The Application of Learning Systems to Support Decision for Stakeholder and Infrastructures Managers Based on Crowdsourcing

Alfonso Bastías, and Álvaro González

Abstract—The actual grow of the infrastructure in develop country require sophisticate ways manage the operation and control the quality served. This research wants to concentrate in the operation of this infrastructure beyond the construction. The infrastructure’s operation involves an uncertain environment, where unexpected variables are present every day and everywhere. Decision makers need to make right decisions with right information. They need to use computational tools called decision support systems (DSS), but now the main source of information came from common users thought an extensive crowdsourcing (DSS), but now the main source of information came from common users.

Keywords—Crowdsourcing, Learning Systems, Decision Support Systems, Infrastructure, Construction.

I. INTRODUCTION

The lifecycle for a civil infrastructure is a least 20 year or more. So in this time many problems show up and the manager need to approach all of them in a rational way. There is evidence the need of a good Information System to management the infrastructure in the lifecycle, a specific DSS with a transactional (OLTP) and analytical database (OLAP). This database can collect data in real time, using the main users of the public infrastructure. The data is processed in real time using learning models based on artificial intelligence that allow a fast analysis and distribute the information in a proper way to take action in order to fix the problem reported.

The concept of “crowdsourcing” is used in this research to gather raw objective data and subjective information about the public infrastructure. Common users of the infrastructure can notify to the authority throughout an internal system that a problem with potential risk is present in the known point in the infrastructure. Different kind of problems can be reported, such us working signal, stoplight, cracked pavement, etc. and the system analyzes this data/information using learning system that let learn and predict future events. The ultimate goal of this research intent to notified to a respective office, such as minister (public works), major officer, fight fighter department, police department, etc.

Today a user will not have a way to report a problem identified in the public infrastructure. The most probable action is calling the police if the problem is critic. However regular problems will not be reported, and an Infrastructure manager will not have the opportunity improve or avoid accidents that may cause major injuries for users.

II. THE CHALLENGE

The research has developed a model and a prototype of mobile application to achieve this goal. Users are able to report in-situ, geo-reference problems, and then the knowledge engine developed in the server side process this information using learning capabilities.

A Geographic Information System (GIS) is feed with the data process by the knowledge engine allow to manage anomalies in the infrastructure’s operation. Know, there is not only a reported problem by txt or calling, but also, but photo/videos uploaded by the people of the city, the main users of the public infrastructure.

The main information platform bases on a server have the hard task with a sophisticate algorithm to evaluate in real time information and redirect to the appropriate track. The commons problems reported are damage in infrastructure (Ways, bridges, sidewalks, highways), car accidents, bad design of routes, urban highway and highways, early alert for catastrophes, water and air contamination, among others. In this platform, the information can be used for authorities to make a prompt decisions (Minister, Cities authorities, Police Department, etc.)

The research explores the application of learning capabilities in GIS/DSS. A model and prototype for a real infrastructure operation will be created. The fundamental contribution of this research is a better understanding of how learning systems and crowdsourcing can improve decision support in infrastructure management. This understanding will be applied in a framework that can be used to develop a new, or adapt an existing, model for decision making. The proposed framework allows GIS/DSS developers to create models with the capability of learning, evolution, and adaptation, with data ongoing in real-time. The core of the system is based on artificial intelligence with learning algorithms for each case, based in the main characteristics of the data classified as subjective and objective information.

The challenge is to develop a general framework to apply learning systems in the engineering management field to
manage infrastructure. The literature review demonstrated specific models/solutions of learning systems, but mostly applied to operational control. This research intends to create a general framework of learning systems to be applied to decision problems primarily at the level of strategic planning and management control. This will require that the control and treatment of subjective and objective information be carefully developed and related directly with learning algorithms already developed by the artificial intelligence field.

III. THE LEARNING SYSTEM

The application of learning systems in decision support is not necessarily new, but it is certainly underdeveloped. The identification of processes and procedures for the proposed framework will be an interesting contribution to the industry because it will provide an ad-hoc framework specifically for the management field. Any improvement of existing static models or even dynamic models will increase the accuracy of decisions, and if decisions are made with greater accuracy, the industry will benefit.

