www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

HUMAN COMPUTATION ENABLED ORGANIZATIONAL LEARNING IN THE FACE OF DEEP UNCERTAINTY: EXAMPLE OF CONCEPTUAL ESTIMATING

REVISED: March 2015
PUBLISHED: April 2015 at http://www.itcon.org/2015/18
EDITOR: Ruikar K.

Jing Du, Assistant Professor,
Department of Construction Science, The University of Texas at San Antonio;
email: jing.du@utsa.edu

Mohamed El-Gafy, Associate Professor,
School of Planning, Design and Construction, Michigan State University;
email: elgafy@msu.edu

SUMMARY: *Parametric estimating has been widely used in conceptual estimating to tackle quality and efficiency related issues. Despite the merits obtained, human involvement is always missing from the decision process and organizational learning is difficult to realize because knowledge of experienced estimators is not well preserved. This paper proposes a paradigm and a corresponding information system that combine the intelligence of humans and computational power of computers, streamline the information flow in the mixture of human-computer-network, and solicit and preserve knowledge of estimators in proposal development. The Human Computation paradigm created by Von Ahn has been tailored to tackle the construction conceptual estimating problems. A Cloud computing infrastructure that supports the proposed paradigm is also proposed. In order to demonstrate the applicability and validity of the proposed paradigm and information system, a case study of the proposal development of a power plant project is introduced. The findings confirm that the proposed paradigm and information system can be used to advance knowledge transfer and organizational learning in construction problems where necessary information is not always available. Although designed for conceptual estimating problems, the proposed paradigm and information system can be applied in other construction engineering and management problems fraught with deep uncertainties.*

KEYWORDS: *Conceptual estimating, human computation, organizational learning, information system*

REFERENCE: *Jing Du, Mohamed El-Gafy (2015). Human computation enabled organizational learning in the face of deep uncertainty: example of conceptual estimating. Journal of Information Technology in Construction (ITcon), Vol. 20, pg. 275-294, http://www.itcon.org/2015/18*

1 INTRODUCTION

Conceptual estimating in proposal development is nontrivial (Globerson & Zwikael 2002). Estimators need to develop estimates under significant time pressure and with only limited information (Alder 2006). It always ends up with inaccurate estimates and unqualified budget which further increase the financial risk in construction execution phase (Carr 1989). Parametric methods has been widely applied in conceptual estimating to enable faster and more accurate proposal development since 1936 (Kwak & Watson 2005). Commonly used methods or algorithms include regression analysis (Kouskoulas & Koehn 1974) and Artificial Neural Network (ANN) (Adeli 2001). In addition, Case Based Reasoning (CBR) has been introduced as an alternative in conceptual estimating, where decisions are made based on the experience from similar historical projects (Kim, An & Kang 2004). The feasibility and validity of these methods have been well proven by many case studies (Kim, An & Kang 2004).

However, a closer examination finds certain problems with conceptual estimating, which can hardly be solved by current parametric estimating methods or CBR. First, human opinions are missing from the implementation of most parametric estimating or CBR frameworks. Most parametric estimating approaches are built on historical data (Walton & Stevens 1997), and require advanced statistical or modeling knowledge in order to enable better handling of the models and proper interpretation of the results (Kwak & Watson 2005). The knowledge set of most engineers and estimators are divergent from this purpose and thus most parametric estimating methods remain a black box for the daily users. In practice, estimators have to accept the results instead of actually controlling the estimating procedure. This partially explains why many estimators hesitate to use the latest statistical approaches (Rush & Roy 2000). Second, given the deep uncertainties of construction projects, such as unforeseen site conditions or financial difficulties (CII 2007; Du, El-Gafy & Ghanem 2012; Du, Liu & Karasulu 2014), historical data can hardly cover the entire uncertain space. Fig.1 illustrates a situation where known information only represents a small portion of the entire space. When project decision makers are facing new conditions, such as a brand new project, parametric estimating fails to yield a satisfactory result (Du & Bormann 2014). It is severer when the statistical model is oversensitive to small changes, such as one can see in *multicollinearity* (Farrar & Glauber 1967). Third, different algorithms or approaches are disconnected. Any algorithm or method has its unique advantage in discovering the pattern of data, and therefore multiple methods should somehow be integrated to generate a more solid conclusion where a confidence band is given. Even though new frameworks are emerging, the authors didn't find any one of them attempts to connect different models in order to reach such a solid conclusion. Most studies focus on estimating and comparing the performance of different approaches and make recommendations for a particular method (Kim, An & Kang 2004). The last problem with current conceptual estimating is that organizational learning is hard to be realized. Knowledge and experience of estimators are mostly piecemeal and anecdotal; there is not existing paradigm that preserves the knowledge as a part of organizational learning.

In order to address the limitations of the present conceptual estimating methods and systems as discussed above, this study aims to answer the following question:

Can we develop an innovative paradigm that streamlines information flow between estimators and statistical models, and preserves knowledge of experienced estimators to realize sustainable organizational learning?

The objectives of this study include: 1) to examine the applicability of *Human Computation* in construction problems: Human Computation was proposed at first as a computer technique in which certain steps of a computational process is outsourced to humans (Ahn 2005). The main idea is to utilize unique abilities of humans and computer agents to achieve symbiotic human-computer interaction (Ahn 2005). Such interaction becomes extreme important in problems which neither human nor computer can solve alone (Ahn 2005). This study examined each element of the general paradigm of Human Computation, and identified the corresponding realizations; 2) to tailor the general Human Computation paradigm into construction problems, and to develop a corresponding information system that supports Human Computation in construction problems. Inspired by the principals of Human Computation, this paper proposed a paradigm that combines the intelligence of human beings and computational power of computers, and streamlines the information flow in the mixture of human-network and computer-network in proposal development. A Cloud computing based infrastructure was also proposed to support the knowledge flow; and 3) to demonstrate the applicability of the proposed paradigm with a case study. In order to test the feasibility of the proposed system, the selected case study should be associated with certain level of complexity and uncertainty. Therefore, this study conducted a case study about the conceptual estimating of a mage-project (a power plant project). The remainder of this paper introduces the theoretical background, the methodology, the paradigm and system, and finally the case study. .

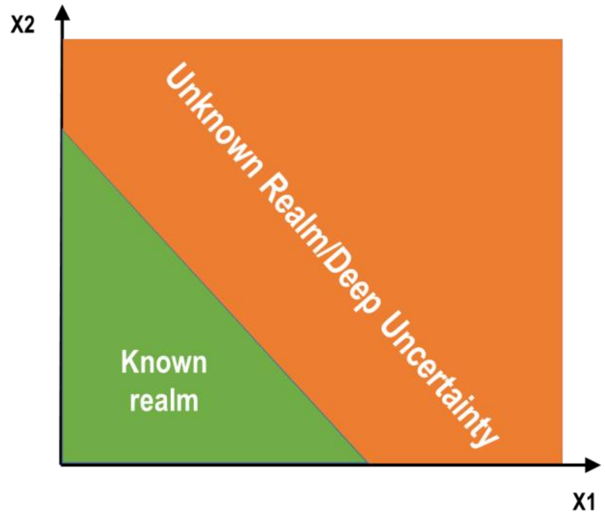


FIG. 1: Deep Uncertainty in Conceptual Estimating

2 THEORETICAL BACKGROUND

2.1 Parametric Cost estimating

Parametric cost estimating can be dated back to 1936, when Wright provided equations to predict the cost of airplanes over long production runs, a theory later be called the "*learning curve*" (Wright 1936). This method was then extended to other industries by Rand company in 1950's and RCA company in 1970's (ISPA 2008). Early in 1994, a joint Government/Industry Committee was established to search ways to increase the use of parametric cost estimating techniques (NASA 1995). This initiative finally resulted in several industry wide parametric cost estimating handbooks (ISPA 2008; NASA 1995). Per the definition of project management body of knowledge (PMBOK), parametric cost estimating is defined as a method that uses "*statistical relationship between historical data and other variables to calculate an estimate for activity parameters, such as scope, cost, budget, and duration.*" (ANSI. & PMI. 2008). It builds on Cost Estimating Relationships (CERs) between the project characteristics and construction project cost, and applies algorithms to determine an approximation of the total construction project cost (ISPA 2008). As a result, the main interest of current parametric estimating research is on searching for better algorithms that can facilitate more accurate discovery of CERs (Kim, An & Kang 2004; Shtub & Versano 1999; Smith & Mason 1997).

Since 1970's, regression method, or normally referred to as multiple regression analysis (MRA), has been introduced to parametric estimating (Kouskoulas & Koehn 1974). It has been used in residential projects (Khosrowshahi & Kaka 1996), early pre-tender estimating (NASA 1995), schedule analysis (Martin Skitmore & Thomas Ng 2003). It is also integrated with expert systems (Bowen & Edwards 1985), and refined by other statistical algorithms to increase the applicability and reliability (Trost & Oberlender 2003). Currently MRA have become a widely used method for estimating cost because it has the advantage of a well-defined mathematical basis as well as measures of goodness of fit to a given historical data set (Kim, An & Kang 2004). The easy-going feature and reliable results make it a popular tool in practice (ISPA 2008). Since 1990's, as the development of Artificial Intelligence and increasing availability of relevant software packages, Artificial Neural Network (ANN) started to gain its popularity in parametric estimating (Adeli & Wu 1998; Ayed 1998; Kim, Seo & Kang 2005). As a non-parametric approach, ANN doesn't require predefined formulas, but allows the data to speak for itself (Kim, Seo & Kang 2005). Because of the advantages of ANN, it is found that ANN fits well to investigating complex nonlinear relationships (Hsu, Gupta & Sorooshian 1995). Recent efforts focus on improving the performance of ANN by integrating other artificial intelligence algorithms such as Genetic Algorithm (Kim, Seo & Kang 2005).

Both MRA and ANN are considered as discovery methods of CERs since they aim to estimate the relationships between project parameters and construction cost, parametrically or non-parametrically (ISPA 2008). In recognition of the importance of historical successful experience, Case Based Reasoning (CBR) is proposed as an alternative to enhance the conceptual estimating (Perera & Watson 1998). It builds on a proposition that new problems can be solved by adopting solutions that were proven to be successful (Ashley 2006). A typical CBR system follows the following steps (Kim, An & Kang 2004): (1) Retrieving one or more stored cases similar to the new case according to the percentage similarity; (2) Adapting the solutions of the retrieved cases according to the differences between the stored cases and the new case, and adopting the adapted solutions to the new case; and (3) Retaining the new solution as a part of the stored cases throughout the test. Because CBR is very similar to how an expert solve a problem, it has been applied for a variety of construction related issues including cost estimating (Kim, An & Kang 2004). In fact, Kim and colleagues (2004) found that CBR has a better performance than MRA and ANN with a respect to long term use. The similarity between CBR and industrial practices has been confirmed by the authors: it is found estimators in a daily basis follow the same procedure to conduct conceptual estimating – by comparing and revising the estimates on the basis of similar historical jobs. As a result, CBR sometimes be considered as an expert system (Kim, An & Kang 2004).

