University of Massachusetts Amherst 1863

Rensselaer Polytechnic Institute 1824

University of Southern California 1880

The Collective Mind Initiative

DARPA

IPTO
INFORMATION PROCESSING TECHNOLOGY OFFICE

The Collective Mind Initiative

The Final Report

By

Norman K. Sondheimer
University of Massachusetts Amherst
Amherst, MA 01003
(413) 545-5654
Sondheimer@cs.umass.edu

William A. Wallace
Rensselaer Polytechnic Institute
Troy, NY 12180
(518) 276-6854
wallaw@rpi.edu

Peter M. Will
Information Sciences Institute
University of Southern California
Marina del Rey, CA 90292
(310) 822-1511
will@isi.edu

Table of Contents

1. Introduction

Military missions are planned for tight coordination of personnel and machines. Equipment failures during a mission imperil success and the lives of the war fighters involved. Our vision is to remove equipment reliability risks from mission planning and execution. We propose to address this challenge with technology supporting the pre-mission, in-mission and post-mission phases of operations.

Technically, we are aiming to achieve the goal of increased mission success rates by forming peer communities of equipment, e.g., the Marines' Light Armored Vehicles involved in Operation Iraqi Freedom might have formed a peer group. We will then use the collective experience of the peer communities to improve availability and reliability.

- Pre-Mission: technology to support decisions on the choice of equipment out of peer communities for availability and reliability using data on analogous missions and theaters.
- In Mission: technology to operate equipment optimally by sharing performance across the members of the peer community assigned to a mission.
- Post-Mission: Learning throughout equipment fleets from failures, remedial actions and successful reliability practices achieved in missions in order to prepare for future missions.

Consider these three scenarios:

1. The Movement from Kuwait to Baghdad: A Marine Expeditionary Brigade must quickly cover large stretches of ground while maintaining a synchronized battle line. In accomplishing both goals simultaneously, its light armored vehicles; amphibious assault vehicles, tanks, howitzers and NBC Fox vehicles are placed at risk of mechanical failure. Fielded Equipment mobility risks are well known. The key Light Armored Vehicle (LAV) mobility risks are known to be differentials, batteries and the planetary gear sets in each of its wheel hubs. For example, the planetary gear sets can overheat, restrict oil flow and destroy themselves. Knowing these limits, the Brigade Commander must integrate the mission objectives and what the drivers are reporting on their vehicle's health, in order to execute a plan for the best possible results. Technology needs to be developed to help the Commander in decision making by providing better information on which to base decisions.

All mobility risks are causally connected to the external environment (e.g., road or off-road surface), demands placed on vehicles (e.g., hours of continuous movement), their configuration (e.g., type of tires), and maintenance history (e.g., most recent oil change, failure history, repair history). Using the network communications becoming available, technology is needed to allow all Brigade vehicles to share their response, identify risks and "best practices". It will need to be able to advise the drivers on the management of the vehicles and the commanders on the management of the Brigade - all automatically. For example, if continuous speed on a section of road is seen dangerously raising the temperature of leading LAV's planetary gear sets, the technology will identify the problem and produce the recommendation that this section of road must be traveled more carefully by trailing vehicles.

2. A Raid on Fallujah: Small force action requires tight coordination of combined arms. Air cover, ground systems and foot soldiers must all function for the operations to succeed. During house-to-house searches for example, soldiers stop every 15 minutes to check status. Their equipment should be doing the same. Technology is needed that will identify systemic problems early. Weapon misfires will be noted and watched more carefully as the mission proceeds. Similarly, problematic radio transmission, overheating of vehicles and other electro-mechanical problems will be checked throughout the mission. If these problems are seen to reach the equivalent of the human physician's "numbers to treat", adjustment to the overall mission's plan will be recommended. Improved self-awareness will increase information superiority.

3. After-Action Lessons Learned: The single Marine Expeditionary Brigade described above would be only one of many moving through Iraqi. Similarly, the weapons systems involved will have been stressed in exercises, as well as regular training. The raid on Fallujah will have been one of many raids conducted. Technology needs to be developed that would collect this aggregate and ensemble data, factor out the differences in environments, generalize the lessons learned and continuously improve health management systems. Each use of equipment will be treated as an experiment whose results are available to the entire Service.

The challenges above are "DARPA-hard". The online decision-making stresses the availability of information and the requirements to decide with partial information. As opposed to laboratory experiments where variability is controlled, each piece of equipment is unique. The environment, demands placed on it, configuration and maintenance will vary in a potentially large number of important ways. To learn and reason at the speed needed to be useful on the battlefield will require significant advances in the methods available today.