The concept of “crowdsourcing” is used to get information about the public infrastructure. Common users can notify to the systems that a problem is present in the infrastructure. Different kind of problems can be reported and the system will analyzed this information using a learning system with will learn a predict events, and trend will be explored and notified to a respective office, such as minister, major officer, fight fighter department, police department, etc. Today a user will not have a way to report a problem identified in the public infrastructure. The most probable action is calling the police if the problem is critic. However regular problems, will not be reported, and an Infrastructure manager will not have the opportunity to do can avoid accidents or major injuries for users, can be reports using a current mobile phones. [1]

The point of departure of this research focuses in the model’s learning capabilities. Adaptive models are models which change during time, adjusting their parameters and functions to increase the information available. In particular, there are three kinds of metadata: input data, time factor data, and output. Given a process with N input-output data pairs \[ D_N = \{ (x(t), y(t)) \}_{t=1}^N \], the main objective of adaptive modeling is to choose a model where \( \hat{y} = f(x, w) \), \( w \in \Omega \). [2]. Techniques for approaching this equation have been developed in the Artificial Intelligence field using special models reviewed later in this proposal.

Artificial intelligence is a branch of computer science that has been developed for the purpose of solving the learning problem. The most recognized approach to applying learning is the neural network. Neural networks learn by definition, through a training and/or adaptation, depending of the learning algorithm used in each case. Previous research has studied the applicability of neural networks in construction field as a general application [3], or as more specific solutions to construction problems such as bidding [4], estimation [5], and project procurement [6]. Of late, neural networks have been mixed with other technologies, such as fuzzy logic and genetic algorithm to produce more sophisticated and realistic learning systems as global solutions [2, 7, 8]

IV. THE GENERAL LEARNING FRAMEWORK

A. Introduction

The base for this research was provided by a rigorous content analysis. This analysis discovered the industry’s need for a framework to apply learning components. There are five components identified for a Decision Support Systems, (i) Data management, (ii) Model management, (iii) Knowledge Engine, (iv) User Interface, and (v) User. This research is focused in one of those five components, the knowledge engine. [9, 10].

The initial components for a general learning framework that will be fully developed in this research through an experimental design, is based on case studies that are applied to the initial framework, identifying patterns and general structure that need to be address in a general solution. The internal structures of the knowledge engine pointed in the learning framework are 1) input, 2) learning engine, and 3) output. The input component entails special treatment for the information. The learning engine includes a series of sub-components of a knowledge database, learning algorithm, and feedback and update treatment. Finally, output is a general measure of information used to make the decision and it should have a high and clear correlation with the decision to be made.

![Basic Learning System Diagram](Fig. 1)

B. The Framework

The general learning framework has its foundation in the basic learning diagram. This initial framework addresses the following research question: What are the primary data sources that can be characterized in the maintenance of the infrastructure that influence the model’s design in the general framework of decision support systems with learning capabilities? There are three main components of this framework: Input, Knowledge Engine, and Output.

1) The Input

The framework identifies three different types of input that might require different treatment depending on how it is obtained and used. In particular, special kinds of treatment are prescribed for tacit knowledge. In each case different

---

1 \( D_N \) represents a set of pair input/output, where \( x(t) \) is the input for time \( t \) and \( y(t) \) is Output for time \( t \).

2 \( w \) is an unknown parameter vector for the model structure \( f(x(t), w) \).
transformation functions might be applied, and the case studies are chosen to represent each type of situation. This research identified three types of input: objective, subjective and contractual input.[11, 12]

**Objective Information** is related to meaningful and quantifiable measurement. No judgment, or minimal user subjectivity, is included in the pre-analysis input. The users who provide this information are not critical because it is “human-factor” independent.

**Subjective Information** is directly related to knowledge experts or tacit knowledge or personal judge, which may vary user to user. The level of the knowledge and experience of the users/expert is a factor considered in the determination of input treatment and level of trust.

The tacit aspects of knowledge are those that cannot be codified, but can only be transmitted via training or gained through personal experience. Tacit knowledge has been described as “know-how”. It involves learning and skill but not in a way that can be written down. Tacit knowledge has been found to be a crucial input to the innovation process.