The above brief review of the evolution of parametric estimating indicates a fact where emerging statistical algorithms are becoming increasingly relied. Even though CBR emphasizes the importance of experience, similarity function calculation is still based on a set of algorithms, such as gradient descent method (Kim, An & Kang 2004).

2.2 Human Dimension in Cost estimating

With the rapid development of parametric estimating, Altayeb (1997) raised a reflective question at a time when computer models have become increasingly popular among estimators: *What is more important, human factors or computer factors?* Altayeb observed that certain abilities of human beings can hardly be substituted by computers, such as expertise and intuition to make enlightened decisions among a group of alternatives. On the other hand, he also noticed the importance of computers for their speed and accuracy of processing, organizing, and exchanging information. In conclusion, additional efficiency from computers can be gained only if human beings are able to control the entire process (Altayeb 1997). Similar arguments never end. Increasing evidence indicate the significance of experience of human beings in the success of cost estimating and competitive bidding, which is defined as knowledge and skills gained from the observation of and practical acquaintance with facts and events (Fu, Drew & Lo 2003; Lowe & Skitmore 1994). It seems in most cases estimates are fundamentally a result of estimators' *cognitive psychology* including risk attitude, experience background, stress, familiarity with projects, requirements of customers, knowledge gained from experience, and professional skills (Fellows & Liu 2000; Leung, Ng & Skitmore 2005; Lowe & Skitmore 1994; Mak & Raftery 1992). As a result, some scholars urge a reconsideration of the role of human dimensions in cost estimating – people's choice (Fellows & Liu 2000; HARTONO 2010).

The argument between human roles and computer roles in cost estimating comes from the understanding of the nature of estimating practice. Estimating has long been treated as a reflection of *scientific positivism,* where the requirement of "objectivity" tends to exclude the involvement of the knowledge of the observer (Polanyi 1962). As discussed above, one natural result of positivist trend in cost estimating is the increasing use of statistical models, which are believed as objective projections of the history to the future. Whereas, as commented by Polanyi (1961), there are indeed "*two kinds of knowledge which combine into the understanding of a comprehensive entity: our reliance on our awareness of the particulars is the personal; our knowledge of the entity, the objective element of knowing*". This perspective highlights the priori knowledge and the empirical components of "objectivity". Tauber (1997) also commented, as science has evolved, the notion of what constitutes "objectivity" has also changed since the standards of the boundary between "objectivity" and "subjectivity" is evolving. If personal knowledge (e.g., intuitions and insights) is excluded from decision making, "*questions arise as to how investigations are instigated, how they are carried out and how conclusions are formulated.*" (Fellows & Liu 2000). Any decisions in cost estimating are directed by cognitive motivation of estimators (e.g., interest, goal directed actions) and aim to answer the basic human initiated investigative questions (Fellows & Liu 2000). Especially when construction systems are regarded as social-technical systems where human and organizational factors exert strong influences on the decision makings (Du 2012; Du & El-Gafy 2010; Du & El-Gafy 2011, 2012; Du & El-Gafy 2014; Du & Wang 2011), human experience and knowledge constitute an important dimensions in construction management practices including cost estimating. Unfortunately, human dimension remains a missing link in current parametric estimating.

This paper urges a rethink to the role of human beings in parametric cost estimating, particularly at a time when artificial intelligence is highly valued and relied on. The real challenge is to enable a seamless synchronization of human intelligence and computer computational capacity. In addition, such synchronization should be able to encourage the solicitation and application of human intelligence, and ultimately allows a sustainable organizational learning. Recently emerged human computation paradigm provides a possible solution.

2.3 Human Computation

A widely accepted definition of Human Computation is given by Ahn (2005); it is "*a paradigm for utilizing human processing power to solve problems that computers cannot yet solve.*" It reverses the traditional computation process where human formalizes problem and computer offers solution – in Human Computation paradigm, computer is used to solicit human opinions, and then integrate and interpret the opinions (Ahn 2005). The original intention of Human Computation is to deal with certain problems which are challenging for computers but trivial for human beings, such as image recognition (Ahn 2005). Recent representative works include visual recognition (Ahn 2005), character recognition (Von Ahn et al. 2003), language understanding (Bernstein et al. 2010), and human communication (Chen et al. 1999) etc. For example, in the reCAPTCHA project (Von Ahn et al. 2008), system makes use of the human cycles to help digitize books which optical character recognition (OCR) software is unable to read: the reCAPTCHA system presents images of words scanned from old books or street pictures for humans to decipher, as part of normal website validation procedures (i.e., the validation words; see Fig.2). Then the results are returned to the reCAPTCHA service, which are integrated to the digitization projects.

FIG. 2: The application of reCAPTCHA; retrieved from Gmail: http://mail.google.com (Aug. 2014)

Scholars found roots of human computation in several existing theories including crowed sourcing (Howe 2008), social computing (Parameswaran & Whinston 2007), and collective intelligence (Malone, Laubacher & Dellarocas 2009). The commonality between these technologies and human computation is concentrating on soliciting human knowledge and facilitate human collaboration with the aid of computers (Quinn, A.J. & Bederson, B.B. 2011). Whereas Quinn and Bederson (2011) describes six properties which distinguish human computation from other similar ideas, which are *motivation*, *human skill*, *quality control*, *aggregation*, *process order*, and *task-request cardinality*. According to the nature of the problem, human computation may have different process order and task-request cardinality, but a typical procedure of human computation can be outlined on the basis of properties proposed by Quinn and Bederson: at first, tasks are requested by the users who benefit from the computation. Then these tasks are formalized and assigned to the workers who will contribute opinions. Depending on the applications, a variety of human skills, specific knowledge or abilities held by the workers are leveraged. A variety of incentives might be provided to motivate people to participate. The opinions provided by workers are then aggregated to combine the efforts for solving the global problem. Statistical processing of data and iterative improvement are performed to validate the solution and remove any sabotage or misunderstanding. Quinn and Bederson (2011) also notice an opportunity of better training computers during the human computation process. For example, in order to train a pattern recognition machine such as a classifier, it requires a large quantity of example patterns along with annotations (answers from humans). In most cases, obtaining enough amounts of annotations is extreme labor-intensive and time consuming. Human participants provide significant quantity of annotations which can be used as inputs for machine learning. Based on the previous discussion, a paradigm of a human computation system can be outlines as Fig. 3.

Human computation has now been applied in a variety of areas, including proofing reading the characters that can hardly be processed by OCR (DP 2015), Galaxy Zoo (a tool that utilizes numerous online players to classify a large number of Galaxies) (Lintott et al. 2008), Stardust@home (a NASA project that encourages volunteers to search images for tiny interstellar dust impacts) (NASA 2015), and BONIC (a distributed computing project at Berkeley that utilizes idle computing resources of volunteers) (Anderson 2004). These applications have shown that Human Computation is able to produce "many novel ideas aimed at organizing web users to do great things" (Quinn, Alexander J & Bederson, Benjamin B 2011). However, human computation has not yet been widely applied in the construction engineering and management areas. A wide range literature search has been conducted by using the keywords of "*construction*" or "*engineering*" plus terms relevant to the use of Human Computation such as "*human computation*", "*crowdsourcing*", "*human-based computation*", "*human-assisted computation*", "*ubiquitous human computing*", "*distributed thinking*". To the best knowledge of the authors, there is no single publication addressing the application of Human Computation in the construction engineering and management areas. In most conceptual estimating, the implicit knowledge and experience of estimators are of the center interest which computers are incapable of (Altayeb 1997). Literature confirms the difficulties and importance of soliciting and conserving knowledge of experienced estimators to enable a sustainable organizational learning (Fu, Drew & Lo 2003). Human computation provides a possibility that combines human intelligence and artificial intelligence, and enables the information flow between human network and computer network. Enhanced by data mining techniques (such as machine learning), human computation is promising in conceptual estimating.

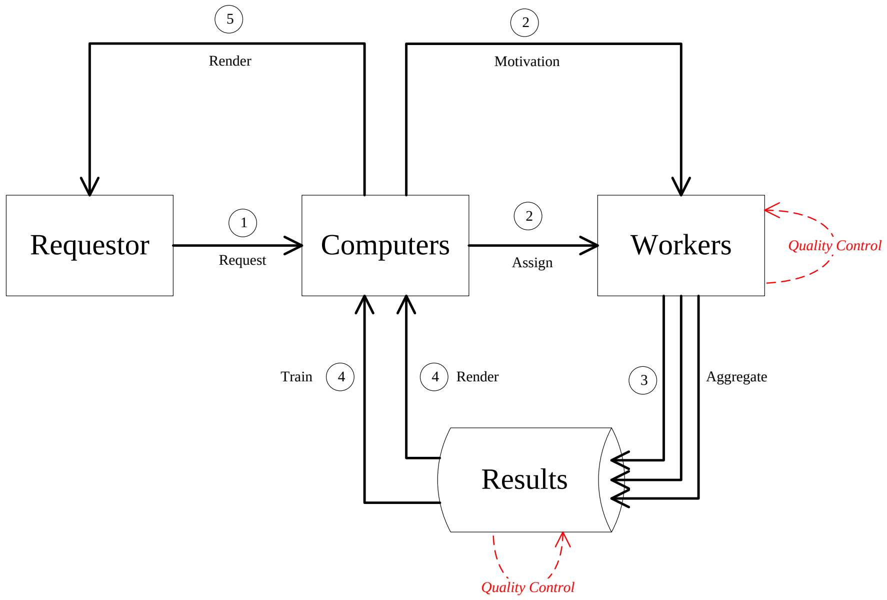


FIG. 3: Paradigm of human computation

3 METHODOLOGY

This study follows a three-step approach:

- **Step1:** To examine the applicability of the Human Computation paradigm in the area of construction conceptual estimating. The Human Computation paradigm created by Von Ahn has been applied in a variety of areas other than construction. Therefore, it is critical to examine if the important components of human computation can fit into the construction problems. Specifically, can the five elements proposed by Von Ahn (*motivation*, *human skill*, *quality control*, *aggregation*, *process order*, and *task-request cardinality)* be used to address construction problems, such as conceptual estimating, completely and effectively? The accomplishment of step 1 will help identify the components of a tailored Human Computation paradigm in the construction context.
- **Step 2:** To tailor the general paradigm of Human Computation into construction problems. On the basis of step 1, each component of tailored paradigm will be examined fully, including the definition, subsystems, methods, and processes. The work process of applying the tailored paradigm will also be developed to integrate the five components. In addition, a Cloud computing based information system will also be proposed to automate the work process. The accomplishment of step 2 will make the system ready for applications.
- **Step 3:** To demonstrate the applicability and validity of the proposed paradigm and system with a case study. Given that complex projects tend to have more issues with deep uncertainties and possibility larger knowledge gap (e.g., the knowledge gap between engineers and estimators in a very complex power plant project), preferably the selected case study should represent certain level of complexity. As a result, this study will examine the conceptual estimating of a power plant project, and how the application of the proposed paradigm and system help streamline the estimating process, make better use of knowledge and enable organizational learning.

