A Distributed Resource Evolutionary Algorithm Machine (DREAM)

Ben Paechter
Napier University
Edinburgh
Scotland
benp@dcs.napier.ac.uk

Thomas Bäck
Informatik Centrum Dortmund
Dortmund
Germany
baeck@icd.de

Marc Schoenauer & Michele Sebag
Ecole Polytechnique
Palaiseau
France
marc@cmapx.polytechnique.fr

A. E. Eiben
Universiteit Leiden
Leiden
The Netherland
gusz@wi.leidenuniv.nl

J. J. Merelo
Universidad de Granada
Granada
Spain
jmerelo@geneura.ugr.es

T. C. Fogarty
South Bank University
London
England
fogarttc@sbu.ac.uk

Abstract- This paper describes a project funded by the European Commission[1] which seeks to provide the technology and software infrastructure necessary to support the next generation of evolving infohabitants in a way that makes that infrastructure universal, open and scalable. The Distributed Resource Evolutionary Algorithm Machine (DREAM) will use existing hardware infrastructure in a more efficient manner, by utilising otherwise unused CPU time. It will allow infohabitants[2] to co-operate, communicate, negotiate and trade; and emergent behaviour is expected to result. It is expected that there will be an emergent economy that results from the provision and use of CPU cycles by infohabitants and their owners. The DREAM infrastructure will be evaluated with new work on distributed data mining, distributed scheduling and the modelling of economic and social behaviour.

1 Introduction

It is hoped that Universal Information Ecosystems "will turn the complex information infrastructure as it is emerging today into a rich, adaptive, responsive and truly open environment" [1] For this to happen there will inevitably be some types of software infohabitant (sharing some of the properties of a chromosome or agent) which must adapt and evolve, for example, to model complex distributed systems, such as large groups of humans, or to help solve problems or search out data. The idea behind the Distributed Resource Evolutionary Algorithm Machine (DREAM) is to provide the technology and software infrastructure necessary to allow these infohabitants to exist and evolve in a virtual world.

It is out intention that the DREAM framework will go beyond the state-of-the-art by providing all of the following features within the same system:

- A tool that is a framework in which to develop instantiations of applications, rather than having the models or problems hard-coded into it.
- A tool designed to allow both the solutions of industrial optimisation problems and the modelling of the behaviour of large systems.
- A tool which allows for free migration of infohabitants through the internet, thus allowing the formation of diverse niches
- A tool which allows the use of spare CPU cycles in an automated and secure manner
- A tool designed to allow behaviour at the macro level to be observed (whereas, for example, *Creatures* only monitors its inhabitants at the micro level)
- A tool designed to be scaleable and open

There are many systems in existence which provide some of the features of the DREAM. Some provide libraries or toolkits for evolution. Some, for example *Swarm* [2], provide simulations using agents. Others, for example *Creatures* [3] provide artificial life simulations with migration over the Internet. Some systems, for example *Network Tierra* [4] also allow some sharing of CPU time. Further systems, for example Beowulf clusters, allow parallel processing using networks of machines. However, no systems currently in existence provide all the features we hope to provide. In particular, few current systems have been developed with scalability and openness as paramount concerns.

The DREAM project is based on two components, a physical one and a virtual one. The virtual world will necessarily be implemented on physical resources. The owners of those resources, or infohabitants working on behalf of end users, will from time to time have problems that need to be solved using great amounts of computing power. At other times they will have computing power that

[1] Funded as part of the European Commission Information Society Technologies Programme. The authors have sole responibility for this work, it does not represent the opinion of the European Community, and the European Community is not responible for any use that may be made of the data appearing herein.

[2] The European Commission IST Pro-Active Initiative on Universal Information Ecosystems, described in [1], defines an "infohabitant" by - "Individuals, organisations, as well as virtual entities acting on their behalf, smart appliances, etc. could be denoted as 'infohabitants' of a Universal Information Ecosystem".

0-7803-6375-2/00/$10.00 ©2000 IEEE.