Measurement and validation of expert knowledge as input to the framework is key factor of model development success, it is recognized that different experts might provide different kinds of output for the same problem; the subjective information is not human-factor independent as objective information. The learning framework’s management of these issues must be supported by external information, in direct relation with the experts’ field experience, their positions in their respective companies, and the types of decisions they make.

**Contextual Information** is related to the information of the decision environment, which affects the model, but is not controlled by the organization or the decision-makers. This information affects the performance of the model positively or negatively, and is used as given information. The contextual information is not addresses in this research; however it is located conceptually in the final framework. Future research will address this particular need.

2) The Knowledge Engine

The knowledge engine is the heart of a DSS. This is the most critical component in the framework. It is connected to external components of DSS such as the knowledge database which remains in the Data and Model Management. It is also connected to the user feeding the user-interface with the output evaluated by the knowledge engine. The learning components, which are part of the knowledge engine in this framework is focused on learning engine and learning algorithms derived from artificial intelligence, a branch of computer science.

However, the learning engine it is supported with additional components derived from other sciences such as business rules, case-based reasoning, and optimization models. All of these sub-components are explained in detail later during the experimental design. The knowledge database must be well structured to store historical and/or relevant information. An appropriate structure and design of the database structure let the knowledge engine collect information which is used by the learning components. The AI component is the most complex part of the system, although its use is recommended due to the native application of learning capabilities, hybrid solutions are also applied and might require a separate construction of the feedback and update process. The learning algorithms use specific technologies such as neural networks, fuzzy logic and/or genetic algorithms to solve the learning approach, imperfect information, and optimization problems respectively. The following research sub-questions need to be conducted for type of problem and convergence: What learning algorithms are appropriate for each case? What test can be used to probe convergence?[13]

In general, all mathematical techniques are able to develop a learning capability, though, some have the ability to use feedback and update their internal parameters more easily than others. This the particular case happen to neural networks, which belongs to the artificial intelligence science, includes learning capabilities in their native form, which setting its learning algorithms for different purposes.

**Artificial Intelligence** (AI) provides a full spectrum of techniques which, by definition, incorporates learning capability and changes the model when new information is provided to the systems through an internal feedback and update process. So, it is recommended for the learning engine being based on AI, in order to avoid complex feedback and update processes.

AI let combine technologies already used separately. For example combining fuzzy logic, neural networks and evolutionary computing solve human inference process, feedback and update process, and optimization, respectively. Those techniques are analyzed and studied in the next section. The combination of FL and NN has begun to be applied in many areas, and it is called Neural Fuzzy Systems[8]. For that reason it is review in detail the neural networks, fuzzy logic and evolution computation, ending with a review of the most common combination of these techniques.

Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. The brain basically learns from experience. This brain modeling also promises a less technical way to develop mechanical solutions.

In most cases, neural networks are adjusted and/or trained, so that a particular input leads to a specific target output. Adjustment of the network is based on a comparison of the output and the target, until the network matches the target, or until an allowed error occurs. There are two types of learning: supervised and unsupervised. Supervised learning requires a teacher. The teacher might be a training set of data or an observer who grades the performance of the network results. Either way, having a teacher is learning by reinforcement. When there is no external teacher, the system must organize itself by some internal criteria designed into the network.

Another important internal characteristic of the neural
Advantages: Neural Networks learn system behavior by using input-output data. The representation of the human thinking process is good enough for solving many problems that are either unsolved or inefficiently solved by existing techniques, included fuzzy logic. The networks incorporate the previous information to meet and advance further solutions.

Disadvantages: The major problem is the nature of the “black box”; the general understanding of what there is in the black box is incomplete compared with the understanding provided by fuzzy logic. It is much harder to determine the right structure and layer of the net to solve a particular problem. Manipulating and learning parameters for learning and convergence become increasingly difficult. Another important disadvantage is the number of data sets necessary to train the neural networks. Depending on the number of components and layers, the neural network might require a high volume of information (datasets), one of the most critical elements for the construction industry, the lack of data.

The learning engine of the neural network is based on a learning algorithm, which is affected by its structure. There are many algorithms that serve different purposes. Neural networks have been applied in the construction industry to address strategic decision and many different aspects have been reported in research journals [3, 5, 6].