The remainder of this paper will introduce the findings.

4 THE PARADIGM: INTERACTIVE PARAMETRIC ESTIMATING

4.1 Architecture

Building on human computation theory, this paper proposes a paradigm for conceptual estimating and relevant organizational learning. The proposed paradigm, interactive parametric estimating (iPE), highlights the collaborative effort of human and computer network in conceptual estimating phase, and supports the solicitation and preservation of knowledge. There are five key modules as shown in Fig.4: *Human Skills*, *Cost estimating Relationships (CER)*, *Validation*, *Optimization*, and *Learning*.

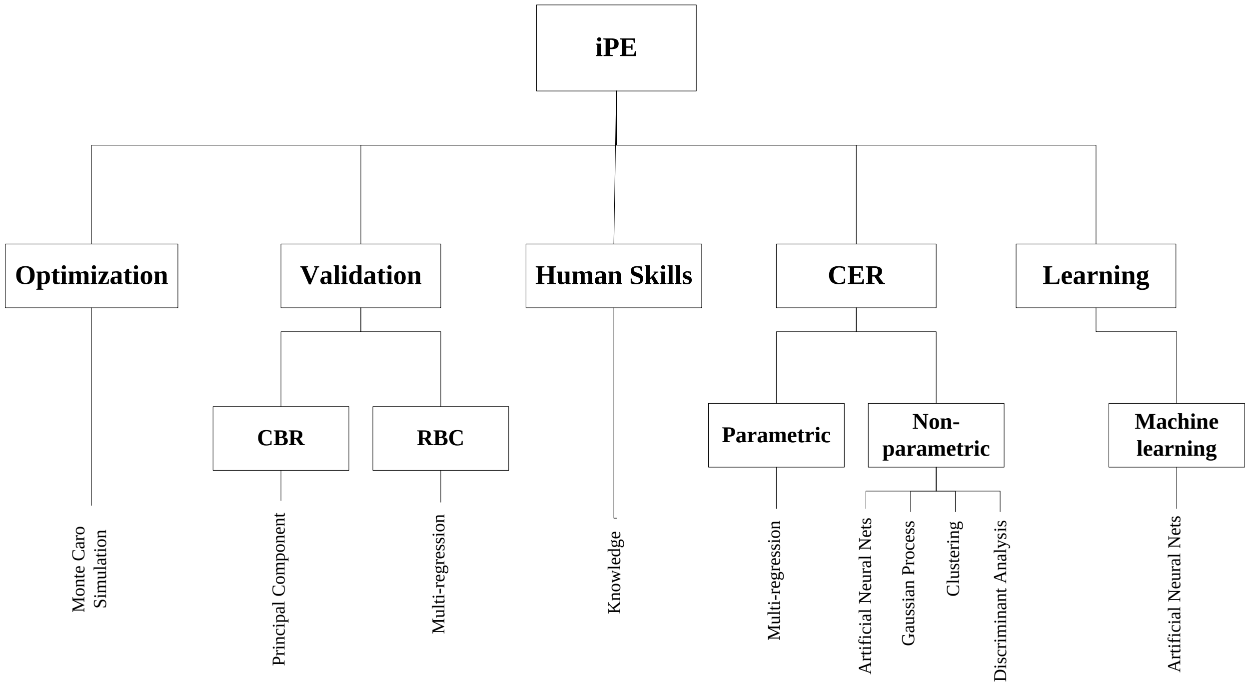


FIG.4: Architecture of proposed paradigm

Each module supports an important functionality to realize the paradigm.

Human Skills, specifically the experience and knowledge held by experienced estimators about construction projects, estimating and relevant factors, play a central role in the proposed paradigm. One of the core purposes of the paradigm is to develop a standard process that facilitates the application of human skills and takes full advantage of human skills in conceptual estimating. Experienced estimators are assigned the right to adjust and/or override the decisions. The other functionalities are applied to solicit, formalize, enhance and preserve the experience and knowledge of experienced estimators.

CER, explored by parametric and/or non-parametric models, is used to discover the patterns of historical data, which serves as the preliminary analysis for decision making. Building on the Grounded Theory (Glaser & Strauss 1968), a variety of statistical models are established to provide multiple sources of opinions which constitute the foundation of preliminary decisions. Information visualization is also utilized to facilitate the decision making. For example, Multidimensional Scaling (MDS) might be utilized to enhance decision makers' understanding on the similarities between projects (Du & El-Gafy 2011).

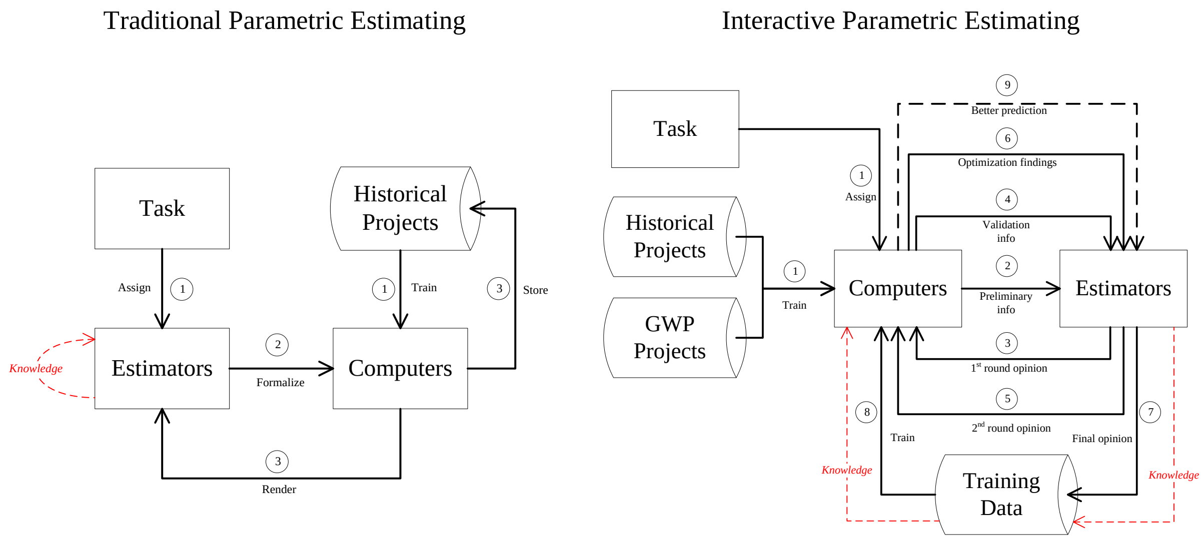
Validation provides third party sources of information to allow the decision makers to adjust the preliminary decisions. Unlike the statistical models of CER, techniques utilized by validation are focused on different perspectives to describe the pattern of historical data. For example, a CER model might focus on revealing the quantitative relationship between project characteristics and craft quantities. In contrast, a validation model attempts to capture the quantitative relationships between crafts (RBC) such as the reasonable ratio of concrete reinforcing to concrete. In another case, a validation model could also provide opinions based on the similarity between a project and historical projects, which utilizes methodology of Case Based Reasoning (CBR). It identifies similar projects in the history, and make connections to the present project. It can be used to facilitate the comparison thinking by linking new projects to archived projects. A method developed by the authors, called "Similarity Measure And Ranking Tool (SMART)" has been used to realize CBR. Limited by the scope of this paper more details about SMART cannot be provided in the present manuscript, but readers can refer to a paper we published in 2014 for more information (Du & Bormann 2014).

Optimization is applied to provide a global insight into the entire estimating process. Discipline based estimating practice tends to lead to local optimized solutions. It is very difficult for craft-based estimators to cooperate together to reach a global optimum. This is either unrealistic in most cases or out of capacity even when it is targeted. For example, using longer pipes in power plant projects might increase related material cost; but it saves direct labor work hours in the installation phase because the installation of longer pipes is often more efficient. Minimizing total install cost (TIC) therefore involves the simultaneous consideration of both factors (material cost and installation labor cost), which is hard to achieve using discipline based estimating (Steffey & Anantatmula 2011). In optimization module, Sensitivity Analysis method (Du & Bormann 2014) is employed to discover the distribution of ultimate TIC with an overall consideration of the material price escalation (Knight & Fayek 2000), risk mitigation (Molenaar 2005; Pyra & Trask 2002), probability distribution of productivity (Schexnayder, Knutson & Fente 2005) and etc. Consequently, decision makers are able to adjust the quantities of particular crafts and choose certain risk mitigations to increase the efficiency.

Learning aims to preservepreserve the knowledge gained in the process of decision making, discussion, and other activities in the conceptual estimating phase. Instead of solely focusing on the actual data of history, the proposed paradigm values the importance of estimators' opinions. It builds on an assumption that certain characteristics of a project are very difficult to capture, which in contrast could be easily captured by implicit knowledge or experience. In many cases, especially when the situations are relatively new, the judgments made by the estimators differ from the predictions made by the statistical models. The proposed paradigm assigned bigger weights to the human-made judgments to tackle deep uncertainties. It requires a method to preserve human-made judgments. Through designed procedure (such as Game With a Purpose) and machine learning techniques (such as ANN), it is expected the implicit knowledge and experience can be structuralized and preserved for organizational learning purpose. It has been found that ANN has a variety of advantages to learn from data, and therefore, it is used in the learning component of the proposed paradigm. In the case study, it was realized by JMP. In order to realize learning, the judgments made by the estimators are used as the "training dataset" of ANN models. The target of ANN modeling is to adjust the structure by data training. In a training iteration (or an epoch), values of input variables are weighted and summed up for an artificial neuron. Then an activation function is applied to convert the summation of weighted input to output activation.

4.2 Analysis Flow

In a common parametric estimating practice (Fig. 5 (a)), statistical models are built based on the historical data. Then different statistical approaches are compared according to their goodness of fit, such as R square or mean absolute error to actual construction cost data (Kim, An & Kang 2004).