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

is idle because the computer is not being used at that time or because the CPU is not being used 100%. The infrastructure proposed here allows the DREAM to be implemented across hardware resources that are shared and distributed. At a global level, this allows much more effective and efficient use of existing hardware. At the local level, individual users are enabled to use much more computer power than they physically possess without having to invest in hardware. An individual DREAM will be implemented across a number of computers and some computers may take part in more than one DREAM. For example, a researcher at a University may allow their desktop PC to take part in a DREAM run by their institution, and another run by researchers worldwide, see figure 1.

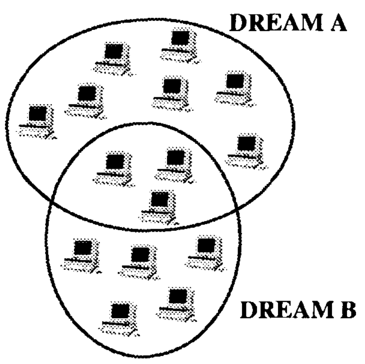

Figure 1 - The Mapping of DREAMs onto hardware.

On the virtual side any number of distinct experiments can be run within a particular DREAM. We can think of each experiment as a distinct "infoworld", see figure 2.

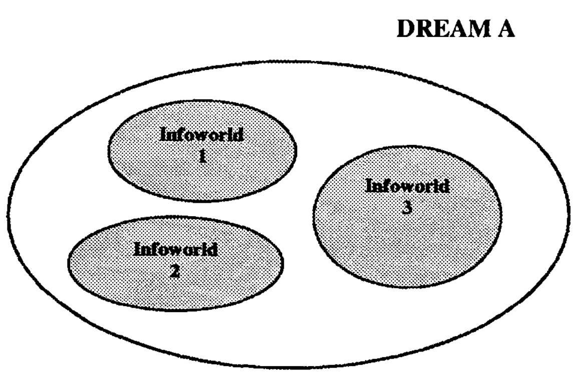

Figure 2 - The Mapping of Infoworlds onto a DREAM

Each infoworld might be used to either solve a problem or to simulate some situation. In the first case, an evolving population of infohabitants will be able to tackle a problem in an adaptive fashion. Individual infohabitants or sub-populations can compete, thereby enforcing quality pressures on each other. By using Darwinian principles in computer simulations, high quality problem solutions can be evolved. While this competition continues, there is also the possibility for co-operation or negotiation between individual infohabitants or sub-populations. This, it is expected, will provide a collective intelligence that operates effectively by dividing the problem at hand and allowing infohabitants to generate a solution jointly. This is especially important for real, very large problems. It is expected that quality solutions will be obtained for problems of a much larger scale than present technology supports.

In the case when it is viewed as a simulation tool, the DREAM virtual society might be used to simulate aspects of real world society. Example aspects that might be simulated are: the way in which interactions at a local level lead to emergent behaviour at the macro level; the way that geographical distribution leads to cliques and clusters that share the same information and ideas within themselves; or the ways in which economies arise from the assignment of resource value to information.

The DREAM framework provides a virtual environment that, by its intrinsic complexity and variety, matches real environments better than currently used systems. It is expected that non-trivial insights in the complex dynamics of society can be achieved.

The intention is to put forward the "microscopic simulation / macroscopic analysis" approach. The DREAM framework enables an enormous scale-up in the size of simulated words. The worlds we intend to simulate are derived (i.e. simplified) from real life. Yet, it is an open research question whether this scale-up reaches such a level of fidelity that we could indeed speak of real human scenarios.

To summarise this introduction, the objectives we hope to achieve are:

- To create the software infrastructure necessary to support the next generation of evolving infohabitants in an open and scalable fashion, using existing Internet infrastructure and existing hardware resources.
- To unify evolution approaches, so that infohabitants can evolve using a number of complementary mechanisms.
- To allow meta-optimisation procedures, so that the algorithms for evolution themselves can be optimised by co-evolving a virtual world with the infohabitants it contains.
- To create the software infrastructure necessary to support the emergent virtual economy that will result from the implementation of virtual machine onto physical resources.
- To demonstrate the usefulness of the infrastructure by using it to implement two applications which can make full use of it.
- To facilitate an improved understanding of the dynamics underlying real-world economic and social systems by simulating these systems with the DREAM.