Fuzzy Logic is an approach to computing based on “degrees of truth” rather than the usual "true or false" (1 or 0) Boolean logic on which modern computing is based. The idea of fuzzy logic was first advanced by Dr. Lotfi Zadeh of the University of California at Berkeley in the 1960s. Dr. Zadeh was working on the problem of computer understanding of natural language. Natural language (like most other activities in life and indeed the universe) is not easily translated into the absolute terms of 0 and 1. Whether everything is ultimately describable in binary terms may be a philosophical question worth pursuing, but in practice much of the data we might want to feed into a computer are in some state in between, and so, frequently, are the results of the computing process. Fuzzy logic primarily concerns the relative importance of precision: How important is it to be exactly right when a rough answer will do?

There are many member functions to define the fuzzy set; each function has its own purpose. The simplest function is represented by straight lines. It is not necessary to create complex member functions; mere superposition of functions is an easy way to represent complex solutions.

Advantages: Fuzzy Logic converts complex problems into simpler problems using approximate reasoning. It can be used to model the uncertainty and nonlinearity of a system. It avoids the complex mathematical models. Fuzzy logic is easy to understand, flexible, tolerant of imprecise data, model nonlinear functions, and is based on natural language.

Disadvantages: As complexity increases, it becomes more challenging to determine the correct set of rules and membership functions needed to describe system behavior correctly. The use of fixed geometric-shaped membership functions limits the system knowledge more in the rule base rather than the membership function base with a high level of requirement of system memory and processing time. “Defuzzification” does not guarantee solutions that operate under all possible conditions. Once the rules are determined, they remain fixed in the system, which is unable to learn. Conventional fuzzy logic can neither generate rules nor incorporate previous information in the rule base, which is very important for pattern recognition.

Evolutionary Computing is the name given to a collection of algorithms based on the evolution of a population toward the solution of a certain problem. Three main types of evolutionary computing techniques have been reported in research studies based on different evolutionary algorithms (EA): genetic algorithms (GA), evolution strategies (ES), evolutionary programming (EP), and genetic programming (GP). Evolutionary computing was introduced in the 1960s by I. Rechenber in his work “Evolution Strategies. GA was developed by John Holland in the 1970s, his main publication “Adaptation in Natural and Artificial Systems” was published in 1975. In 1992 John Koza used GA to develop programs to perform certain task, calling the method Genetic Programming (GP), using LISP as main language programming [14-16].

Genetic algorithm takes its name from genetic behavior. It is inspired by Darwin’s theory of evolution, and by the role of chromosomes in reproduction. All living organisms consist of cells, in each cell there is the same set of chromosomes. Chromosomes are strings of DNA and determine the shape of the whole organism. A chromosome consists of genes, blocks of DNA. Each gene encodes a particular protein. Basically, it can be said that each gene encodes a trait, for example the color of the eyes. Possible settings for a trait (e.g. blue, brown) are called alleles. Each gene has its own position in the chromosome. This position is called a locus.

The complete set of genetic material (all chromosomes) of a type of organism is called its genome. A particular set of genes within a genome is called a genotype. The genotypes are, with later development the basis for the organism's phenotype, its physical and mental characteristics, such as eye color, intelligence etc.

During reproduction, recombination (or crossover) first occurs. Genes from parents combine to form a whole new chromosome. The newly created offspring can then be mutated. Mutation means that the elements of DNA are a bit changed. These changes are mainly caused by errors in copying genes from parents. The fitness of an organism is measured by the success of the organism in its life (survival).

Genetic algorithms have been applied in the construction field in isolated ways [17], but are more commonly found in fusion systems, combinations of different technologies of.
Evolution Strategies were invented in the early 1960s by Rechenberg and Schewefel. ES are typically used for continuous parameter optimization, with a strong emphasis on mutation for creating offspring. Mutation is implemented by adding some random noise drawn from Gaussian distribution. Also, the mutation parameters are changed during a run of the algorithm.

One of the main contributions to evolutionary computing is self-adaptation, which occurs in the context of changing fitness landscapes. In cases where the objective function is changing, the evolutionary process is aiming at a moving target. When the objective function changes, the present population needs to be reevaluated since they have been adapted to the old objective function.