1.1 The Collective Mind Initiative

Collective Mind for Equipment readiness was proposed to DARPA as a subject in the overall AI mission of the then new IPTO Office. The Principal Investigator was Norman Sondheimer of University of Massachusetts Amherst with William Wallace of Rensselaer Polytechnic Institute and Peter Will of University of Southern California Information Sciences Institute as co-PIs. DARPA funded the work at the level of a study. Peter Will's portion of the work was funded through AFRL as the contract agent.

The tasks were to solicit ideas from the best University, Industry and Military people for their ideas on the Collective Mind problem, to generate support for the idea from the Military and report their support to DARPA.

The overall goal of the Collective Mind Initiative is to show that

- Improved Equipment[1] Performance,
- Weapons Effectiveness, and
- Mission Critical Readiness

---

[1] We use "equipment" to denote any military platform

can come from amassing and sharing collective knowledge derived from the community of equipment via on-board information sharing that embodies the functions and utility of agents. The knowledge found from a fleet of equipment is to be used to improve the overall performance of the fleet, each single piece of equipment in the fleet and all equipment with similar components.

Finding solutions to the challenges in the Introduction is at present, human intensive. Our goal is to make it automatic by the development of new Artificial Intelligence Information Intensive techniques that fit the new DOD concepts of Net Centric Warfare and apply over all of the Services.

We proposed to address the opportunity by the development of an innovative technology: the Collective Mind with Collective Learning, Collective Reasoning, Collective Behavior and Collective Improvisation.

The approaches we propose to address these challenges of Collective Mind capitalize upon the existing data in the form of designed engineering models and existing field data; exploit the structure of platforms (equipment) as a network of interacting subsystems; exploit the heterogeneous experience of the various platforms (equipment), and be able to reconfigure to undertake unique missions (See Section 2.3 to 2.6). The knowledge produced as a result of these efforts will then be used to improve maintenance procedures in general, provide focused help to the individual maintainer and will be integrated into sophisticated reasoning systems for planning and scheduling – and for determining mission modifications that could be proposed to improve mission operations.

1.2 Outline of the Report

Following the Introduction, we describe how we involved academic and industrial researchers and practitioners, and gathered their thoughts on both the opportunities and challenges faced in developing the technology needed to implement the Collective Mind for the services.

Out of these discussions came the suggestion that a proof of concept experiment be conducted using an existing corpus of data on equipment reliability. The equipment selected for analysis was locomotives manufactured by General Electric (GE) and the analysis was performed by GE Global Research. An overview of this work is provided in Section 3 and the complete report is included in Appendix B. The report concludes with a list of military “customers” who were contacted during the study and a discussion of the potential impact of Collective Mind.

2. Research Agenda

2.1 Introduction

We ran several workshops on the Collective Mind as well as holding many smaller meetings. The first workshop was attended by potential users, military and industrial, and the remaining workshops were attended mainly by researchers from Academia and Industry [see Appendix A for locations and list of attendees].

Our definition of Collective Mind has four components: learning, reasoning, behavior and improvisation. Collective Learning is the process of understanding our dynamic environment by pooling data from multiple sources. Collective Reasoning is the transformation of the Collective Learning into codified knowledge, i.e., models. Collective Behavior is global behavior that emerges from locally interacting individuals in the collective. Collective Improvisation is the reworking of knowledge to meet unique mission requirements.

The application focus of our work on Collective Mind is on Mission Critical Readiness of Military Equipment. The details of that domain motivate the working definitions of Collective Mind and serve to give metrics on its efficacy.

2.2 Background

We postulate a scenario in which a mission is proposed for the asset; assume that the asset has not been asked to undertake the mission in the past, or if it has, the proposed mission may be in a new environment.

The technical challenge is to improve the mission readiness by ensuring that each and every asset (component, equipment, platform, etc.) employed in the mission is capable of performing the operations specified to achieve the objectives of the mission. This involves not only the selection of assets, but also their preparation for the mission and even their maintenance during operations. The first step is to search and see if there are any assets configured like that needed for the mission that have performed similar mission in a similar environment – and have a similar history. Here the collective is the set of all assets that are the same as the set needed for the mission. We therefore need for every asset a "model" that predicts its performance based upon the condition of its equipment, its past performance and the conditions expected during the ensuing mission. More specifically, we need the vector of performance variables that are some function of the attributes that represent the condition of the asset – perhaps in terms of its components. The condition of the equipment is in turn a function of its operation in the past (the missions it has performed including the environment it has performed in) and its maintenance history.

The Collective Mind for this case is the knowledge from the collective set of all assets as well as the ability to learn from the collective as these assets continue to operate in the missions.

If we consider each asset (component, equipment, platform…) as an agent with its own reasoning and learning capability, the Collective Mind is a system of interrelated actions by these agents. The actions are purposeful and the agents are attentive to the actions of other agents. The agents construct their actions, understand that the system consists of connected action by themselves and others, and interrelate their actions within the system. In order to accomplish the foregoing, actions by the agents must be contributions to the goals of the system; there must be a common representation for each agent to understand the actions of others and the results of those actions; and the system must recognize the need for an agent to subordinate its actions to those of the system. In this conceptualization, the actions are really the mental processes of the Collective Mind. The Collective Mind is in how the agents contribute and represent all actions, and produce improved group behavior.