(a) (b)

FIG.5: The flowcharts of traditional and interactive parametric cost estimating

The “best” model is finally selected and applied to the new projects to conduct parametric estimating. This process builds on a weak assumption that the future will stay consistent as the history, and thus the historical relations between project parameters and construction cost can be used to future conditions without refinement. Even though the conclusions obtained from this process can be used to support the decision makings of estimators, estimators are not actually involved in. Soft knowledge is not well solicited and used to deal with unique conditions of new projects which probably have not been captured by historical data and relevant statistical models.

iPE, in contrast, combines the advantages of statistical models and soft knowledge of estimators to enable a seamless information flow in the entire estimating process. Fig.5 (b) describes the analysis flow of iPE:

(1) A variety of statistical models are built based on historical data and GWPs, i.e., hypothetical projects for soliciting opinions only;
(2) Opinions from all the statistical models are presented to the estimators in a single graphical demonstration so the differences can be perceived by the estimators;
(3) Based on the preliminary opinions provided by statistical models, estimators make the first round judgments;
(4) The first round judgments of estimators are compared with third source information, such as the reasonable ratio between crafts, or similar historical projects facilitated by CBR;
(5) Estimators revise their judgments based on the information obtained in step 4;
(6) The second round judgments from different disciplines are fed into the optimization module, where Sensitivity Analysis simulation is conducted to investigate the optimal estimates for each craft;
(7) Estimators make their final judgments (3rd round) based on the findings of optimization analysis; as a result, the characteristics of new projects, reasonable relations to historical projects, and global optimization are all considered in the decision making;
(8) The final decisions of estimators are then used as training data to improve the performance of statistical models (such as ANN models);
(9) Improved statistical models not only reflect the fact of the history, but also reflect the perceptions and opinions of estimators towards the future, and therefore more reliable estimating can be made based on new models.

In this way three iterations of information exchange are realized between computers and humans, and implicit knowledge and experience of experienced estimators are solicited to revise the final decisions and train the models for future use.

4.3 Information System

The proposed paradigm can be realized with an information system as shown in Fig. 6. Multiple users, i.e., estimators, can access a central database server and all available tools (e.g., predictive models) via thin clients (e.g., personal computers). The central server processes the requests and operations of estimators and performs the major computational tasks including statistical modeling and learning. It is also able to return the results to estimators through a Cloud-based infrastructure. This information system ensures a master database and streamlined information flow.

To be noted the configuration illustrated in Fig. 6 represents a private cloud environment, i.e., a cloud infrastructure that can be provisioned for open use only by company employees (Subashini & Kavitha 2011). It applies the *client - server* model of centralized computing (Nieh, Yang & Novik 2000). There are three major components in this design:

- Client: A thin client that relies heavily on network resources for computation and storage. The user interface of the information system is realized in MS Excel or web browser, as both are user-friendly for most of the construction practitioners. Taking web browser based user interface as an example. It can easily be implemented with the Hypertext Transfer Protocol (HTTP) via port 80. Request made by the users can be sent to the central server, which designated addresses of clients via IP. The analysis results are sent back to the clients and shown in the user interface through the same HTTP port.
- Internet: The computer network that uses the standard Internet protocol suite (TCP/IP) to serve as the communication channel. A web server that supports not only HTTP, but also server-side scripting using Active Server Pages (ASP), PHP, or other scripting languages;
- Central server: A Hadoop server (HDFS) that assumes the mass storage and simulation related computation. Specifically, we formulate the simulation tasks as MapReduce jobs and execute them in Hadoop, which is an open source implementation of the MapReduce framework (Dean & Ghemawat

2008). Unlike typical parallel computing models such as MPI, and General-Purpose GPU Computing, the MapReduce framework removes many obstacles to porting existing single-threaded codes onto parallel architectures. It hides the complexity of parallelization, data distribution, fault-tolerance, and load balancing from the developer. More importantly, the MapReduce framework is also well integrated with existing commercial cloud computing infrastructure. For instance, Amazon Elastic MapReduce (Amazon EMR) is a web service that makes it easy to quickly and cost-effectively process vast amounts of data. Amazon EMR uses Hadoop to distribute the data and processing across a resizable cluster of Amazon EC2 instances. The present information system utilizes the on-demand scalability and pay-per-usage model of Cloud Computing infrastructure for accelerating the analysis tasks. The Apache Hadoop project is an open-source implementation of MapReduce. It includes the Hadoop distributed file system (HDFS), designed for storing data files on a distributed network of computers, and Hadoop MapReduce, the parallel computation engine. Although Hadoop is written in Java, developers can write jobs in any other programming language using a utility called Hadoop Streaming. Hadoop Streaming implements Map and Reduce functions as interfaces to external user-specified applications. External Map and Reduce applications communicate with Hadoop Streaming through standard Unix streams. They read input key-value pairs via standard input and write back their output via standard output.

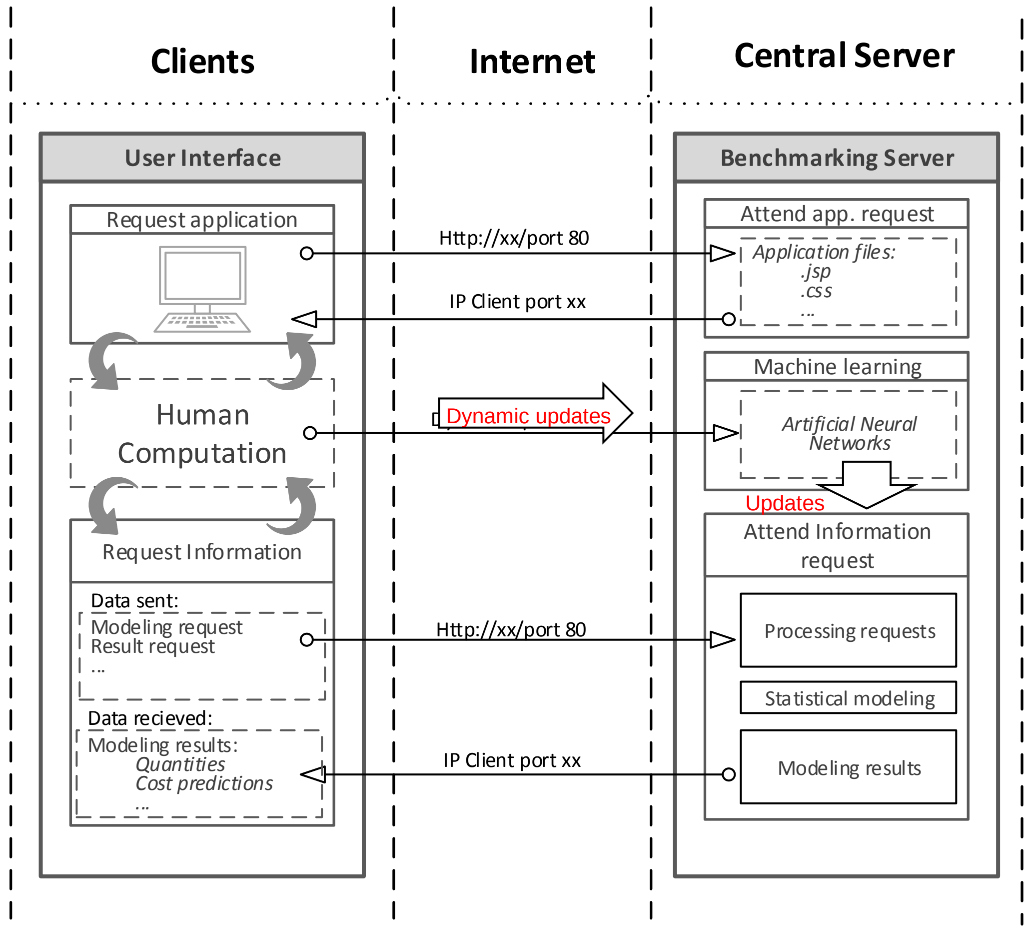


FIG. 6: Configuration of the proposed information system

The first step in implementing this information system is to set up the application, i.e., configuring the client for executing the necessary application files. Although the main computation is done on the remote HDFS server, some application files need to be set up in the client to allow local operations, such as determining the crafts of interest. One solution is through web browser application. Users access the HCBL server through the URL of *SaaS (Software as a Service)*, which sends the necessary files for realizing local user interface, such as JSP (Java Script Application), to the users' terminals. Port 80 is used to send requests to the HCBL server as a pure endpoint for the HTTP. It is compatible with MS Internet Explorers and therefore easier for the client side.

Other ports, such as Port 20, can be used to send back application files. On the back end, the installed application files query the HCBL server to execute computation. On the front end, users have the ability to view and operate simulation. The best way to implement this kind of interactive applications is using the model-view-controller (MVC) pattern (Krasner & Pope 1988). MVC is a classic design pattern that depends on three major components: models for maintaining data, views for displaying purposes, and controllers for handling user initiated events (Krasner & Pope 1988).

It is also worth noting that the human computation results, i.e., updated expert opinions, will be dynamically fed back to the central server, where ANN models will be built to preserve the knowledge. Where it is possible, the newly built ANN models will be used in the next round human-computer communications to replace the models that solely built upon historical data. A feedback is established to strengthen the opinion generation and evaluation.

5 CASE STUDY

5.1 Overview

A tool named iPET (interactive parametric estimating tool) was developed based on the proposed paradigm, and was tested in the proposal development of a power plant project. The case study was coordinated by an EPC (engineering, procurement and construction) contractor in the U.S (denoted as company A). The project used was a combined cycle power plant project budgeted for more than 2 million direct labor hours, 328 million dollars and planned to finish in 35 months.

Conceptual estimating for this type of projects is complicated involving the consideration of many unique factors such as actual geographic conditions of the job site. However, for the selected project only limited information was available and the time frame for developing the proposal was very short. A series of face-to-face interviews with those in key roles at company A (including vice-president, department heads, coordinators and estimators) have identified 12 project parameters (Tab.1) for parametric estimating of 28 crafts (Tab.2). Finally 47 historical jobs dated from 1993 to 2010 were selected to build statistical models.

TABLE 1: Selected project parameters

No.	Symbol	Description	Type
1	MW	The megawatt reading of the project	Numeric
2	Type	The nature of the job	Nominal
3	ENG	Engineering company in the project	Nominal
4	Config	The mechanical equipment configuration of the project	Nominal
5	Equip	Total number of mechanical equipment	Numeric
6	Layout	Job site layout classification	Nominal
7	Stack	Measurement of center line (feet) of outside stacks in a multiple unit configuration	Numeric
8	Vendor	The vendor of steam turbine	Nominal
9	Status	Refers to whether the installation of a project is at an existing facility	Nominal
10	TowerCells	Total number of cooling tower cells	Numeric
11	MainRack	End-to-end length in linear feet of the main pipe rack connecting the multiple units	Numeric
12	RadiusSum	Linear feet measurement from a central location to the major work areas of control room, water treatment facility, storage tanks and cooling tower	Numeric

TABLE 2: Crafts used for testing developed tool

No.	Craft	Description
1	Concrete PFC	Volume of concrete pouring, finishing and curing
2	Embedded steel	Volume of embedded steel
3	Concrete formwork	Area size of concrete formwork
4	Concrete reinforcing	Rebar
5	Total steel	Weight of major structural steel erected on site
6	Steel grating	Area size of steel grating
7	Steel Handrail	Weight of steel used for handrails

No.	Craft	Description
8	Steel pipe racks	Weight of steel pipe racks
9	Complete piping	Total length of all main pipes
10	U/G piping	Length of underground pipes
11	Critical piping	Length of critical pipes
12	Main circulating water system	Length of circulating water system
13	Small bore pipe	Length of small bore pipe
14	Large bore pipe	Length of large bore pipe
15	Total cable	Total length of main cables
16	Cable 600V and above	Length of cables for 600V and above transmission
17	Cable below 600V	Length of cables for below 600V transmission
18	Control cable	Length of cables used for controlling system
19	Instrument cable	Length of cables used for instruments such as sensors
20	Cable tray	Length of cable tray
21	A/G conduit	Length of above ground conduit
22	U/G conduit w/civil	Length of underground conduit
23	U/G conduit concrete	Concrete used for underground conduit
24	Iso Non Seg Bus	Length of isolated non-segregated phase bus (power)
25	Lighting	Count of lights
26	Install instruments	Count of installed instruments
27	Instrumentation Tubing	Length of instrument tubes
28	Instrument Calibration	Count of calibrations

5.2 The Model

The model was built following the procedure as described in Fig.5 (b):

Preliminary statistical models: Regression and ANN models were developed for each of the 20 crafts based on 47 historical project information. The models were made into an Excel spreadsheet tool to demonstrate results to estimators (Fig.7).