2 Collective Intelligence

The DREAM sets up the essential components of a technology that can realise a new style of computing: an

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

evolving infohabitant society embodying a collective intelligence. Such a collective intelligence has an intrinsically distributed, adaptive, and self-organising nature. Competition between individual infohabitants or between sub-populations gives a selection pressure which guides evolution towards evolving high quality solutions to problems. Co-operation and negotiation between individual infohabitants and between sub-populations complements the competition to provide a collective intelligence for effective problem solving by division of the problem – a key to solving real, very large problems.

An evolving infohabitant society will be physically placed in a distributed hardware and software environment. Technically, this environment is created by integrating numerous computers through advanced communication technology. Such an infrastructure provides a very natural basis for holding a variety of local environments with different local conditions reflecting different aspects of a problem. The results of individual and collective learning processes on local level can be disseminated globally by migrating and recombining pieces of knowledge.

3 Virtual Economy

The mutual use of other people's hardware and software in the distributed environment will give rise to a virtual economy, where some virtual currency unit is needed for periodical clearance among the owners of infohabitants (the participating individual users and organisations). Such a virtual economy is expected to take effect and to have a significant impact on the future information society. Besides the technological and scientific objectives on computational level, the DREAM project constitutes a long term, real life experiment of such a virtual economy. It is hoped that the insights gained can be translated into recommendations regarding future policy measures.

We intend to tackle the problem of accounting for use of computer power by infohabitants and their owners, and for provision of computing power by others or the same owners at different times. This will require the introduction of a currency unit which can be earned and spent by human users. This currency may also be used by infohabitants to trade information, so that an infohabitant has an incentive to find out things that will be useful to other infohabitants. Additionally, migration to, or existence in, some virtual niches, might also involve costs to infohabitants. Of course, the use of any currency, its storage, and transmission, poses its own security problems. However, DREAM is not concerned with this issue, we will make use of off-the-shelf solutions.

Some political issues may become apparent here, such as "should information be free?". We need to ensure that people feel secure to give their spare CPU time in the knowledge that they can draw on these "banked resources" when required. We must also consider the relationship between hard resources (i.e. those provided by an infohabitant's owner) and earned resources (i.e. those which an infohabitant had acquired through trading with other infohabitants). The relationship between the resource wealth of an infohabitant and its fitness will also require investigation.

The emergent economy that will result, is not only necessary for the Universal Information Ecosystem to fit into the real world, it might also provide novel insights into the ways in which economies work, as many macroscopic features will become available on demand (see section 6 for details).

4 Infoworld Evolution

Evolution is a central component of natural and artificial adaptive systems. We must also achieve this property within the infoworld by incorporating the feature of evolution. By means of evolutionary processes in combination with huge numbers of infohabitants, the system as a whole is expected to produce qualitatively new behaviour yielding new solutions or new insight into real-world problems and into simulations of real-world scenarios.

The evolutionary component of the infoworld essentially splits into variation operators (mutation, crossover of infohabitants), co-evolutionary operators on the level of sub-populations of infohabitants, and evolutionary processes across different niches. The variation operators will be defined for each class of infohabitants; these might go beyond the well-known standard representations such as real-valued vectors (evolution strategies), binary or discrete-valued vectors (genetic algorithms), permutations of integers (order-based genetic algorithms), or even computer programs in LISP-like languages (genetic programming).

For the standard representations, the corresponding mutation and recombination operators can be taken from existing work on evolutionary algorithms, while a more complex representation of e.g. an infohabitant's survival strategy, an infohabitant's foraging strategy, an infohabitant's norms and values, or others, will require an internal language with corresponding specialised variation operators. Based on the partners' experience of the above mentioned evolutionary algorithms, new operators or operator combinations (for mixed representations, or even no representation) will be designed, implemented and experimentally tested.

In addition to the variation operators, also corresponding selection operators need to be realised. These selection operators are necessarily local and become active only among a local group of infohabitants. Criteria for differentiating fitness or survivability among infohabitants as well as local strategies for selection need to be developed and implemented.