A question is: … does the collective achieve the desired behavior? Formally,

a) Let $f_i(x_i)$ be the function (or task) to be optimized by an individual agent A_i, where $i=0,\ldots,N$ and N is the number of individuals in the collective, and x_i is the parameter of A_i.
b) Let $G(f_1(x_1), f_2(x_2),\ldots,f_i(x_i), \ldots,f_N(x_N))$ be the function that must be optimized by the entire collective.

We are interested in the characteristics of G. Furthermore, when G is given, can the system automatically determine $f_i(x_i)$ for the individuals? In the simplest case, when G is an additive (sum) function, then every individual should simply maximize $f_i(x_i)$ so that G will be maximized. However, in real world applications, G can be much more complex, and some individuals must "sacrifice" themselves (i.e., minimize their own $f_i(x_i)$) in order to maximize the value of G in the collective situations. Ideally, given a new application G, the collective should be able to automatically generate $f_i(x_i)$ for each individual. In an ant colony, the may be built in the ant's body, or may be it is dynamically adjusted depending on the task they perform. For example, when ants must build a bridge to surmount a chasm, some ants will "sacrifice" themselves to make the bridge while others will move over it. Another example is the voting game, where individuals vote "yes" and "no", but their reward depends on the percentage of yes votes of the people of whole population. This global reward function is collected by a "Referee" and is not known by the individuals (Tung & Kleinrock 1993). Similarly, if we make certain assumptions on the relationship between individuals (such as they act as constraints), then there are some studies that can perform distributed optimization when G is given and fixed (Modi, et al. 2003).

2.3 Collective Learning

Machine learning is traditionally for single agents. Recently, there is the new trend of multi-agent learning. However, the topic of Collective Mind has several unique features that demand a new paradigm for machine learning that we call Collective Learning. The new learning problems include:

- How do individuals learn the structure (some call it topology) of the organization dynamically? Existing approaches such as hidden Markov models or Bayesian Networks are mostly about learning parameters and they avoid this structural learning problem because it is too hard.

- How do individual agents learn a model of the environment and the same time a model of other individuals? Traditionally, these two modeling activities are fixed together, and most subsume the other. In the Collective Mind, these two may be related, but they definitely have different characteristics and require different learning techniques (Shen 1994).

The Collective Mind envisions integrating domain knowledge from many sources with real-time data feeds from deployed platforms to support rapid problems identification and response. Sources of domain knowledge include the following;

- The "anatomy" of each platform (What are the components and subsystems? How are they physically located? How are they connected?);
- The "physiology" of each platform (How does each subcomponent of the platform contribute to the overall functioning of the platform? Typically, this is divided into separate models for each subsystem.);
- The maintenance history of each component (When manufactured. History of maintenance actions. Results of previous tests.);
- The deployment history of each platform (What missions has it participated in? Where? Under what environmental condition? How long mothballed? Where?).

The real-time data feeds include the following:

- On-board sensors on each vehicle;
- Maintenance events;
- Debrief from crew after each mission (or each shift?).

Collective Learning involves a distributed set of learning platforms that must learn continually but that only occasionally have opportunities to communicate with each other. Under such conditions, it is not feasible to pool all of the sensor readings from all of the platforms in real-time. Instead, each platform must form its own hypotheses and then, when the opportunity arises, communicate its hypotheses to the other platforms (along with the key supporting data and observations).

Existing learning methods do not have good ways of making use of domain knowledge. Existing learning methods are designed for off-line batch training (e.g., constructing an optical character recognizer by training on a database of 1 million labeled hand-written characters). Existing learning methods are designed for learning from a single combined database, rather than by combining hypotheses from many other learning agents.

Making Learning Knowledge-Guided: Domain knowledge can guide learning in two ways. First, it can suggest the space of hypotheses to consider [Dzerosi & Lavrac 1994]. For example, a learning system that only had sensor readings must learn to relate overall platform failure directly to the history of sensor readings. It might explain a platform failure in terms of the accumulation of several episodes of operation in high ambient temperatures. But a knowledge-based learning system could explain the platform failure in terms of failure of the engine caused by the added load placed on the engine by the air conditioning system resulting from the crew needing more air conditioning to operate successfully in high ambient temperatures. A knowledge-based learner can relate sensor readings to individual subsystems and then explain overall platform failure in terms of the failure of certain subsystems.