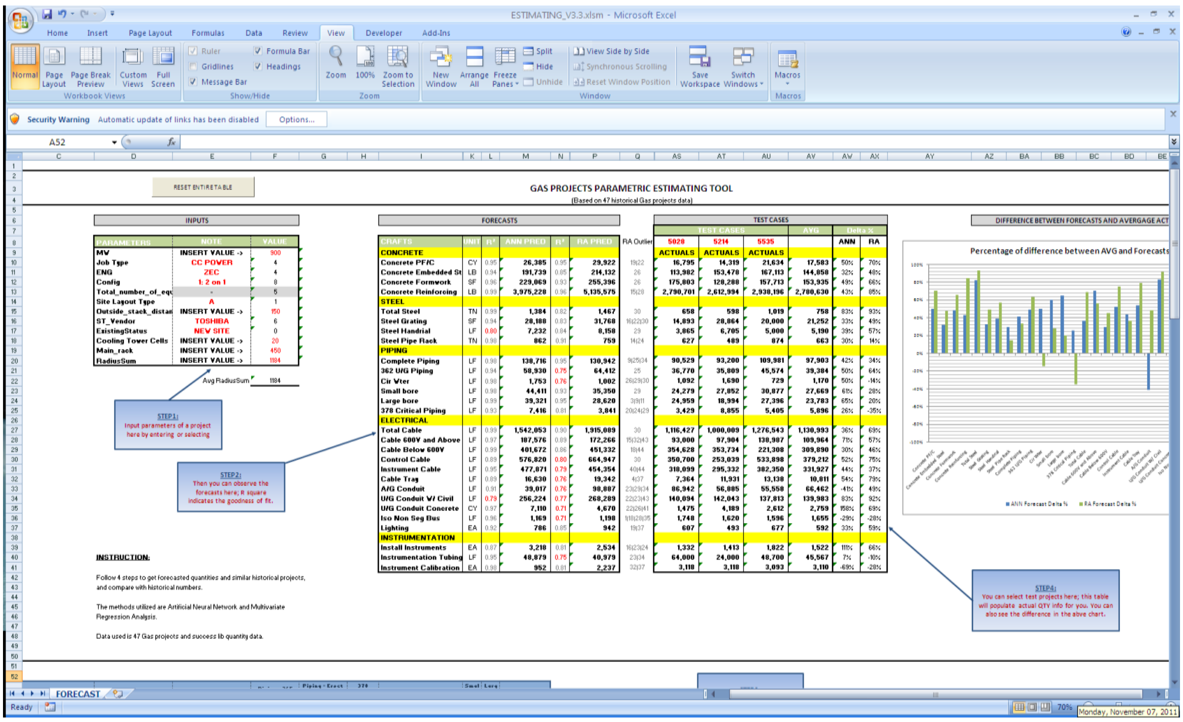


FIG. 7: Presenting statistical models to estimators using Excel Spreadsheet

Integrated demonstration: estimators make the first round decision on the basis of multiple sources of opinions instead of a single one, facilitated by the "bubble charts". As illustrated in Fig. 8, each bubble represents the forecasting result of a model. The height of the bubble shows the goodness of fit (R-square), and the size of the bubble represents the 90% confidence interval of the forecasting. The red dot shows the judgment made by the estimators. Estimators were able to handle this tool without much difficulty. Their judgments were tracked as the first round judgments.

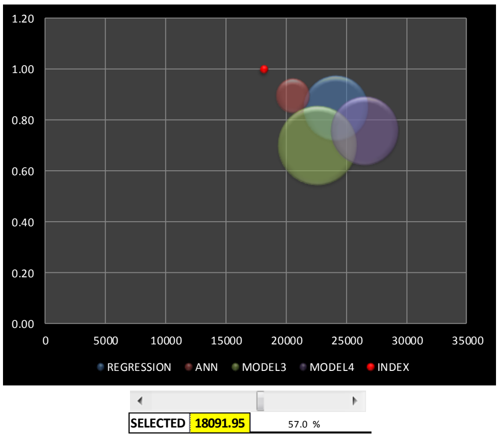
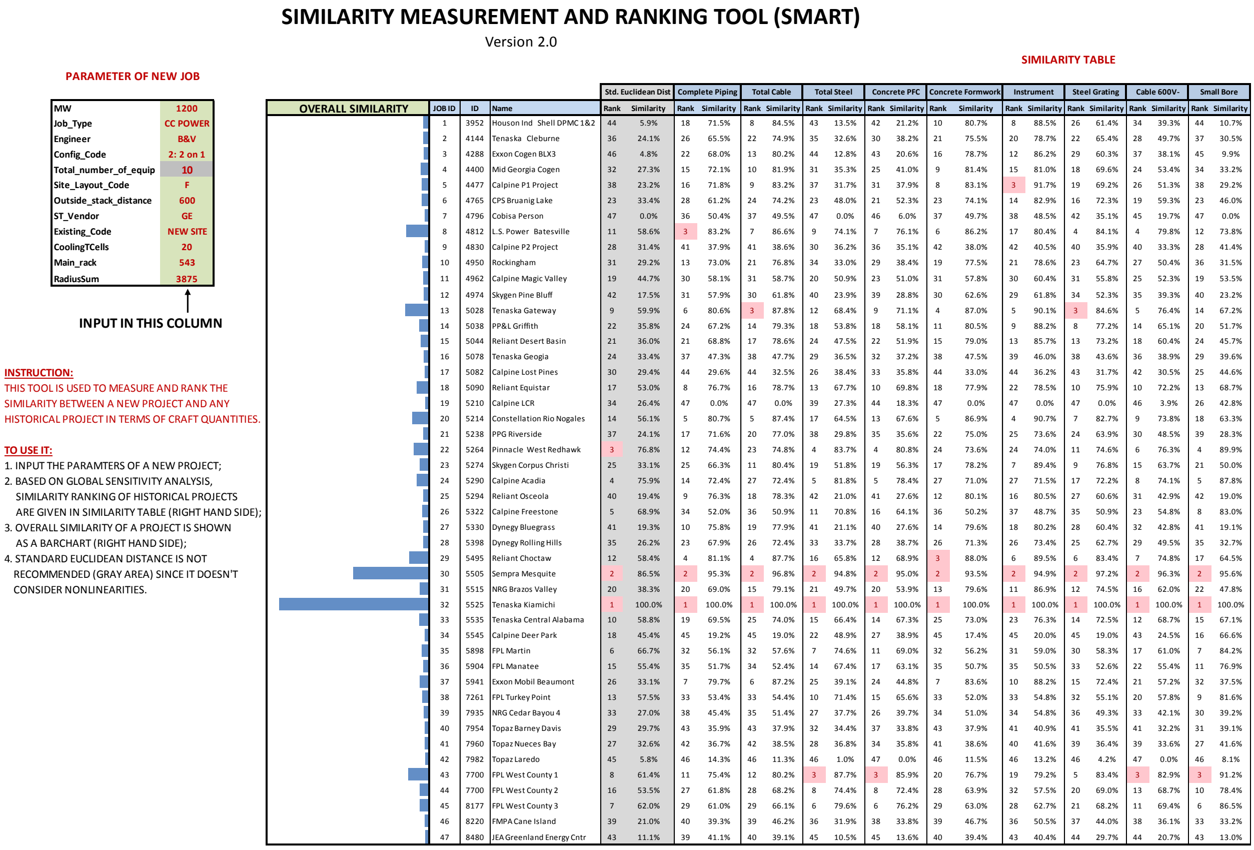


FIG. 8: Visualizing different model results with bubble charts

Validation: Then a graphical demonstration of the relationships between crafts (RBC) was presented to the estimators for their reference. In this demonstration, the quantitative relationship between any pair of craft, for example, concrete and steel usage, is shown as a regression formula. Estimators were able to evaluate if their first round estimates present reasonable ranges. Meanwhile, a tool developed to identify most similar historical projects (Du & Bormann 2014) was used (Fig.9), so estimators could also compare their estimates against historical experience. Estimators revised their judgments by considering all third-party resources;



SIMILARITY MEASUREMENT AND RANKING TOOL (SMART)

Version 2.0

PARAMETER OF NEW JOB

MW	1200
Job_Type	CC POWER
Engineer	B&V
Config_Code	2: 2 on 1
Total_number_of_equip	10
Site_Layout_Code	F
Outside_stack_distance	600
ST_Vendor	GE
Existing_Code	NEW SITE
CoolingTCells	20
Main_rack	543
RadiusSum	3875

INPUT IN THIS COLUMN

INSTRUCTION:
THIS TOOL IS USED TO MEASURE AND RANK THE SIMILARITY BETWEEN A NEW PROJECT AND ANY HISTORICAL PROJECT IN TERMS OF CRAFT QUANTITIES.

TO USE IT:
1. INPUT THE PARAMTERS OF A NEW PROJECT;
2. BASED ON GLOBAL SENSITIVITY ANALYSIS, SIMILARITY RANKING OF HISTORICAL PROJECTS ARE GIVEN IN SIMILARITY TABLE (RIGHT HAND SIDE);
3. OVERALL SIMILARITY OF A PROJECT IS SHOWN AS A BARCHART (RIGHT HAND SIDE);
4. STANDARD EUCLIDEAN DISTANCE IS NOT RECOMMENDED (GRAY AREA) SINCE IT DOESN'T CONSIDER NONLINEARITIES.