A major aspect of the work is to incorporate the feature of self-adaptation into the evolving aspects of the system. By means of self-adaptation, collective intelligence is known to emerge from the interaction of few individuals in evolution strategies already, such that self-adaptation in a system with large number of heterogeneous infohabitants is a fundamental issue to achieve qualitatively new behaviour on the level of the whole system.

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

Locally evolving populations can benefit strongly from segregation into sub-populations with different goals, which co-evolve by competition between the corresponding sub-populations (e.g., a population of test cases might co-evolve with a population of problem solvers so as to find improved solvers as test case hardness increases in the co-evolving population). Also, different populations might exhibit different goals, such that an additional feature of high-level evolution can be incorporated into the system by modelling co-evolution. Incorporating this feature into the infoworld guarantees that different goals of infohabitants can be taken into account by segregated, but interacting populations.

The infoworld is naturally spatially distributed over the Internet, with an interconnection structure among the infohabitants which is neither predefined nor static over time. Instead, the interconnection structure (an underlying graph) randomly emerges and evolves over time, depending on processes of birth, death, migration, variation, and selection. It is this open structure of the infoworld which forms one of its fundamentally new capabilities.

Existing work on spatially distributed populations is completely focused on fixed, regular and static interconnection topologies and is therefore a restricted version of the infoworld. The specification and implementation of the control mechanisms of the infohabitants over their spatial interconnection is necessary, including the implementation of the essential operators to build the spatial structure. Currently, we consider operators allowing migration of infohabitants, communication between infohabitants, recombination between infohabitants, and selection among infohabitants as essential for building and changing this structure.

5 Methodology

The DREAM as a framework for applications has two separate parts: the infoworld and its infohabitants. The infoworld must deal with generic infohabitants, and be able to make them evolve independently of their internal representation. This means that the infohabitants must have a generic interface that allows them to evolve, or at least that the infoworld must know how to make them evolve. To make that possible, an infohabitant abstract data type will be designed; all real infohabitants should be instances of this data type.

The system must be able to work in a wide range of platforms and operating system, and thus must be based on standard industrial tools. As a result we propose to use object frameworks to implement all objects (infohabitants, infoworld, evolution operators). Such frameworks allow one to define the interface using Interface Definition Languages (which are part of the Common Object Request Broker Architecture framework [5]), without worring about the actual implementation, and also allow to program objects in a variety of languages, since the IDL can be compiled to many different languages. This will allow to program different parts of the infoworld in different languages, and allow the user to work with the language with which he or she is more familiar, enhancing the openness and universality.

6 User Interface Considerations

Evolving systems can be scaled up enormously. Current experimental research in the areas of evolutionary computing, artificial life and artificial societies is based on populations of a couple of dozens, hundreds, or at most thousands. The DREAM aims to provide the technology to use millions of infohabitants. This has two major scientific consequences:

- A big leap regarding the size and complexity of the problems that can be tackled
- The necessity for new tools for scientific analysis of the nature of collective intelligence, self-organisation, emergent behaviour, and complexity. By being able to simulate virtual societies at large scale, the DREAM also provides a powerful tool to underpin a new kind of experimental social and economic research.

It is unclear and unpredictable how such a massively distributed, open-ended adaptable information ecosystem will behave over time. The emerging features of the system are likely to be completely innovative, and observation tools need to be designed carefully because we do not know beforehand what kind of information is important to observe. Furthermore, control and steering mechanisms of current evolutionary computation techniques might be insufficient to cope with such a new level of complexity. It hence becomes necessary to have a flexible graphic tool allowing one to design new observables on the fly, in order to verify globally some hypotheses that could only be guessed by partial human observation.

The input interface will allow the user to define precisely the characteristics of the target infoworld, and the kind of infohabitants that will evolve in that world. For the infoworld, this includes the description of the time and space scales, the resources that are available to the infohabitants as well as the response of the infoworld to the existence and actions of the infohabitants. The design of the infohabitants will include their class, their genotype (what are they made of?), their phenotype, especially their interfaces with the infoworld (how do they act on their environment, and react to environmental changes?) as well as with the other infohabitants (how do they communicate?) and the diversity-modification operators (how do they evolve?).