Second, domain knowledge can constrain the space of possible explanations [Clark & Matwin 1993]. If we consider the space of all mappings from raw sensor readings to platform failures, this is an immense space. Rapid learning from small amounts of data requires that the space of mappings be highly constrained. Domain knowledge can constrain the space by recasting it in terms of components and subsystems rather than just raw sensor readings. It can also suggest the direction of possible effects. For example, operating an engine at higher RPMs tends to reduce engine life; operating an engine at extreme temperatures tends to reduce engine life, etc. Without this kind of background knowledge, a learning system would need to consider (and reject) the hypothesis that lower RPMs and normal temperatures reduce engine life!

Existing research in Inductive Logic Programming and Probabilistic Relational Models shows how to use domain knowledge to define the space of possible hypotheses. However, these methods have not been scaled up to large problems or to problems involving very noisy, sensor-based data.

There is only a small amount of research showing how domain knowledge can constrain the space of possible hypotheses considered by the learning system. This research is largely *ad hoc*. Substantial work is needed to develop good modeling languages for describing the domain knowledge and good ways of converting the domain knowledge into constraints on the hypothesis space.

Making Learning Real-Time: There are two challenges to creating real-time learning systems. The first challenge is to design online versions of existing learning algorithms. There is a lot of existing work on online (real-time) algorithms for training neural networks, linear threshold units, and decision trees [Rosenblatt 1958; Littlestone 1987; Utgoff 1989]. Most batch search and optimization algorithms can be converted into online algorithms in principle. The challenge is to find practical, efficient online versions of these methods.

The second challenge is to make those online algorithms adaptive, by which we mean that they can deal with changing worlds in which new kinds of failures occur, new kinds of sensors become available, old sensors cease to be available, and the probabilities of different faults change because of changes in missions and the ways that platforms are being used in missions.

Existing research in expert-weighting and portfolio algorithms have been shown (theoretically) to adapt rapidly to changes in the phenomena being predicted [Cesa-Bianchi, et al. 1993; Vovk & Watkins 1998]. A DARPA program could transition this work into real-world systems and show how it can be applied in noisy real-time settings.

Making Learning Collective: The challenge for Collective Learning is for multiple learning agents to pool their learned knowledge *without* pooling all of their sensor data. Existing research suggests the following directions to pursue.

- First, research on ensemble learning methods learns to take a weighted vote of the hypotheses learned by each individual learning agent [Kuncheva 2004]. This has been hypothesized, and has been shown to give results comparable to those obtained by training a single system on the entire data set [Chan & Stolfo 1996].

- Second, research in support vector machines (and related algorithms) shows how to identify the key data points that support a hypothesis [Cristianini & Shawe-Taylor 2000]. These points are known as the "support vectors", and they are sufficient to reconstruct the hypothesis perfectly. An interesting direction would be for the multiple agents to exchange their support vectors and then use all of these support vectors in learning.

2.4 Collective Reasoning (includes Planning and Scheduling)

In Collective Mind, since knowledge and information are distributed among many individuals, it makes reasoning and planning/scheduling much harder. One unique advantage this will offer is that damages to any individuals will not paralyze the entire organization. The individuals should know where to ask for and deliver information, know how to recover information when some nodes died, know what to communicate among themselves in order to make a good plan, and know how to evaluate a new plan/schedule collectively.

Another aspect of Collective Reasoning is how to divide a global task into a "workflow" of smaller tasks so that each small task can be performed by some individuals, and when those small tasks are finished, the results should be assembled in such a way that a global solution can be readily obtained. This is the divide and conquer problem and typically the "Task Allocation" problem.

An important opportunity for research is to integrate learned knowledge into sophisticated reasoning systems. The Collective Mind requires this, because the results of individual component and vehicle prognoses must be used by the mission + maintenance scheduler to decide when and how to schedule platforms for missions and for maintenance. Uncertainty in prognoses must be translated into uncertainty in mission success and/or the need for maintenance.

There are at least two approaches to incorporating uncertain predictions into complex reasoning: propagation of uncertainty and ensembles; propagation of uncertainty computes a posterior distribution over random variables of interest (e.g., mission success, expected equipment losses) based on the distributions of other random variables (i.e., the prognostic predictions); ensemble methods construct a set of non-stochastic "alternative scenarios" or alternative models and compute schedules based on each scenario. The resulting schedules are then analyzed to identify consensus scheduling decisions and/or ways of modifying the schedule so that it will succeed under all scenarios. Ensemble methods have been very popular in classification learning, but there has been little research on ensembles for reasoning.

Another research opportunity is to determine and propose modifications to missions both in planning and scheduling of operations. The space of potential mission modifications is huge, so some way is needed to constrain the Collective Mind from proposing ridiculous modifications.