SIMILARITY TABLE

				Std. Euclidean Dist		Complete Piping		Total Cable		Total Steel		Concrete PFC		Concrete Formwork		Instrument		Steel Grating		Cable 600V-		Small Bore	
OVERALL SIMILARITY	JOB ID	ID	Name	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity
	1	3952	Houson Ind Shell DPMC 1&2	44	5.9%	18	71.5%	8	84.5%	43	13.5%	42	21.2%	10	80.7%	8	88.5%	26	61.4%	34	39.3%	44	10.7%
	2	4144	Tenaska Cleburne	36	24.1%	26	65.5%	22	74.9%	35	32.6%	30	38.2%	21	75.5%	20	78.7%	22	65.4%	28	49.7%	37	30.5%
	3	4288	Exxon Cogen BLX3	46	4.8%	22	68.0%	13	80.2%	44	12.8%	43	20.6%	16	78.7%	12	86.2%	29	60.3%	37	38.1%	45	9.9%
	4	4400	Mid Georgia Cogen	32	27.3%	15	72.1%	10	81.9%	31	35.3%	25	41.0%	9	81.4%	15	81.0%	18	69.6%	24	53.4%	34	33.2%
	5	4477	Calpine P1 Project	38	23.2%	16	71.8%	9	83.2%	37	31.7%	31	37.9%	8	83.1%	3	91.7%	19	69.2%	26	51.3%	38	29.2%
	6	4765	CPS Bruanig Lake	23	33.4%	28	61.2%	24	74.2%	23	48.0%	21	52.3%	23	74.1%	14	82.9%	16	72.3%	19	59.3%	23	46.0%
	7	4796	Cobisa Person	47	0.0%	36	50.4%	37	49.5%	47	0.0%	46	6.0%	37	49.7%	38	48.5%	42	35.1%	45	19.7%	47	0.0%
	8	4812	L.S. Power Batesville	11	58.6%	3	83.2%	7	86.6%	9	74.1%	7	76.1%	6	86.2%	17	80.4%	4	84.1%	4	79.8%	12	73.8%
	9	4830	Calpine P2 Project	28	31.4%	41	37.9%	41	38.6%	30	36.2%	36	35.1%	42	38.0%	42	40.5%	40	35.9%	40	33.3%	28	41.4%
	10	4950	Rockingham	31	29.2%	13	73.0%	21	76.8%	34	33.0%	29	38.4%	19	77.5%	21	78.6%	23	64.7%	27	50.4%	36	31.5%
	11	4962	Calpine Magic Valley	19	44.7%	30	58.1%	31	58.7%	20	50.9%	23	51.0%	31	57.8%	30	60.4%	31	55.8%	25	52.3%	19	53.5%
	12	4974	Skygen Pine Bluff	42	17.5%	31	57.9%	30	61.8%	40	23.9%	39	28.8%	30	62.6%	29	61.8%	34	52.3%	35	39.3%	40	23.2%
	13	5028	Tenaska Gateway	9	59.9%	6	80.6%	3	87.8%	12	68.4%	9	71.1%	4	87.0%	5	90.1%	3	84.6%	5	76.4%	14	67.2%
	14	5038	PP&L Griffith	22	35.8%	24	67.2%	14	79.3%	18	53.8%	18	58.1%	11	80.5%	9	88.2%	8	77.2%	14	65.1%	20	51.7%
	15	5044	Reliant Desert Basin	21	36.0%	21	68.8%	17	78.6%	24	47.5%	22	51.9%	15	79.0%	13	85.7%	13	73.2%	18	60.4%	24	45.7%
	16	5078	Tenaska Geogia	24	33.4%	37	47.3%	38	47.7%	29	36.5%	32	37.2%	38	47.5%	39	46.0%	38	43.6%	36	38.9%	29	39.6%
	17	5082	Calpine Lost Pines	30	29.4%	44	29.6%	44	32.5%	26	38.4%	33	35.8%	44	33.0%	44	36.2%	43	31.7%	42	30.5%	25	44.6%
	18	5090	Reliant Equistar	17	53.0%	8	76.7%	16	78.7%	13	67.7%	10	69.8%	18	77.9%	22	78.5%	10	75.9%	10	72.2%	13	68.7%
	19	5210	Calpine LCR	34	26.4%	47	0.0%	47	0.0%	39	27.3%	44	18.3%	47	0.0%	47	0.0%	47	0.0%	46	3.9%	26	42.8%
	20	5214	Constellation Rio Nogales	14	56.1%	5	80.7%	5	87.4%	17	64.5%	13	67.6%	5	86.9%	4	90.7%	7	82.7%	9	73.8%	18	63.3%
	21	5238	PPG Riverside	37	24.1%	17	71.6%	20	77.0%	38	29.8%	35	35.6%	22	75.0%	25	73.6%	24	63.9%	30	48.5%	39	28.3%
	22	5264	Pinnacle West Redhawk	3	76.8%	12	74.4%	23	74.8%	4	83.7%	4	80.8%	24	73.6%	24	74.0%	11	74.6%	6	76.3%	4	89.9%
	23	5274	Skygen Corpus Christi	25	33.1%	25	66.3%	11	80.4%	19	51.8%	19	56.3%	17	78.2%	7	89.4%	9	76.8%	15	63.7%	21	50.0%
	24	5290	Calpine Acadia	4	75.9%	14	72.4%	27	72.4%	5	81.8%	5	78.4%	27	71.0%	27	71.5%	17	72.2%	8	74.1%	5	87.8%
	25	5294	Reliant Osceola	40	19.4%	9	76.3%	18	78.3%	42	21.0%	41	27.6%	12	80.1%	16	80.5%	27	60.6%	31	42.9%	42	19.0%
	26	5322	Calpine Freestone	5	68.9%	34	52.0%	36	50.9%	11	70.8%	16	64.1%	36	50.2%	37	48.7%	35	50.9%	23	54.8%	8	83.0%
	27	5330	Dynegy Bluegrass	41	19.3%	10	75.8%	19	77.9%	41	21.1%	40	27.6%	14	79.6%	18	80.2%	28	60.4%	32	42.8%	41	19.1%
	28	5398	Dynegy Rolling Hills	35	26.2%	23	67.9%	26	72.4%	33	33.7%	28	38.7%	26	71.3%	26	73.4%	25	62.7%	29	49.5%	35	32.7%
	29	5495	Reliant Choctaw	12	58.4%	4	81.1%	4	87.7%	16	65.8%	12	68.9%	3	88.0%	6	89.5%	6	83.4%	7	74.8%	17	64.5%
	30	5505	Sempra Mesquite	2	86.5%	2	95.3%	2	96.8%	2	94.8%	2	95.0%	2	93.5%	2	94.9%	2	97.2%	2	96.3%	2	95.6%
	31	5515	NRG Brazos Valley	20	38.3%	20	69.0%	15	79.1%	21	49.7%	20	53.9%	13	79.6%	11	86.9%	12	74.5%	16	62.0%	22	47.8%
	32	5525	Tenaska Kiamichi	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%	1	100.0%
	33	5535	Tenaska Central Alabama	10	58.8%	19	69.5%	25	74.0%	15	66.4%	14	67.3%	25	73.0%	23	76.3%	14	72.5%	12	68.7%	15	67.1%
	34	5545	Calpine Deer Park	18	45.4%	45	19.2%	45	19.0%	22	48.9%	27	38.9%	45	17.4%	45	20.0%	45	19.0%	43	24.5%	16	66.6%
	35	5898	FPL Martin	6	66.7%	32	56.1%	32	57.6%	7	74.6%	11	69.0%	32	56.2%	31	59.0%	30	58.3%	17	61.0%	7	84.2%
	36	5904	FPL Manatee	15	55.4%	35	51.7%	34	52.4%	14	67.4%	17	63.1%	35	50.7%	35	50.5%	33	52.6%	22	55.4%	11	76.9%
	37	5941	Exxon Mobil Beaumont	26	33.1%	7	79.7%	6	87.2%	25	39.1%	24	44.8%	7	83.6%	10	88.2%	15	72.4%	21	57.2%	32	37.5%
	38	7261	FPL Turkey Point	13	57.5%	33	53.4%	33	54.4%	10	71.4%	15	65.6%	33	52.0%	33	54.8%	32	55.1%	20	57.8%	9	81.6%
	39	7935	NRG Cedar Bayou 4	33	27.0%	38	45.4%	35	51.4%	27	37.7%	26	39.7%	34	51.0%	34	54.8%	36	49.3%	33	42.1%	30	39.2%
	40	7954	Topaz Barney Davis	29	29.7%	43	35.9%	43	37.9%	32	34.4%	37	33.8%	43	37.9%	41	40.9%	41	35.5%	41	32.2%	31	39.1%
	41	7960	Topaz Nueces Bay	27	32.6%	42	36.7%	42	38.5%	28	36.8%	34	35.8%	41	38.6%	40	41.6%	39	36.4%	39	33.6%	27	41.6%
	42	7982	Topaz Laredo	45	5.8%	46	14.3%	46	11.3%	46	1.0%	47	0.0%	46	11.5%	46	13.2%	46	4.2%	47	0.0%	46	8.1%
	43	7700	FPL West County 1	8	61.4%	11	75.4%	12	80.2%	3	87.7%	3	85.9%	20	76.7%	19	79.2%	5	83.4%	3	82.9%	3	91.2%
	44	7700	FPL West County 2	16	53.5%	27	61.8%	28	68.2%	8	74.4%	8	72.4%	28	63.9%	32	57.5%	20	69.0%	13	68.7%	10	78.4%
	45	8177	FPL West County 3	7	62.0%	29	61.0%	29	66.1%	6	79.6%	6	76.2%	29	63.0%	28	62.7%	21	68.2%	11	69.4%	6	86.5%
	46	8220	FMPA Cane Island	39	21.0%	40	39.3%	39	46.2%	36	31.9%	38	33.8%	39	46.7%	36	50.5%	37	44.0%	38	36.1%	33	33.2%
	47	8480	JEA Greenland Energy Cntr	43	11.1%	39	41.1%	40	39.1%	45	10.5%	45	13.6%	40	39.4%	43	40.4%	44	29.7%	44	20.7%	43	13.0%

FIG. 9: Similarity Measure And Ranking Tool (SMART)

Optimization: Then a Sensitivity Analysis was conducted with a consideration on three controllable factors: cost escalation, risk mitigation options, and distribution of productivity. What if scenario analysis was also performed to evaluate different scenarios, for example, a different design of HRSG (heat recovery steam generator) or different physical location of control rooms. Tornado charts were utilized to explore the optimal decisions;

Final decision discussion: a meeting was held where estimators from different disciplines discussed the opinions obtained so far from different sources, and worked to reach a final widely agreed decision which considers both historical data and uniqueness of the new project;

Learning: the final decision is fed into a series of ANN models for training. In that, the ANN models are capable of capturing the knowledge and experience of estimators.