Some basic building blocks will be chosen in some predefined list of standard types, then assembled and structured using simple graphical moves. This construction will be totally recursive, i.e. newly created building blocks can immediately be used among the original ones to build even more complex ones.

The output interface will allow the user to monitor different observables, at the infohabitant level, at the infoniche level, and at the infoworld level. Any feature

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

defined in the input interface is eligible to become an observable, and should be made available either directly for each infohabitant or through statistical measures for the infoniches and the infoworld.

Candidate measures include those from standard statistics - from simple means and averages to Principal Component Analyses and non-linear multi-regressions, in order to detect emergent correlations. But more recent measures that were specifically designed and/or used in Evolutionary Computation will also be considered, addressing fitness landscape description, genetic diversity or operator usefulness, for example.

The user should be able to select any measure of any topic at any time and at any level. For instance, he can ask for some existing observables for a given infohabitant, or summaries of some observables for certain types of infohabitant at a given infoniche, or global summaries of such observables over the infoworld. All these features will then be displayed in real time, or stored to be later examined off-line.

The expected result from the graphical output interface is to facilitate the detection of emergent properties of the infoworlds/infohabitants. It is well known that graphical plots can make things more visible. The underlying hope is that yet unknown emerging properties will be first detected by human analysis of a very small number of infohabitants. The subsequent possibility of immediate and easy monitoring of interesting measures of a higher grain (without having to compile or reprogram anything) allows one easily to confirm (or deny) early hypotheses.

7 Implementation Considerations

Each virtual machine or infoworld is by itself an implementation of a machine-within-a-machine; so we must understand and design it as such. We must create a toolbox to design these machines so that:

- They work as intended (that is, they carry out the problem they were designed to solve).
- They do not affect their hardware environment (that is, they carry out their work safely).
- They use the resources they are given, and at the same time, by using them, they generate resources (they use "virtual currency" to run, and generate virtual currency if lent for a certain project).
- The user knows their state and what they are doing in each moment (a proper user interface should be designed, that allows proper visualisation of the evolution process and the infohabitants, and allows the user to affect it).
- They must all work in the same way, i.e. they must have common interfaces that can be used independently of the inner workings (in the same way that any kind of car can be driven by any driver).
- All these components must be in an easy-to-use toolbox, and must be compatible with any kind of evolving infohabitant.

The infoworld must be designed to have the following properties:

- It must be defined as an environment in which the user's desired evolution is taking place. To that end, it can be a co-evolutionary environment (the infohabitants themselves define fitness), or a pure optimisation environment, in which a (possibly complex) function is going to be optimised.
- The infoworld exists also as an infohabitant in a universal ecology of many infoworlds. It should be possible to parameterise its workings to make it evolve along with the infohabitants it holds. This would allow users to pursue complex optimisation problems like meta-optimisation (find the best algorithm for an optimisation task) or changing fitness environments (the fitness, which is defined by the infoworld, could be defined according to external inputs).
- The infoworld must communicate with other infoworlds and with the user; the interface must show some infoworld properties that can be read and/or changed by the user. The infoworld, for security reasons, must define with which infoworlds it is able to communicate and which infohabitants it is able to accept. It should not be possible for any infohabitant to migrate to any infoworld.
- The infoworld must provide a distributed and scalable accounting and clearing mechanism for resources as provided by infohabitant owners and used and traded by infohabitants.
- The infoworld must contain completely the infohabitants it holds; security measures will be implemented to avoid problems, such as an infoworm escaping from the infoworld and affecting the host machine.

8 Proof of Principle Applications

Three proof of principle applications will be implemented in order to test the DREAM framework. The application areas are carefully chosen such that they together span a wide scale. Namely, we cover three types: optimisation (scheduling application), modelling (data mining application), and simulation (economic and social behaviour). The deliberately different focuses are meant to provide feedback on the general underlying paradigm, from different angles.