Evolutionary Programming (EP) was initially developed to simulate evolution as a learning process with the aim of generating artificial intelligence. A simple prediction task is to guess the following input symbol in an input stream. That is, considering n inputs predict the (n+1)th and articulate this prediction by the nth. Usable mutation operators are generally given to generate: (1) changing and output symbol, (2) changing a state transition, (3) adding a state, (4) deleting a state, and (5) changing an initial state.

For historical reasons EP has been associated with prediction tasks and the use of finite state machines (FSM) as their representation. Since 1990 EP variants for optimization of real valued parameter vectors have become more frequent and even positioned as standard EP. Today is considered a very open framework in terms of representation and mutation operations.

Genetic Programming (GP) is the youngest member of the evolutionary computing family, and differs from other EC strands in its application. The GP is designed to use parse trees as chromosomes. The mutation operation typically acts by selecting a node of the tree. The resulting chosen sub-tree is from one parent into the other. Depending on the problem at hand, and the user’s perceptions on what the solutions must look like, this can be the syntax of arithmetic expressions, formulas in first-order predicate logic, or code written in a programming language.

Combinations of Techniques. Fuzzy logic, neural networks, and evolutionary computing are techniques that have been used separately and successfully in isolated solutions, but the use of each of them has advantages and disadvantages. A good combination of these techniques makes a model robust and strong but probably much costlier in its implementation. Fuzzy logic captures the imperfect information; neural networks have the learning engine in native form, and the genetic algorithms have the component for optimization in its model. A combination of those technologies offers a clear benefit to systems used in the construction field. The combined use of those technologies has been reported in research journals [2, 7, 8].

An interesting study applying a mixture of fuzzy logic, neural networks and genetic algorithm to construction, called “Evolutionary Fuzzy Neural Inference Model (EFNIM)”, was created by Cheng and Ko. There are multiple ways to combine those techniques. A better understanding of the algorithms, their advantages and disadvantages, and the related decision problems will be a critical factor in the further development of learning engines. [7]

3) The Output

The output must be a measurable variable in order to feedback the system, if not, conversion need to be applied in order to make the information workable inside the learning components. The output is the ultimate information used by the decision maker to make the decision. In the general learning framework, the output is used to calibrate the model, increasing the accuracy of it.

The definition of the output is a critical task for an implementation of a learning system. The type of output can also be classified as objective or subjective information. Objective output is related to information directly compared with information gathered from the field, such as final project cost, final schedule, cost or schedule growth, etc. By the other hand, subjective output is related to information not gathered from field without an analysis with a human factor involved. Examples of subjective outcomes are usually related to a discrete or continue subjective scale, such as: Relative success of a project, relative recommendation to use a procurement method or technology. Use subjective outcome to feedback the system become is a challenge task that designer of DSS should avoid if it is possible.

The real output is compared with the estimated output to adjust the knowledge engine. If the real output, defined also as a target, is subjective, the feedback process requires either an extra step of conversion, or simply adds additional external data to define a set input/output/target.

IV. AN APPROACH OF THE GENERAL LEARNING FRAMEWORK

The initial general learning framework is populated with specific components. In particular, the input identifies three types of information. Additionally, the knowledge engine requires a database to manage the new information which is connected to external DSS components, but represented in the final diagram with dash lines. Finally, the learning components identified appropriate techniques to be used in learning algorithms based on Artificial Intelligence.
The connection between the input and output is a conceptual and a mathematical model based throughout on learning algorithms; the mathematical techniques are the main component of the learning engine and are required to address not only algorithms but also to address decision problem; this issue is addressed in the experimental design. In general, core engines are oriented to some particular technique related to some specific type information – subjective, objective, or context - with the time those engines had evolved and became able to work in a wider spectrum of data input than originally created [18, 19].

In order to construct an appropriate learning model, has been developed a workflow to achieve convergence in the result represented in the Fig. 3. The selection of the appropriate learning engine remains in the input/output characteristics and the DSS designers’ skills. There are implicit benefits identified for each technique used. Advantages and disadvantages need to be addressed by the decision maker prior to the model development. The Figure shows a tendency used today after the rigorous content analysis has been fully developed in the literature reviewed. The tendency shows a progressive use of artificial intelligence in which subjective and objective information let unstructured problems be solved more appropriately than traditional techniques, such as linear programming or regression models.
V. THE PROTOTYPE AND THE APPLICATION

There are a wide alternative of platform to develop a solution to achieve the main goal of this research. One of the most relevant aspects is the data collection that came for the users of the public infrastructure. The nature of the data and the penetration of smartphones make feasible the construction of a software application for this input information [20].