This can be viewed as the problem of reasoning about "utility function" of the commander. We can imagine that several tradeoffs are operating in a battle theater: (a) mission goals, (b) safety of troops, (c) reliability of supply, (d) loss of equipment and (e) speed. There has been recent work on inferring multi-attribute utility functions by observing the choices made by humans, (e.g., in auctions or in video games). We would also like to develop systems that are instructable, so that commanding officer can say "Timing is critical; don't propose any modifications that delay the mission." This learning should begin during exercises and continue into the battlefield. Once the utility function has been acquired, the scheduler can generate and evaluate alternative mission modifications and choose the modifications and choose the modification that has the highest utility.

2.5 Collective Behavior

A collective should yield an emergent whole that is qualitatively more than the sum of the parts. It's functionality ought not to be something that one of the parts could do by itself if only it were bigger. Any such system faces a dual challenge.

- How does the behavior of the individual elements yield the emergent behavior of the whole? (Example from our case: how does local awareness of a platform's own state roll up into a global assessment if the state of readiness of the fleet?).
- How can the global functionality be applied to the problem, given that the individual elements are the only sensors and effectors that the system has? (Example: if the system learns that alternators with exposure to high temperature and high humidity have unusually high failure rates, how does that knowledge affect local decisions at the company platoon level?).

This issue is at the heart of the need for compositionality, which may be the critical issue. What makes compositionality difficult is that both individual behaviors of the equipment and their interactions are typically nonlinear. If one adopts a centralized approach, these can be addressed relatively simply, but such a solution does not scale well and is not robust against attack.

Potential ways for addressing these issues draw heavily on simulation and concepts from statistical mechanics as a body of knowledge about how global properties emerge from locally interacting entities. These approaches are also relevant when we assume that the computational and communications environment may be constrained – just as in the field.

The major challenge is to determine how the required learning and reasoning functions can be achieved under such constraints, providing graceful degradation (rather than catastrophic failure) as communications and computational power are incrementally degraded. Conventional learning and reasoning algorithms do not decompose neatly onto such an architecture, or at least they have not been shown to be decomposable this way. Swarming approaches, by contrast are ideally suited to such architectures because of three of their characteristics:

- The individual processes are small compared with the overall system, so they can easily run on cycles scavenged from other applications on embedded processors.

- Each process interacts only with others that are co-located with it in some topology. The best fit comes when this topology is isomorphic with the physical distribution of the platforms, but even if it is not, it does provide a way to limit the interactions among processes and thus function in an environment with bounded communications.
- Their emergent dynamics are robust to incremental changes. They tend to degrade gracefully over wide parameter ranges. (There are, naturally, limits beyond which they cannot function, and characterizing these is critical to a program in this area, but they offer a much broader range of operability than do conventional mechanisms.)

<u>Integration of Monitoring/Diagnosis with Scheduling and Planning</u>: The modular vision of "first we assess the state of our platforms, then we plan the mission" is unrealistic in a highly dynamic environment in which platform state and mission constraints change constantly. We need new mechanisms that can incrementally learn and plan in tandem. Such mechanisms must have the "any-time" characteristic: they quickly produce an approximate answer, and can give more detail if more time and resources are available to them. In the context of closely coupled learning and planning, often an early approximation to one half of the problem (say, learning) can help constrain the space that the planning function must search, and the planner's early results can in turn help focus the learner, leading to more rapid convergence than would be possible in a sequential system.

Swarming algorithms are typically any-time, and have been demonstrated in both classification and planning tasks.

<u>Environmental Integration</u>: The issue here is discriminating between two possible interpretations of an aberrant sensor reading: system malfunction (the system is out of spec) vs. environmental or historical stress (the environment is out of spec). No matter how complete our models of our systems may be, complex electromechanical systems will always have emergent properties that surprise us. We need ways to use other platforms that share the same situation as an implicit engineering model to distinguish (local) equipment failure from (shared) environmental stress. More generally, we need to compare behaviors across platforms that may not be co situated right now, but that are near to one another in the space of shared histories.

This challenge relies directly on the notion of proximity among platforms in some topology (physical space-time; history space), and so lends itself naturally to swarming methods, one of whose hallmarks is locality of interaction.

2.6 Collective Improvisation

"Making do" and "taking the initiative" are desirable actions of military personnel, tasked with supporting battlefield operations.

Improvisation involves reworking knowledge in time to meet the requirements of a given situation. Reworking refers to revising or abandoning planned-for procedures. Time is central to improvisation since the improviser's decision cannot be undone once it is done. Finally, meeting requirements means accounting for constraints in the decision setting while acting to meet the goals of the response. The question of *when* to improvise involves recognizing when planned-for procedures cannot or should not be applied. In problem-solving terms, it may therefore be conceptualized as a categorization problem, in which the ability of likelihood of a decision maker to categorize correctly is influenced by a number of factors, such as penalties associated with making an incorrect choice. The question of *how* to improvise involves developing and deploying new procedures in real-time. It may be conceptualized as a search and assembly problem, influenced by factors such as time available for planning, risk in the environment and the results of prior decisions.

