sub case is retrieved by each of the agents in conflict. This retrieval is guided by a set of simple numerical-valued constraints on the shared interface parameters. Interface parameters are those features of a case that are shared by more than one agent. All the relevant agents have to reach an agreement on the values of the shared parameters. We will discuss CBR-TEAM in more detail below.

The user gives a problem specification which consists of minimum head size for the pump in the required design. There are three agents named Motor-agent, Pump-agent and Vbelt-agent which
are responsible for design of the motor, pump and Vbelt components respectively. Each of these agents retrieves a suitable design from a catalogue of manufacturer specified designs. Indexing into a catalogue is based on design requirements and also the requirements of the components of other agents that may interact with the components of this agent. An agent does an *initiate* if it had no subcase to *extend*. If a conflict is detected a *send* operation is performed with the constraint that was violated. When an agent receives feedback from others, it immediately *assimilates* the feedback. The retrieved local cases are placed in a central blackboard and hence there is no need for an *inform* operator. Any partial case generated by an agent is completely visible to all the other agents. During the initial phase of retrieval, the agents may have only partial information on the requirements of other interacting components. So, each of the agents chooses the lowest cost design based on the information available to it. Trying to assemble these components into an overall design may lead to conflicts due to mismatches in the parameters that are shared by two or more components. For example, Vbelt and Pump have required-pump-power as a shared parameter and both the Vbelt agent and the Pump agent impose their own set of constraints on this parameter. A mismatch on this parameter involves one agent assigning a value to the parameter that violates the constraints in another agent. Each of the agents negotiate with the other agents in conflict to resolve any mismatches. This is done by posting locally generated requirements on the shared parameters that are involved in the conflict to the relevant agents. The agents then do another round of retrieval using the previous information and the new requirements from other agents to get better cases to be assembled into a design that does not produce the same conflict.

CBR-TEAM does not allow relaxation of hard constraints. It relaxes only soft constraints. Soft constraints are arbitrarily divided into four levels, 1 to 4, with 4 being the level that involves least loss of desirability if a constraint at this level is relaxed. Soft constraints are tied to the cost attribute of a component. Lower cost components satisfy more soft constraints. Violations of any hard constraints, considered to be at level 0, leads to failure. Another feature of CBR-TEAM is *unilateral relaxation*. When an agent finds that its progress has been stagnated for a long time, it unilaterally relaxes its local requirements so as to allow retrieval of subcases with poorer similarity to the requirements.

Note that the retrieval process is iterative and can happen many times until a case of required quality is obtained.

Below we give a simplified example trace of CBR-TEAM during a design session\(^{10}\). Problem specification, known to all the agents (that understand it) is the minimum head for the pump.

An illustrative subset of constraints for the Motor agent includes:
1. motor-horse-power ≥ pump-required-power, *level* 0
2. motor-drive-speed ≥ 1200, *level* 0
3. pump-required-power ≤ 5.0, *level* 1
4. pump-required-power ≤ 4.0, *level* 3

For the Pump agent:
1. motor-horse-power ≥ pump-required-power, *level* 0
2. pump-required-power ≤ 30, *level* 0

For the Vbelt agent:

\(^{10}\)There are a number of other constraints, but we just show an illustrative subset here.
Initially, all three agents have a local partial view of the problem solving situation. The agents have certain local constraints on the parameters, and based on these constraints they retrieve a template and develop the best component design. The Pump agent retrieves a PUMP TEMPLATE “model4-impeller43-wfr1” to build PUMP “pp_pump_1”. The Motor agent attempts to extend this by retrieving MOTOR TEMPLATE “motor10” and building MOTOR ”mp_motor_1”. However, it detects that this leads to a violation of soft constraints and hence enters into a negotiation with all the relevant agents (those that know about the parameters involved in conflict). In CBR-TEAM, the negotiation process is simple. The agent that detects any conflict on a parameter simply communicates the constraints whose violation led to the detection of that conflict. In this example, the Motor Agent detects and communicates conflict on the required power parameter. It sends out the constraint — (pump-required-power ≤ 4.0, 3) to the Pump agent. The Vbelt agent fails to retrieve any templates based on the local constraints.