Today’s market is governed by two main alternatives as operative systems: iOS© from Apple™ and Android© from Google™. Although there are other systems available for mobile devices, those were not considered for the prototype construction.

The data collected is sent it to the main server, which has two databases, one transactional (OLTP) and another analytical (OLAP). Both work together to perform a predict model based in a Linux Operative System. In this server is locate the Learning Engine programmed in C++ and linked with Matlab© for the neural network engine. The model construction was based on the diagram shown in the Fig. 4.

![Fig. 4 Data analysis for model development](image)

The users mobile application start when a problem in the street or highway is observed, this problem is reported using the smartphone with the developed application, including multimedia data, such as photography, video, sound or just text. Then this information is processed, tabulated and classified by the learning engine. A smart criteria define if the problem is valid, and require immediate attention, redirecting the report to the respective agency or department.

![Fig. 5 Example of user experience](image)

Finally there is a web application for administrator and responsible for the public and private infrastructure used for people of the city. The access to the web application requires a valid authentication controlled by user and roles. Also an API (Application Programming Interface) is part of the future research in order to allow the connection among multiples platforms.

VI. CONCLUSION

A huge amount of data can be gathered by the users, experts and not experts. To filter and to classify this information is the successful factor of the research. Today’s there is a high penetration of mobile device that can be used to report problems with the extra useful data, such us a photo geo-referenced.

By other hand, the public infrastructure is managed by department which not have enough personnel to validate or certificate that is operating appropriated. A problem reported by a common user can be much faster that a problem reported by an employee of that department. Identify a problem its potential risk of accident and reported to the responsible in order to be fixed has a high positive impact in the society. This makes infrastructure more safe and secure just with an early alert.

REFERENCES

Alfonso Bastias, Ph.D. actually is Director of Civil Engineer and Associate Professor of The School of Engineering at Universidad del Desarrollo, Santiago - Chile. He is an Industrial Engineering from P. Catholic University of Chile. He holds a Master of Science in Civil Engineer at P. Catholic University of Chile. In 2006 he got his Ph.D. in Civil Engineering, with Emphasis in Construction Engineering and Management at University of Colorado, Boulder (USA). He has experience in research areas such as: Project Management, Project Control, Risk management, Information Technology and Knowledge Management. His has been working with private, public companies and research Centre such us National Regulatory Research Institute (NRRI), National Cooperative Highway Research Program (NCHRP), U.S. General Services Administration (GSA), Federal Highway Administration (FHWA), California Department of Transportation (Caltrans) y Construction Industry Institute (CII). During his doctoral studies he conduit Research through CII (Construction industry Institute) with Army Special Forces, ExxonMobil, General Motors, NASA Jet Propulsion Lab, Washington Group, British Petroleum, Fluor, HB Zachry, NASA Facilities, Ontario Power Generation Inc., CSA, y Aker Kvaerner Songer among others.

Álvaro González graduated with a Civil Engineering and MSc degree from the Catholic University of Chile in 2002, specializing in Design and Construction of Infrastructure. Then he worked as a consultant and researcher in highway engineering at DICTUC, where he carried out a number of technology transfer projects, including the first cold in place recycling of asphalt pavements using foamed bitumen. In 2005 he moved to New Zealand to pursue PhD studies on the performance of stabilized pavements under accelerated loading at the Canterbury Accelerated Pavement Testing Indoor Facility (CAPTIF) from the New Zealand Transport Agency. In 2009, after completing his PhD degree at the University of Canterbury, Álvaro joined ARRB Group Ltd (formerly the Australian Road Research Board) as a research engineer and project leader in Melbourne, Australia. Since February 2011 Álvaro is a Lecturer at Universidad del Desarrollo in Santiago, Chile, where he is developing a number of research projects in road construction, pavement recycling and technology applied to construction.