Collective Improvisation is an approach to supporting battlefield personnel when the need to develop and deploy new procedures arises. In Collective Improvisation, past knowledge – which may be contained in databases such as ontologies and may be operated upon by decision logics – is re-examined and reorganized in order to meet new requirements. Results of these improvisations are then fed back into the system, thereby completing the learning loop.

Related prior work lays the foundation for collective improvisation. Hayes-Roth and colleagues have developed a series of blackboard-style architectures to support and in some cases model improvisation. Models built upon these architectures are enabled with dynamic control, thereby allowing execution of real-time control plans which specify a sequence of tasks, parameter values and constraints. In order to support improvisation and capture the learning involved, the blackboard or any architecture must have access to and understanding of models of the physical systems of the platforms involved in battlefield operations.

3. Collective Mind Experiment

3.1 Background

The Collective Mind concept was studied, as described, by several workshop groups comprised of experts from both the Academic, Industry and Military communities. The technical opportunities excited the Academicians, Industry researchers, and the practical opportunities excited the Military. It became clear that a proof of concept experiment, if it could be performed within the Study, would serve to crystallize the thinking of all communities and show immediate tangible results. This section presents an overview of the experiment. Appendix B contains the final report.

A good experiment necessarily requires a good corpus of experimental data. While military data exists, the best and most accessible corpus of data was found in GE Service Transportation Rail Operation. GE's locomotives are sophisticated electromechanical machines that contain big diesel engines, have to operate in all climates, all weathers and terrains and move heavy equipment. They can be considered therefore as reasonable surrogates to military vehicles.

We selected Improving Mission Reliability, a component of Mission Operations, as the objective for the "Proof of Concept" experiment. Mission reliability was broadly defined as – given a mission of duration X-days, what percentage of units assigned to that mission are able to complete the mission without a critical failure. The motivation for high mission reliability in both commercial and Military environments is two-fold. First, it gets the mission performed; second, it makes mission planning and execution more predictable and effective; third, it reduces the logistics footprint required to support a certain level of readiness. In the military domain, this may mean picking the best five vehicles to conduct a reconnaissance mission in swampy terrain; in the commercial sector, it may imply selecting the best five locomotives to deliver time-critical shipments from coast to cost. This problem is accentuated in the case of new mission types or new equipment platforms when insufficient data exists on how the equipment will behave in that environment.

Our experiment involved an extensive corpus of field data, but many missions involve newly introduced equipment. The paradigm for new platforms focuses on the continuum and tension between engineering test cell projections made before deployment, and retrospective statistical measurement of performance found or measured after a substantial number of missions. During this 'gap' – the first wave of missions on new platforms, both the commander and the maintainer are selecting and repairing units 'blindly' with respect to their equipment's expected behavior. The Collective Mind approach tries to compensate for the scarcity of operational experience on any single unit by learning from ground performance of 'peer' units with current or past similar deployment experience.

3.2 The Experiments

The Study funded an experiment (in reality, a series of experiments) conducted by GE Global Research that applied peer-based learning to predict time-to-failure performance in locomotives. GE as part of their normal business keeps an extensive and perhaps unique data set of field failures and repair actions from customers' locomotives. The data was obtained from GE locomotives owned and operated by GE Rail and Union Pacific. The data is obtained from *normal* computer control systems used in the locomotive and delivered back to GE by a variety of means including a satellite dish on each train. That is no special sensors were installed for the normal course of business or for our experiments. The data for our experiment combined design, utilization and repair information on 1100 locomotives over a 2 year time period. Any individual locomotive's time-between-failures appears chaotic and unpredictable.

This project utilized existing *systems* on the locomotives. No new sensors were added to collect these data. This Collective Mind approach capitalized upon existing data collection methods and did not design *a priori* new sensors or other data collection systems.

The data for the study was collected from four different sources:

1) *Locomotive Design & Engineering Data from GE Rail:* GE Rail manufactured the locomotives in this study. As the OEM, GE Rail possessed engineering data on locomotive models, configurations, date of manufacture, date of service, the date service was installed, upgrades, and software modifications.

2) *Locomotive Recommendation Data from GE Rail EOA^{TM} remote monitoring and diagnostics service:* For each locomotive, there was a time-stamped record of when the Expert on Alert (EOA^{TM}) system detected abnormal patterns in the fault data leading to a recommendation being issued by GE Rail Locomotive Services. A red or yellow recommendation indicated a problem that was serious and required a fix in the next 7-10 days at most.

3) *Locomotive Maintenance Data from Repair Shops:* Each red or yellow recommendation used in the experiments was associated with maintenance feedback from railroads or GE repair shops which indicated the exact repair action that successfully fix the problem. Therefore the data included only maintenance intervals where a genuine problem existed on the locomotive that was verified by the maintenance personnel.