The following figures show the designs formed by the Pump and Motor agents during this step. The Pump agent uses “model4-impeller43-wfr1” to build PUMP “pp_pump_1” which basically consists of slots filled in by parameters from the template and some other slots filled in by functions for calculating case attributes.

```
PUMP pp_pump_1 an object of class PUMP
MODEL: "model4-impeller43-wfr1"
WATER-FLOW-RATE: 104.125
MAX-HEAD: 288.63
AVAILABLE-HEAD: 257.5425
PUMP-REQUIRED-POWER: 10.987118
RUN-SPEED-RANGE: (2700.0 3300.0)
PUMP-RUN-SPEED: 3000
COST: 228.0625
WEIGHT: 50.503906
EVALUATION: (:PUMP-AGENT :GOOD)
ACCEPTABILITY: :ACCEPTABLE
```

The Motor agent retrieves MOTOR-TEMPLATE “motor10” from the template database to build MOTOR ”mp_motor_1”.

```
MOTOR "mp_motor_1" an object of class MOTOR
MODEL: "motor10"
PUMP-REQUIRED-POWER: 10.987118
SPEED-RANGE: NIL
WEIGHT-RANGE: NIL
HORSEPOWER: 15.0
MOTOR-DRIVE-SPEED: 2400
COST: 650.0
WEIGHT: 150
EVALUATION: (:MOTOR-AGENT :FAIR)
ACCEPTABILITY: :UNACCEPTABLE
```

The Vbelt agent fails to retrieve any templates based on the local constraints. These constraints are not tight enough to retrieve any templates that pass the tests setup by the local filters.

Note that during the time the Motor agent is extending a partial case seeded by the Pump agent, the Pump and Vbelt agents themselves could be initiating other seed subcases due to the
asynchronous nature of the distributed search.

Given the increased awareness of the Motor agent’s requirements, the Pump agent attempts a new round of retrieval and generates PUMP “pp\_pump\_4”. It takes the Motor agent’s communicated constraints into consideration while doing this. The Motor agent thus successfully adds its subcase MOTOR “mp\_motor\_4” and the Vbelt agent adds VBELT “vbelt\_1” giving the final design.

The Pump agent retrieves PUMP-TEMPLATE “model4-impeller43-wfr2” from the pump template database to build PUMP “pp\_pump\_4”.

<table>
<thead>
<tr>
<th><strong>PUMP pp_pump_4</strong></th>
<th>an object of class PUMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL:</td>
<td>“model4-impeller43-wfr2”</td>
</tr>
<tr>
<td>WATER-FLOW-RATE:</td>
<td>0.5</td>
</tr>
<tr>
<td>MAX-HEAD:</td>
<td>288.63</td>
</tr>
<tr>
<td>AVAILABLE-HEAD:</td>
<td>288.63</td>
</tr>
<tr>
<td>PUMP-REQUIRED-POWER:</td>
<td>0.1</td>
</tr>
<tr>
<td>RUN-SPEED-RANGE:</td>
<td>(2700.0 3300.0)</td>
</tr>
<tr>
<td>PUMP-RUN-SPEED:</td>
<td>3000</td>
</tr>
<tr>
<td>COST:</td>
<td>228.0625</td>
</tr>
<tr>
<td>WEIGHT:</td>
<td>50.503906</td>
</tr>
<tr>
<td>EVALUATION:</td>
<td>(:PUMP-AGENT :GOOD)</td>
</tr>
<tr>
<td>ACCEPTABILITY:</td>
<td>:ACCEPTABLE</td>
</tr>
</tbody>
</table>

The Motor agent retrieves MOTOR-TEMPLATE “motor1” from the template database to build MOTOR “mp\_motor\_4”.