5.3 The Results

The result shows that estimators continued improving their estimates based on the updated decision making process. Take "complete piping" for example, two statistical models were used to give the preliminary information, which was 138,716 LF (model 1) and 130,942 LF (model 2). Based on the preliminary information, estimators gave the 1st round judgment as 125,000 LF. Most estimators believed that statistical models were too conservative. But after validation by RBC and CBR, estimators slightly adjusted the estimate to 135,000 LF, as this project required more steel than what were expected, and piping has a strong correlation with steel usage. Furthermore, the optimization session revealed certain risk factors that were ignored, especially that the air cooling tower would be used in this project which was a relatively new technology to the company. In addition, sensitivity analysis revealed that piping was highly dependent on the distance between main racks, which in this project was difficult to reduce. After optimization session, estimators increased the estimate to 138,000 LF. As shown in Fig.10, estimators gradually adjusted their judgments to better capture the unique feature of the new project. The final estimate was then used to rebuild ANN models. It is expected that these new models outperform those solely based on historical data in evaluating new projects.

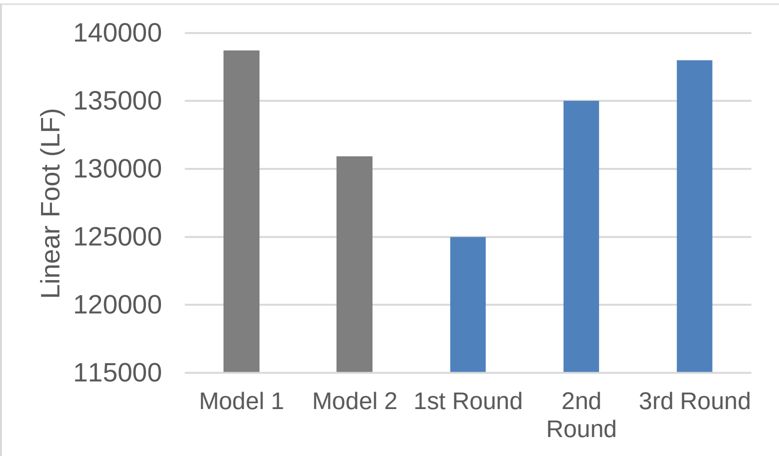


FIG. 10: Complete piping estimating example

Fig.11 illustrates the complete opinion evolution of 28 crafts. The black line (100% line) represents the predicted quantities by averaging model results. It is used as a base line in Fig.11. The first, second and third opinions given by the estimators are shown as "1st", "2nd" and "3rd" lines. As Fig.11 shows, in general, the estimators tended to give estimates lower than the model predictions in the first round, and increase the values after validation and optimization (second and third round). It is very similar to what has been reported by Mak and Raftery (1992), who found that estimators have systematic biases in estimating, especially shown as risk-seeking. The proposed paradigm improved the decision-making to certain extent by encouraging a more conservative estimating. And the opinion change reflects the impacts of dynamically updated information in the decision-making process of estimators. ANN models were built for each of the 28 crafts to learn the latest opinions of estimators (i.e., using the opinions of the estimators as the "training dataset" of the ANN models) that was able to preserve the valuable risk attitude shifts.

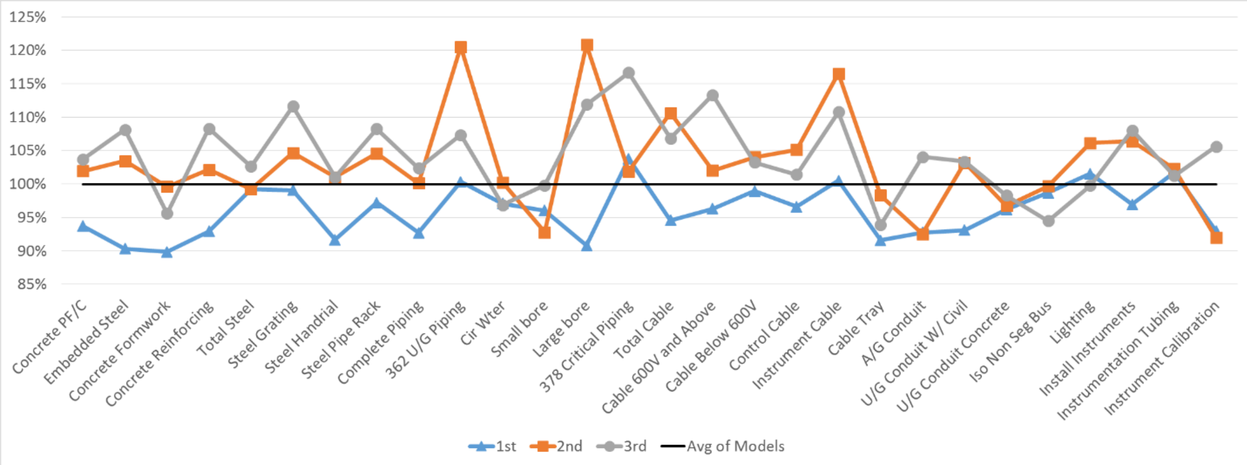


FIG. 11: Evolution of opinions

6 DISCUSSION

From the case study the readers might notice the difference between the final decision and the preliminary results of statistical models based on historical data. However, this doesn't mean that the final estimating results cannot reflect historical cases; on the contrary, final estimating results capture both historical pattern and estimators' perception to future conditions that are unique to this new project. Most estimators agreed that final numbers were more reasonable given many implicit and unknown contingencies. The most significant feature of the proposed paradigm is that it allows a seamless flow of information between human beings and computer models, while highlighting the role of estimators. There are three iterations of information exchange happening between estimators and computer; while in each one of them estimators have the right the revise and override the decisions gained from last round. Perceptions to future "unknown" conditions are highly valued, which makes better sense and allows greater flexibility. In addition, most estimators expressed concerns on relying on a single source of opinion, such as results of linear regression analysis. The proposed paradigm, on the other hand, combines the advantages of a variety of statistical algorithms which allow a consideration of probability and confidence intervals. Results of multiple statistical models, including both parametric and non-parametric models, are demonstrated in a single chart, so estimators can easily adjust their numbers with an overall consideration. Before final decisions, opinions from different disciplines/crafts were linked together and optimized using Sensitivity Analysis based simulation. This offers a unique perspective to the estimators to envision and refine their work as a whole. Finally on the basis of the final decision, the computer was able to remember the opinions of estimators, and evaluated and documented the implicit knowledge of the estimators, such as the proper risk attitude towards a particular situation. In that, knowledge is solicited and sustainable organizational learning is realized.

The proposed paradigm is not focused on improving the "accuracy" or "efficiency" of the conceptual estimating and as a result, traditional measures can hardly be used to evaluate if the implementation is a success. In contrast, the proposed paradigm aims to provide a better method to solicit human experience and to preserve knowledge, especially when situations are relatively new. It builds on the hypothesis that parametric modeling (i.e., statistical models used in many conceptual estimating) cannot work well in the face of deep uncertainties due to the lack of historical data. As a result, the success measure of the proposed paradigm is 1) *Is it able to solicit human opinions in a seamless way?* and 2) *Can the solicited knowledge be preserved for organizational learning?* To test the first criterion we provided Fig. 11 to show the opinion evolution of estimators on 28 estimated crafts. It shows that estimators in the three rounds of opinion expression gave results much different from the model predictions (black line). The largest difference was more than 20%. It is not our intention to evaluate which is more accurate – as in a brand new situation the only way to validate is to compare against the final actual cost. But Fig.11 clearly shows that the proposed method is able to solicit unique human opinions which differ from those of the parametric models. To test to second criterion, we evaluated the effectiveness of ANN in learning human opinions. Most of ANN models gave R-square (which is a good indicator to show the predictability of a model) bigger than 0.9. It suggests that ANN models are able to preserve knowledge. Given the above evaluation criteria, users should be able to evaluate the proposed paradigm.

There are several limitations of current proposed paradigm. A potential limitation is results of different statistical algorithms in preliminary analysis are not differentiated. Even though the goodness of fit indexes (e.g., R-square) are shown in the analysis process, estimators tended to ignore the difference of explanatory abilities between models but focused on the selection of the "middle" value. To overcome this issue different statistical algorithm should be prioritized according to their explanatory ability to historical data prior to rendering the results to estimators. The other weakness is the solicitation and preservation of estimators' knowledge is occurred only when final decisions are made. Discussions or opinions happened during the analysis process are not well recorded. Standard ways to document and formalize estimators' thinking flow during the conceptual estimating should be developed in the future. The third limitation, which is fundamental to the understanding of knowledge evolution, is the influences of posteriori knowledge on the perceptions of estimators to new projects. Estimating is basically a priori process which is mainly built on historical knowledge. Perception is obtained by considering the history and decisions are made within the existing empirical system. Posteriori knowledge (e.g., performance factors, a ratio of the actual construction costs versus budget) is used to update the empirical system, often shown as database updating. However, ignored by most, Posteriori knowledge is also an important feedback that can significant affect the perceptions of estimators (Fig. 9). For example, estimators always tend to adjust their risk attitudes towards the future projects based on lessons learned from previous projects, such as recent "unsuccessful bids", or increasing performance factors. Such adjustment of perception will in turn impact the performance of future bidding and estimating. This negative feedback remains happening until some sort of equilibrium is reached which probably is a signal for new dynamics. Neither traditional parametric estimating methodology nor the proposed paradigm considers such negative feedback loop between human perception and posteriori knowledge. Future work will pay special attention on calculating the influences of actual performance of new projects on human perception besides on database.

7 CONCLUSION

Making a fast and reliable conceptual estimating is extremely important for proposal development but it is nontrivial most of time due to the insufficient information and limited time window. Parametric cost estimating is a widely applied means to enhance conceptual estimating since it captures the patterns of historical projects and there is a big chance that these patterns remain suitable for future conditions. However, it is also noted that each project is unique in nature and future conditions of certain projects are sometimes unexpected. As a result experience and knowledge of estimators become important for conceptual estimating. Traditional parametric methodology, however, tends to isolate the involvement of estimators. Estimators can hardly input their opinions in the analysis process until results of statistical models are given. In addition, due to a different knowledge set, most emerging statistical models remain a "black box" for estimators and thus become impractical.

The authors found the key for a successful conceptual estimating to be combining human intelligence and artificial intelligence. To achieve this target it requires a significant change on current parametric estimating methodology which relies too much on CERs based on historical data. Inspired by the principals of human computation, this paper proposes an innovative paradigm named iPE (interactive parametric estimating). iPE facilities three iterations of information exchange between computers and estimators, and allows estimators to take charge of the entire estimating process. Assisted with a variety of parametric and non-parametric models, multiple opinions are considered and rendered to estimators, who can future revise the decisions based on experience, historical rules and global optimization strategies. On the end of the conceptual estimating, the opinions and implicit experience and knowledge of estimators are preserved using machine learning techniques, such as ANN as described in section 4.1. The ultimate goal is if we outline enough opinions for projects with different situations, the computer will know itself the proper estimates for new projects in the future. The interactive estimating tool (iPET) proves the applicability of the proposed paradigm. Better organizational learning is expected by a collaborative effort between human network and computer network.