4) *Locomotive Utilization data from a selected railroad:* Each locomotive maintains an on-board record of a number of utilization-related parameters that are collected when a locomotive reaches a railroad yard. These parameters include odometer miles, total megawatt-hours, hours spent motoring, hours spent in dynamic braking, cumulative engine hours, cumulative engine hours moving, percentage of time spent in each of the eight notch settings (analogous to gear settings) and others.

5) *Diagnostics* are done in GE's Diagnostic center staffed by a group of extremely experienced engineers who have final decision power over the machine computed

recommendation produced via case-based reasoning. The diagnosis and repair recommendations are then sent to the local depot or to the train, wherever it is (known through on-board GPS).

3.3 The Collective Mind Computational Approach

Peer-based learning methodologies were investigated since they provide a transparent, adaptable model mechanism. The particular approach taken was to focus on the representation and reasoning mechanisms of instance-based reasoning. Instance-based reasoning (IBR) relies on a collection of previously experienced data that can be kept in their raw representation. IBR is an analogical approach to reasoning, since it relies on finding previous instances of *similar* problems and uses them to create an ensemble of local models. Hence the definition of similarity plays a critical role in the performance of IBR's and is a dynamic concept. The concept of similarity is not crisply defined, creating the need to allow for some degree of vagueness in its evaluation. This issue was addressed by evolving the design of a similarity function in conjunction with the design of the attribute space in which the similarity was evaluated. After developing several exploratory peer-based models, a fuzzy instance-based classifier was used that was designed by an evolutionary search (instead of by a manual process). Specifically the following steps were used:

1) *Retrieval* of similar instances from the Data Base;
2) *Evaluation of similarity measure* between the probe and the retrieval of instances;
3) *Creation* of *local models* using the most similar instances (weighted by their similarity measures); and
4) *Aggregation of outputs* of local mode to probe.

No additional sensors were used for this experiment. All data came from on-board sensors used by the control systems that regulate the various subsystems in the locomotive. This constraint implies that it will not be necessary to re-instrument existing or new platforms to re-apply a similar process and obtain comparable results.

The experimental results reveal that consulting a unit's peers, the collective, provides a significant increase in the ability to characterize the behavior of that unit in terms of completing the next mission. The peer-based approach is robust and degrades gracefully with the information loss that is likely to be present in the battlefield. *In addition, 'rules of thumb' such as using the newest units on a mission were actually shown to be damaging rather than beneficial.*

With this limited data, the use of Evolved Peers provided the best overall accuracy (60.35% = over 3 times better than random selection) for past performance. When the selection was limited to a small fixed number of units, Evolved Peers provided an accuracy of 63.5% (over 10 times better than random selection) for the past performance. Finally, Evolved Peers provided the best overall accuracy (55% = 2.7 times better than random selection and 1.5 times better than best heuristics) for future performance.

The collective (peer-based) approaches have shown great robustness to information loss. This will enable mission reliability for minimally instrumented platforms operating with limited bandwidth.

The experiment showed the applicability of peer-based learning methodologies with evolutionary algorithms to select the best attributes for representing peers and to define similarity measures for identifying the most similar peers for a given unit. By evolving the models over different time slices, it has been shown the ability to dynamically adapt the neighborhoods of peers using incremental operational and maintenance data. In future work, structural design of the attribute space (for the definition of peers) could be extended by using genetic programming in lieu of evolutionary algorithms, and attribute selection and weighting to attribute construction. The fitness function to tradeoff classifier accuracy and confidence could be improved by adding measure of representation parsimony and find Pareto fronts for different tradeoffs.

Generating more sophisticated local models for predictions could also extend the approach. The present assumption was that each peer had a rather “feeble” track-history, which motivated the peer approach to begin with. In situations where the peers have a richer track-history, more complex models, whose parameters could be obtained using a local search method, could be developed. In addition to the aforementioned technical extensions, one or more experiments should be conducted using data describing equipment usage more typical of military operations. For example, the data should include instances of irregular use of equipment, equipment use in diverse environments in a variety of missions, equipment operating conditions that range from none to little usage, normal operations and stressed and overload.

3.4 Conclusions

The GE experiment showed significant improvement in fleet performance, so much so that GE has already tested some of the ideas in commercial practice. The military representatives at the meeting where the results were reported gave the work a very good report saying that the technology far exceeded any technology used in the military today. Since this experiment has been completed, it has been brought up many times in Military maintenance circles.

4. Technology Transfer to the Military

A key imperative from DARPA was to elicit and enlist support from a real military customer.