<table>
<thead>
<tr>
<th><strong>MOTOR mp_motor_4</strong></th>
<th>an object of class MOTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL:</td>
<td>“motor1”</td>
</tr>
<tr>
<td>PUMP-REQUIRED-POWER:</td>
<td>0.1</td>
</tr>
<tr>
<td>SPEED-RANGE:</td>
<td>NIL</td>
</tr>
<tr>
<td>WEIGHT-RANGE:</td>
<td>NIL</td>
</tr>
<tr>
<td>HORSEPOWER:</td>
<td>1.0</td>
</tr>
<tr>
<td>MOTOR-DRIVE-SPEED:</td>
<td>2400</td>
</tr>
<tr>
<td>COST:</td>
<td>100.0</td>
</tr>
<tr>
<td>WEIGHT:</td>
<td>33</td>
</tr>
<tr>
<td>EVALUATION:</td>
<td>(:MOTOR-AGENT :EXCELLENT)</td>
</tr>
<tr>
<td>ACCEPTABILITY:</td>
<td>:ACCEPTABLE</td>
</tr>
</tbody>
</table>

The Pump agent and the Motor agent have retrieved satisfactory design templates and the Vbelt agent needs to attempt a retrieval of a Vbelt template. The power parameter from PUMP “pp\_pump\_4” and the drive speed parameter from MOTOR “mp\_motor\_4” are the shared parameters for the Vbelt-agent that are used to determine the retrieval of VBELT-TYPE-TEMPLATE 3VX.
that in turn is used by the Vbelt agent to design VBELT “vbelt_1”.

<table>
<thead>
<tr>
<th>VBELT vbelt_1 an object of class VBELT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BELT-TYPE: 3VX</td>
</tr>
<tr>
<td>BELT-LENGTH: 25.0</td>
</tr>
<tr>
<td>DRIVE-DIAMETER: 2.5</td>
</tr>
<tr>
<td>LOAD-DIAMETER: 2.2</td>
</tr>
<tr>
<td>NUMBER-OF-BELTS: 1</td>
</tr>
<tr>
<td>REQUIRED-POWER: 0.1</td>
</tr>
<tr>
<td>MOTOR-DRIVE-SPEED: 2400</td>
</tr>
<tr>
<td>LOAD-SPEED: 2727.2727</td>
</tr>
<tr>
<td>BELT-FORCE: 7.3234771990515d0</td>
</tr>
<tr>
<td>LOAD-PULLEY-WEIGHT: 1.0621228</td>
</tr>
<tr>
<td>BELT-LIFE: 334905.70545380004d0</td>
</tr>
<tr>
<td>COST: 34.864</td>
</tr>
<tr>
<td>WEIGHT: 2.3162665</td>
</tr>
<tr>
<td>EVALUATION: (:VBELT-AGENT :EXCELLENT)</td>
</tr>
<tr>
<td>ACCEPTABILITY: :ACCEPTABLE</td>
</tr>
</tbody>
</table>

This completes the design consisting of PUMP “pp_pump_4”, MOTOR “mp_motor_4” and VBELT “vbelt_1”.

4 Experimental Results

In order to get a better insight into the proposed negotiated retrieval mechanism for distributed casebases, we built an abstract version of CBR-TEAM, called the ABSTRACT-TEAM. It captures the essential aspects of CBR-TEAM, while stripping it of domain-specific complexities. The rest of this section will discuss ABSTRACT-TEAM in some detail and follows it with some experimental results demonstrating the benefits of negotiated retrieval.