8.1 Economic and Social Behaviour

The objective here will be to apply the DREAM as a simulator for human societies, focusing on economic and social aspects and restricted to macro-scale analyses. This simulator is meant to evaluate a given policy (direct problem), and to support the design of new policies (inverse

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

problem). In economic and social simulations, the system is expected to facilitate simulations at a new quality level, with an expected explanatory power which goes far beyond current simulations, approaching capabilities of real-world scenarios.

Classical economy is based on Adam Smith's invisible hand principle: the stable global optimum is reached when all agents look after their own well-being[6]. However, this does not hold in a number of cases, among which are the famous "Tragedy of the Commons" [7] or the "Iterated Prisoners' Dilemma". These are paradigms for many current situations, where the concern is the agents' interaction, rather than the quantity of goods or the price levels.

One limitation of existing economical simulations (for example, J. March, based on models of risk takers and avoiders, or Schelling, using games of life to model, for example, the effects of racism), is that they do not take into account the intrinsic opportunism of social and economical agents: they anticipate other individuals' behaviour, thus modifying their own, which in turn amounts to modifying the rules of the game. Communication among the DREAM infohabitants, as well as the size of the population will allow for such aspect of co-evolution to be taken into account.

At first we will try to see the effect of different culturally-founded behaviours (for example competition-oriented, altruist or diversity-oriented) on the global steady state of the whole infoworld. Further, the inverse problem will be approached: are there some rules that can be enforced in order to modify the final state of the infoworld? In other words, in the case of tax policy for instance, an expected outcome would be to replace the old cultural and historical arguments with new simulation based strategies.

A typical example might concern road traffic load balancing: the direct problem is to determine how the construction of a new road will impact the overall traffic; the inverse problem deals with designing new roads (or toll policies etc.) in order to minimise the overall travel time of the population under a range of load conditions.

Another example application concerns tax and welfare policies. The infoworld can be initialised with infohabitants representing groups that differ regarding their wealth status and wealth-gaining behaviour. The effects of various tax collection and redistribution (welfare) policies can then be monitored on-line and aggregated in relevant statistics. An evolutionary approach also facilitates the emergence of individual responses to various policies, such as tax paying/not paying behaviour.

The evolutionary approach is well suited to modelling economic and social processes on the large scale: DREAM will allow one to simulate the actual behaviour of a huge number of infohabitants, as opposed on the one hand to econometric models dealing with distributions, and on the other hand with small discrete models. As individuals affect each other in achieving their goals, anticipate other individuals' choices and modify their own choices accordingly, evolution offers the infohabitants an opportunity to create bunches of "niches". This fragmentation of the milieu should allow more precise characterisations of the possible drifts of the infoworld (how could a sub-population side-step or subvert a general directive). Typically, any policy might produce some counter intuitive effects, due to the opportunism of economic/social agents.

This work is based on the authors' experiences with agent-based simulations following [8][9]. Using a cellular-automata based approach with a regularly arranged, 2-dimensional spatial structure, numerous experiments have recently been conducted at Leiden University within a student project seminar on Evolutionary Economics. In particular, the experiments focused on an empirical investigation and comparison of various taxation schemes in trade networks. The implementation was based on the SWARM simulation software developed at the Santa Fe Institute.

At Ecole Polytechnique, two MSc theses in Evolutionary Sociology have been supervised by Michele Sebag and Marc Schoenauer since 1997: the theses focused on the modelling of car drivers with different behavioural preferences (competition-oriented, altruist, diversity-oriented) and the impact of the individual preference distribution on the overall traffic jam.

These simulations, although somewhat related to reality, suffered from the limitations of the model and the available computing power. In particular, too small a number of interacting agents seems to prevent the system from reaching a 'critical mass' where synergistic effects on the macroscopic level can be observed. Furthermore, some aspects in the individual or interaction models were forcedly simplified: In the trade network study, only von-Neumann or Moore-neighbourhoods have been considered, which means that individuals interact with only four or eight direct neighbours; furthermore they all follow the same strategy; In the traffic network case study, individuals plan their trajectory on the basis of the previous day traffic, and do not modify their trajectory on the fly. The models were thus severely restricted in contrast to the evolving diversity of behaviours in real-world societies.