We made visits/presentations to the following organizations, in each case explaining our concept, listening to their feedback and evolving the concept to make it more suitable for technology transfer

Department of Defense Condition Based Maintenance + (OSD CBM+)
Department of Defense Office of Force Transformation (OSD OFT)
US Air Force Expeditionary Logistics for the 21st Century (eLog21)
US Air Force Knowledge Services (AFKS)
US Air Force Research Laboratory (AFRL)
US Army AMRDEC
US Army Future Combat Systems (FCS)
US Army Logistics Transformation Agency (LTA)
US Army Materiel System Analysis Activity (AMSAA)
US Army Objective Force
US Army Research Laboratory
US Army RDECOM SMS IPT
US Army TARDEC
Joint Strike Fighter Program Office
US Marine Corp Systems Command (MARCORSYSCOM)
US Navy Office of Naval Research

A public presentation was made to all services at the 2004 DoD Maintenance Symposium in Houston, Texas.

5. Potential Impact

The Collective Mind Workshops showed that the Collective Mind topic had Military relevance and interest.

The Collective Mind Workshops showed that the Collective Mind topic had Academic and Industrial Research interest.

The conclusion is that the objective of the DARPA study has been met; both military and research and development communities endorse the concept. A DARPA program could stimulate important research in Collective Mind.

Now is the time to challenge learning researchers to develop *robust* engineering methodologies for deploying learning systems. Machine learning has so far taken place under "laboratory conditions" where Ph.D. researchers hand-craft the systems to make them work. As a result, while machine learning provides a revolutionary new method for constructing intelligent systems (such as handwriting recognition, speech recognition, and artificially-intelligent simulated characters in games), the resulting systems do not learn after they are deployed. Military system must learn after deployment. This is the key to making all kinds of computer systems—from word processors to telephones to robots—adaptive to the needs and environments of their users. Without real-time learning, our hand-crafted intelligent systems will remain brittle and hard-to-use.

Machine learning is currently too slow. Most systems must be trained on thousands or millions of training trials. Introducing new equipment at a rapid pace, the core of Transformation demands fast learning. Learning can be much faster if domain knowledge is available to guide and constrain the process. Success in the Collective Mind project will produce a widely-applicable knowledge-guided learning technology.

Machine learning currently does not scale. Indeed, even human learning takes place only within the head of each individual person, and society spends billions of dollars to combine and communicate this learned knowledge to other people. The Collective Mind project envisions a learning technology that is able to rapidly combine knowledge learned separately by many distributed agents so that each agent can become a "super agent" that benefits from everything learned by the other agents. This might allow computers to learn very rapidly and identify patterns that people, with our limited ability to combine learned knowledge, cannot detect.

References Cited

Cesa-Bianchi N., Freund Y., Helmbold D.P, Haussler D., Schapire R.E., and Warmuth M.K., "How to Use Expert Advice", in *Proceedings, 25th Annual ACM Symposium on Theory of Computing*, pp.382-391, New York: Assoc. Computing Machines, 1993.

Chan P.K. and Stolfo S.J., "Sharing Learned Models among Remote Database Partitions by Local Meta-learning", *Proceedings of KDD-96*, 1996.

Clark, P. and Matwin, S., "Using Qualitative Models to Guide Inductive Learning", In: *Machine Learning - ECML'94, European Conference on Machine Learning, Catania 1994, Lecture Notes on Artificial Intelligence*, pp. 360-365, Springer Verlag, 1994.

Cristianini, N. and Shawe-Taylor, J., An Introduction to Support Vector Machines, Cambridge University Press; 2000.

Dzeroski, S. and Lavrac, N., "Inductive Learning in Deductive Databases", *IEEE Transactions on Knowledge and Data Engineering*, 1994, **5**(6), pp. 939-949.

Kuncheva, L.I., Combining Pattern Classifiers: Methods and Algorithms, Wiley & Sons; 2004.

Littlestone, N., "Learning Quickly When Irrelevant Attributes Abound: A New Linear Threshold Algorithm", *Machine Learning*, 1988, **2**(4), pp. 285-318.

Rosenblatt, F., "The Perceptron: a Probalistic Model for Information Storage and Organization in the Brain", *Psychology Revie*w, 1958, **65**, pp. 386-408.

Pragnesh Jay Modi, Wei-Min Shen, Milind Tambe. "A Dynamic Distributed Constraint Satisfaction Approach to Resource Allocation". *"Distributed Sensor Networks: A Multiagent Perspective"*, Kluwer Publishers, 2003.

Shen, W-M., Autonomous Learning from the Environment, W. H. Freeman, Computer Science Press, 1994. (Foreword by Herbert A. Simon)

Utgoff, P.E., "Incremental Induction of Decision Trees", *Machine Learning*, 1989, **4**, pp. 161-186.

Vovk, V.G. and Watkins C., "Universal Portfolio Selection", *In Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT-98*), ACM Press: New York, pp. 12-23, 1998.