Agents in ABSTRACT-TEAM have to assemble a mutually consistent composite case from local cases retrieved from local casebases. Each agent possesses constraints which may be defined on both local and non-local features. Each agent has a local casebase which is a list of feature vectors. At this point of our work, the control for negotiated retrieval was intentionally kept simple to avoid obfuscating some of the more important issues that we wanted to gain insight into. During Phase I, each of the agents retrieves best local case in parallel, based on local constraints derived from problem specification and initiates a seed partial case. In this phase, each agent constrains its local retrieval to avoid cases which violate constraints on local features. During Phase II, an agent informs the other agents about the existence of a new partial case. During Phase III, each agent tries to merge its local partial case with partial cases from the other agents. Each agent obtains the relevant projections from other agents and checks the feature values against its local constraint set for violations. If a violation is detected, the corresponding constraint is conveyed to the relevant agents and is immediately assimilated. A new iteration of negotiated retrieval is then initiated, but this time with an enhanced view of the problem-solving requirements for at least one agent.

For the experiments reported below, we used the example discussed in Section 2. A case of $A_1$ consisted of three features, and a local casebase was built by randomly generating feature vectors and their corresponding costs. Cases for $A_2$ had four features and a local casebase was similarly generated. Constraints that could be communicated to other agents had a representation similar to that in CBR-TEAM; numeric-valued constraints on single features. At present $A_1$ has a single constraint that it could communicate to $A_2$ and $A_2$ has two constraints that it can communicate to
\( \mathcal{A}_1 \). ABSTRACT-TEAM in the experiments below had only hard constraints and hence did not deal with optimizing any attributes on a case. Composite case had seven features corresponding to three from \( \mathcal{A}_1 \) and four from \( \mathcal{A}_2 \). Problem specification consists of required ranges on certain features\(^{11} \).

Figure 2 shows the time taken by agents executing negotiated retrieval to obtain a mutually acceptable best case\(^{12} \). Thirty runs at each casebase size are plotted. The casebases are initialized to be different for each run. The mean time shows an almost linear rise with the casebase size. The rectangular boxes show interquartile ranges — the top and bottom of a rectangle shows the 75th and 25th quantiles. The median or the 50th quantile is shown as a line in the body of the box.

![Figure 2: Negotiated Retrieval Time versus Casebase Size](image)

We next compared the negotiated retrieval against a simpler algorithm which involves retrieval and conflict detection but no explicit feedback through negotiation and exchange of constraints. Thus, detection of conflict simply leads to another round of retrieval. We again conducted thirty runs at each casebase size, noting the time difference between the simple retrieval algorithm and the negotiated retrieval algorithm for each casebase. Figure 3 shows that the difference rises sharply with the casebase size with the simple retrieval algorithm taking increasingly larger times compared to negotiated retrieval\(^{13} \).

\(^{11}\)Note that our description of ABSTRACT-TEAM does not highlight the fact that in a more general system some of the constraints are emergent - some of the requirements on the solutions emerge as the problem solving progresses. In addition, the constraints exchanged could be context-dependent meaning that they are applicable only in certain contexts which are known only during the problem solving and not apriori. Thus simply exchanging all the constraints at the beginning of problem solving is not a feasible method for the problem solving process described here.

\(^{12}\)The exact units of time are unimportant here. We used (get-internal-run-time) provided by Harlequin’s Lispworks development environment.

\(^{13}\)CBR-TEAM has certain constraints that can be communicated and certain others that cannot be, due to its limited form of representation on communicable constraints. If the detected conflict is due to a non-
5 Related Work

Much of the CBR literature is concerned with casebases of a single agent. There has been some work in breaking a single case into pieces and reasoning with these pieces in a single agent context. MEDIATOR[15] represents large cases monolithically with pieces embedded as parts which can be located within a larger case. A case is indexed both by its own indices and those of its pieces. This sort of representation is different from the the distributed casebase scenario we discussed in the paper. In a multi-agent system, there may be no monolithic case until the subcases are retrieved and assembled.