A DREAM model will consist of two parts: at the local level, the infohabitant is modelled through a set of possible actions, preferences and beliefs. At the global level, the infoworld is described by the communication and perception facilities, and the general information shared by the infohabitants. These notions are borrowed from multiple agent systems and artificial life on one hand, computational economy and management research on the other hand

The first part of this work is thus to provide accurate and robust simulators, in order to predict the behaviour of the population under a given policy/environment. Realistic test cases in the field of economic and social applications will be identified. The relevance of the models will be evaluated from their adequacy with observed scenarios (e.g. prediction of perturbations due to environmental changes).

The natural follow-up of such model is to determine the policy leading the infoworld into a target state (equilibrium,

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

periodic, or even chaotic!). The second part of this work is thus concerned with the policy optimisation. The interest of this second phase derives from the fact that the global optimum of the population is not necessarily reached when each infohabitant pursues its individual optimum. This discrepancy, contradicting the key axiom of classic economics (the "invisible hand" metaphor), arises most frequently in domains concerned with resources sharing (the Tragedy of the Commons).

8.2 Distributed Data Mining

A clearly visible trend in the last decade is that more and more organisations are collecting and utilising more and more data. Storing, organising and effectively retrieving data can certainly provide improvements in management and operational activities. A higher level of support can be achieved by processing the data and discovering more or less hidden relationships among data records and attributes. Such activities, commonly called data mining, require advanced techniques from statistics and machine learning [10][11][12].

Usually there are two technical bottlenecks data mining applications have to face: the huge amount of data (many records and/or many attributes), and the very complex, hardly comprehensible relationships between data features. The next generation of data mining tools has to be able to both types of problems. To put it simply, a high level of both efficiency and expressive power needs to be achieved simultaneously. The main objective here is to use the DREAM framework to perform data mining efficiently, while maintaining a model syntax with a high expressive power.

Distributed data mining allows for a significant gain in efficiency compared to sequential approaches. Current distributed data mining applications usually split either the models or the data. In the master-slave model, partial hypotheses are tested against the whole data set. In the pipeline model the data is distributed, but the hypotheses are tested against every subset of the whole data. This approach leads to speedup if the different data subsets are evaluated in parallel, but still has a search dynamic identical to that of standard sequential data mining. The DREAM framework allows us to distribute both the data and the hypotheses. Infohabitants can represent partial hypotheses, and they can be partially evaluated on a subset of the whole dataset. The global communications among infohabitants will allow, for instance, the rapid elimination of unpromising infohabitants, while letting promising ones breed and migrate to other subsets of the dataset. Different models of distributed data mining will be implemented in the infoworld framework. The basis of the pursued approach is to distribute both the data and the population of models trying to describe the data. A merger of techniques from distributed evolutionary computing and distributed data handling needs to be developed. The main objective is to create the technology to analyse very large amounts of data (millions of records).

Another strength of the DREAM based approach to data mining is the high flexibility concerning model formats. An evolving population of models is applied to create models that describe the given data at an appropriate level of accuracy. This, however, does not imply that we restrict ourselves to genetic programming - the most commonly practised evolutionary approach in data mining applications. Rather than seeing evolutionary data mining as a competitor to other approaches, we aim at coexisting and co-operating sub-populations of infohabitants of different styles, such as species of decision trees, ensembles of neural networks, rough data models etc. The two main tasks here are to prevent overfitting and to find the right balance between expressive power of the models (i.e. flexible formalisms facilitating high quality) and simplicity (facilitating "readability" to humans).

This implementation will be tested on very large real-world data sets from the area of marketing and finances. We expect that access to real world data will be only granted under clear security and confidentiality conditions. This work will therefore also provide experience on such aspects of working within the DREAM framework.

8.3 Distributed Scheduling

The proof-of-principle application of distributed human resource scheduling will look again at methods of partitioning scheduling problems for distributed processing. New methods of scheduling and rescheduling by co-operation, negotiation and trading of infohabitants will be developed, that would not be possible without the DREAM infrastructure. New methods of dealing with the competing multi-objectives of the problem will also be investigated.