Barletta and Mark[2] break cases into pieces where each piece is a sequence of actions used to recover from a hypothesized fault. JULIANA[25] and CELIA[23] implement a distributed case representation that is closer in spirit to the work described here. Cases are broken into pieces called snippets, each of which represents the pursual of a goal. These snippets are linked to other causally related snippets. Following the links can reconstruct a full case. Each snippet has a pointer to the case header, its goal and the context in which it is embedded. During problem solving, an individual snippet can be retrieved and used to pursue a subgoal related to its goal. The next subgoal can be pursued using a snippet from an entirely different case. In all these systems, both the global context and the internal context of a piece in a case are carefully extracted and used as indices for that piece to avoid inconsistencies among pieces participating in a problem solving run. However, in multi-agent systems the agent casebases may be developed in disparate situations making it impractical to follow this strategy. So negotiated retrieval tries to deal with incompatibilities between pieces by detecting and resolving them at the retrieval time.

Rissland et al[24] discuss a CBR system called FRANK for report generation. Based on user’s communicable constraint, then the corresponding agent can only reject a composite solution without being able to give any feedback, just as in the simple retrieval algorithm above. However, if the conflict detected is communicable, then the behavior is similar to negotiated retrieval. Thus, the overall performance of the system will lie somewhere in between the spectrum whose ends are represented by negotiated retrieval and the simple retrieval algorithm.
preferences and requirements, a report type is selected. Each report type is associated with groups of strategies which act as indices into a library of plans. A plan is selected and instantiated to generate a report. Among other things, FRANK includes various types of responses to a failure to retrieve an adequate set of plans. It exploits information about failures to select better alternatives. The system tries alternative plans under the present group of strategies for report generation. Failing that, the systems tries alternative grouping of strategies. If this leads to a failure, the system changes the report type that it is trying to generate. This sort of broadening of allowed strategies upon failure is similar to the broadening of retrieval in our algorithm.

The negotiated retrieval algorithm benefits from a long history of thought in DAI and Multi-agent systems. Lesser[18] discusses the importance of interactions among subgoals and data distributed across a set of agents in distributed problem solving. Selectively sharing relevant partial results of a local search or data can enhance the global problem solving quality and efficiency without paying a large price for communication. In some recent studies, Decker and Lesser[12] further quantify these observations and discuss various trade-offs involved in communication vs enhanced problem solving efficiency. However, much of the past work in DAI has primarily focussed on exploiting interactions among subgoals to resolve control uncertainty. This paper deals with explicit detection of inconsistencies among the local solutions and subsequent focussed resolution of these inconsistencies[7]. In negotiated retrieval, feedback on the causes of conflicts detected leads to an improved retrieval in subsequent rounds where these conflicts are avoided. An agent which assimilates feedback from other agents enhances its view of global problem solving requirements leading to an improved retrieval process.

Sycara’s work on PERSUADER[26] deals with a negotiation model for multi-agent compromise where a mediator agent tries to persuade parties (labor and management) in conflict. The model generates proposals and counterproposals seeking to reduce the differences between the interests of the parties involved by incrementally modifying their beliefs to converge on common interests. PERSUADER draws upon a variety of techniques like case-based reasoning and multi-attribute utility theory to construct compromise solutions through incremental modifications to solution parts rather than by a composition of partial solutions.

6 Conclusion & Future Work

This paper makes an initial foray into methods for performing retrieval of cases distributed across multiple agents. Constraints emerge dynamically, as a result of the on-going problem solving activity. Negotiated retrieval performs focused exchange of information on these constraints to achieve a more coordinated distributed case retrieval. Our future work includes deploying and testing the system on real domains like networked information retrieval.

7 Acknowledgments

The first author would like to thank Qiegang Long for providing the initial encouragement to work on the ideas presented here and also for providing constructive criticism at various stages of the work. The authors would like to thank K. Decker, D. Nieman, Tim Oates, Z. Rubinstien and D. Skalak for reading draft versions of the paper and providing useful inputs.
References