The requirement is to schedule the distributed human resources in a manner that uses those resources, and associated physical resources, in a more effective and efficient manner, and in a manner that improves the quality of life of the people involved.

The first task will be to discover the nature of the partial partitioning of distributed human resource scheduling problems, so that effective partitioning of the problem space between infohabititants and infohabitant communities can be achieved. Partial partitioning occurs in problem such as these when resources are linked far more tightly to a particular group than to others outside the group, but there is not a total partitioning. [13][14]. For example, suppose we wish to schedule the meetings of all academics in Europe. Most meetings will involve academics from only one institution, however some meetings will involve an academic from another institution and occasionally meetings will involve people from several institutions. We will therefore investigate the possibility of scheduling by allowing different infohabitants to take ownership of different partial partitions of the problem (for example the schedule for a particular institution). The infohabitants will then co-operate, negotiate and possibly trade to attempt to resolve conflicts between schedules.

Human resource scheduling is a multi-objective optimisation problem. There are generally a number of different competing objectives that need to be taken into account. For example, we may want to maximise the

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.

number of completely free days a person has, while minimising travel costs, and maximising availability of people at each meeting. We intend therefore to investigate the possibility of different infohabitants having different objective functions, for example, some thinking that free days are most important while others consider that travel costs are paramount. These different objective functions could also be applied to different infoniches, so that an infohabitant that has a good schedule so far as costs are concerned might not do so well if it travelled to an infoniche where costs were not considered particularly important.

Another aspect that will be considered is that of re-scheduling when changes occur. This will involve infohabitants renegotiating with each other when some aspect of the problem changes, for example somebody remembers that their partner's birthday is coming up and they cannot travel abroad on that day.

9 Conclusion

We have outlined the objectives of a three year research project in this paper. Inevitably, some of the claims are not yet well-founded and some objectives might turn out to be too ambitious. It is our intention to publish technical research papers throughout the project, which look in detail of the various issues. It will be these publications that answer the questions about the soundness, or otherwise, of our planned approach.

10 References

[1] European Community Universal Information Ecosystems Proactive Initiative Objectives: www.cordis.lu/ist/fetuie.htm

[2] Swarm home page: www.swarm.com

[3] Creatures 3 Home Page:www.creatures3.com

[4] Tierra Home Page: www.hip.atr.co.jp/~ray/tierra/tierra.html

[5] Object Management Group Home Page: http://www.omg.org/

[6] Smith, A. "The Wealth of Nations", 1776.

[7] Hardin, G. "The Tradegy of the Commons", in Science No. 162 pp1243-1248, 1968.

[8] Epstein and Axtell, "Growing Artificial Societies: Social Science From The Bottom Up", MIT Press, 1996

[9] Simon, H., "Bounded Rationality Model", MIT Press, 1982

[10] Fayyad, U. Piatetsky-Shapiro, G., Smyth, P, and Uthurusamy, R., editors, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, 1996

[11] Weiss, S. M., and Indurkhya, N., "Predictive Data Mining: A Practical Guide", Morgan Kaufmann, 1997

[12] Witten, I. H. and Frank, E., "Data Mining: Practical Machine Learning Tools and Techniques with Java, Implementations", Morgan Kaufman, 1999.

[13] Paechter, B., Rankin, R. C., Cumming, A, and Fogarty, T.C. "Timetabling the Classes of an Entire University with and Evolutionary Algorithm", Parallel Problem Solving from Nature (PPSN) V, Springer LNCS 1498, 1998.

[14] Ross, P., Corne, D., and Terashima, H., "The Phase Transition Niche for Evolutionary Algorithms in Timetabling", Practice and Theory of Automated Timetabling, Burke and Ross Eds. Springer Verlag 1996. ISSN 0302-9743.

Authorized licensed use limited to: Edinburgh Napier University. Downloaded on August 02,2010 at 13:48:30 UTC from IEEE Xplore. Restrictions apply.