This study differs from previous studies on conceptual estimating in three ways: First, for the first time it highlights the importance of the involving human knowledge in the conceptual estimating, in addition to the use of historical data. Previous studies have overemphasized the importance of data and correspondingly, the use of parametric models (Kwak & Watson 2005; Walton & Stevens 1997). These studies underestimated the deep uncertainties in most construction related problems which can hardly be addressed by historical data or statistical models. The present study urges a critical thinking about the traditional conceptual estimating methods and calls for more efforts to improve the active participation of human beings in the conceptual estimating. Second, for the first time it introduces the application of Human Computation in the conceptual estimating. A wide range literature search has been conducted by using the keywords of "*construction*" or "*engineering*" plus terms relevant to the use of Human Computation such as "*human computation*", "*crowdsourcing*", "*human-based computation*", "*human-assisted computation*", "*ubiquitous human computing*", "*distributed thinking*". To the best knowledge of the authors, there is no single publication addressing the application of Human Computation in the construction engineering and management areas. Given the increased complexity of modern construction projects, and the deep uncertainties construction decision-makers are facing, Human Computation provides a possible solution to a wide range of construction problems. Last, this study proposed a paradigm, framework and system that can be used to enhance knowledge transfer and organizational learning in construction organizations with conceptual estimating as an example. It has been identified by many literatures, such as (Chinowsky, Molenaar & Realph 2007), that the construction industry has an aging workforce, and the loss of corporate knowledge can be significantly greater since knowledge is not necessarily codified into standard practices. Previous conceptual estimating studies have been focused on the improving the accuracy of prediction models, but overlooked the necessity of preserving knowledge in the process of estimating. The proposed paradigm and system are expected to enable better organizational learning.

REFERENCES

Adeli, H 2001, 'Neural networks in civil engineering: 1989-2000', *Computer-Aided Civil and Infrastructure Engineering*, vol. 16, no. 2, pp. 126-142.

Adeli, H & Wu, M 1998, 'Regularization neural network for construction cost estimation', *Journal of Construction Engineering and Management*, vol. 124, p. 18.

Ahn, Lv 2005, 'Human Computation', PhD thesis, Cargegie Mellon University.

Alder, MA 2006, 'Comparing time and accuracy of building information modeling to on-screen takeoff for a quantity takeoff of a conceptual estimate', Citeseer.

Altayeb, SA 1997, 'Human factor versus the computer factor', *Cost Engineering(Morgantown, West Virginia)*, vol. 39, no. 11, pp. 15-16.

Anderson, DP 2004, 'Boinc: A system for public-resource computing and storage', in *Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on*, pp. 4-10.

ANSI. & PMI. 2008, *A Guide to the Project Management Body of Knowledge: PMBOK Guide (Fouth Edition)*, Project Management Institute, Inc.

Ashley, K 2006, 'Case-based reasoning', *Information Technology and Lawyers*, pp. 23-60.

Ayed, A 1998, 'Neural network model for parametric cost estimation of highway projects', *Journal of Construction Engineering and Management*, vol. 124, p. 210.

Bernstein, MS, Little, G, Miller, RC, Hartmann, B, Ackerman, MS, Karger, DR, Crowell, D & Panovich, K 2010, 'Soylent: a word processor with a crowd inside', in pp. 313-322.

Bowen, P & Edwards, P 1985, 'Cost modelling and price forecasting: practice and theory in perspective', *Construction Management and Economics*, vol. 3, no. 3, pp. 199-215.

Carr, RI 1989, 'Cost-estimating principles', *Journal of Construction Engineering and Management*, vol. 115, no. 4, pp. 545-551.

Chen, M, Hearsty, M, Hong, J & Lin, J 1999, 'Cha-cha: A system for organizing intranet search results', in pp. 5-5.

Chinowsky, P, Molenaar, K & Realph, A 2007, 'Learning organizations in construction', *Journal of Management in Engineering*, vol. 23, no. 1, pp. 27-34.

CII 2007, *Contracting to Appropriately Allocate Risk*, Construction Industry Institute.

Dean, J & Ghemawat, S 2008, 'MapReduce: simplified data processing on large clusters', *Communications of the ACM*, vol. 51, no. 1, pp. 107-113.

DP 2015, *Distributed Proofreaders Offcial Website*, viewed March 2 2015, <http://www.pgdp.net/c/>.

Du, J 2012, 'Investigation of Interpersonal Cooperation in Construction Project Teams: An Agent-Based Modeling Approach', PhD Disseration thesis, Michigan State University.

Du, J & Bormann, J 2014, 'Improved Similarity Measure in Case Based Reasoning with Global Sensitivity Analysis: An Example of Construction Quantity Estimating', *Journal of Computing in Civil Engineering*, vol. 28, no. 6, p. 04014020.

Du, J & El-Gafy, M 2010, 'Virtual Organizational Imitation for Construction Enterprises (VOICE): Managing business complexity using Agent Based Modeling', in *Construction Research Congress 2010*, Banff, Canada.

Du, J & El-Gafy, M 2011, 'Feasibility Analytical Mapping (FAM) for the Bidding Decision: A Graphic Bidding Decision Making Model Based on Multidimensional Scaling and Discriminant Analysis', *International Journal of Construction Education and Research*, vol. 7, no. 3, pp. 198-209.

—— 2012, 'Virtual Organizational Imitation for Construction Enterprises: Agent-Based Simulation Framework for Exploring Human and Organizational Implications in Construction Management', *Journal of Computing in Civil Engineering*, vol. 26, no. 3, pp. 282-297.

Du, J & El-Gafy, M 2014, 'Using Agent-Based Modeling to Investigate Goal Incongruence Issues in Proposal Development: Case Study of an EPC Project', *ASCE Journal of Management in Engineering*.

Du, J, El-Gafy, M & Ghanem, A 2012, 'Adding Value to Change Order Management Process Using Simulation Approach: A Case Study', in *48th Associated Schools of Construction (ASC) Annual International Conference*, Birmingham, UK.

Du, J, Liu, R & Karasulu, Y 2014, 'Cloud Based Interactive Probabilistic Simulation for AEC Industry', in ASCE (ed.), *ICCCBE 2014*, Orlando, FL.

Du, J & Wang, Q 2011, 'Exploring Reciprocal Influence between Individual Shopping Travel and Urban Form: Agent-Based Modeling Approach', *Journal of Urban Planning and Development*, vol. 137, no. 4, pp. 390-401.

Farrar, DE & Glauber, RR 1967, 'Multicollinearity in regression analysis: the problem revisited', *The Review of Economic and Statistics*, pp. 92-107.

Fellows, R & Liu, AMM 2000, 'Human dimensions in modelling prices of building projects', *Engineering Construction and Architectural Management*, vol. 7, no. 4, pp. 362-372.

Fu, W, Drew, D & Lo, H 2003, 'Competitiveness of inexperienced and experienced contractors in bidding', *Journal of Construction Engineering and Management*, vol. 129, p. 388.

Glaser, BG & Strauss, AL 1968, *The discovery of grounded theory*, Weidenfeld and Nicolson.

Globerson, S & Zwikael, O 2002, 'The impact of the project manager on project management planning processes', *Project Management Journal*, vol. 33, no. 3, pp. 58-64.

HARTONO, B 2010, 'Investigating risky decisions of construction contractors in competitive bid mark-ups'.

Howe, J 2008, *Crowdsourcing: How the power of the crowd is driving the future of business*, Century.

Hsu, K, Gupta, HV & Sorooshian, S 1995, 'Artificial neural network modeling of the rainfall-runoff process', *Water resources research*, vol. 31, no. 10, pp. 2517-2530.

ISPA, ISoPA 2008, *Parametric Estimating Handbook, 4th Edition*, <http://www.galorath.com/images/uploads/ISPA_PEH_4th_ed_Final.pdf>.

Khosrowshahi, F & Kaka, AP 1996, 'Estimation of project total cost and duration for housing projects in the UK', *Building and environment*, vol. 31, no. 4, pp. 375-383.

Kim, G, An, S & Kang, K 2004, 'Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning', *Building and environment*, vol. 39, no. 10, pp. 1235-1242.

Kim, G, Seo, D & Kang, K 2005, 'Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates', *Journal of Computing in Civil Engineering*, vol. 19, p. 208.

Knight, K & Fayek, AR 2000, 'A preliminary study of the factors affecting the cost escalation of construction projects', *Canadian Journal of Civil Engineering*, vol. 27, no. 1, pp. 73-83.

Kouskoulas, V & Koehn, E 1974, 'Predesign cost-estimation function for buildings', *Journal of the Construction Division*, vol. 100, no. 4, pp. 589-604.

Krasner, GE & Pope, ST 1988, 'A description of the model-view-controller user interface paradigm in the smalltalk-80 system', *Journal of object oriented programming*, vol. 1, no. 3, pp. 26-49.

Kwak, YH & Watson, RJ 2005, 'Conceptual estimating tool for technology-driven projects: exploring parametric estimating technique', *Technovation*, vol. 25, no. 12, pp. 1430-1436.

Leung, MY, Ng, ST & Skitmore, M 2005, 'Critical stressors influencing construction estimators in Hong Kong', *Construction Management and Economics*, vol. 23, no. 1, pp. 33-43.

Lintott, CJ, Schawinski, K, Slosar, A, Land, K, Bamford, S, Thomas, D, Raddick, MJ, Nichol, RC, Szalay, A & Andreescu, D 2008, 'Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey', *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1179-1189.

Lowe, D & Skitmore, M 1994, 'Experiential learning in cost estimating', *Construction Management and Economics*, vol. 12, no. 5, pp. 423-431.

Mak, S & Raftery, J 1992, 'Risk attitude and systematic bias in estimating and forecasting', *Construction Management and Economics*, vol. 10, no. 4, pp. 303-320.

Malone, TW, Laubacher, R & Dellarocas, C 2009, 'Harnessing crowds: Mapping the genome of collective intelligence', *MIT Center for Collective Intelligence.[www-document]< htt p://cci. mit. edu/publications/workingpapers. htm>[Retrieved June 10, 2009]*.

Martin Skitmore, R & Thomas Ng, S 2003, 'Forecast models for actual construction time and cost', *Building and environment*, vol. 38, no. 8, pp. 1075-1083.

Molenaar, KR 2005, 'Programmatic cost risk analysis for highway megaprojects', *Journal of Construction Engineering and Management*, vol. 131, p. 343.

NASA 1995, *Parametric Cost Estimating Handbook*, <http://cost.jsc.nasa.gov/pcehhtml/pceh.htm>.

—— 2015, *Stardust@home offcial website*, viewed March 1 2015, <http://www.nasa.gov/mission_pages/stardust/spacecraft/collector-index.html>.

Nieh, J, Yang, SJ & Novik, N 2000, 'A comparison of thin-client computing architectures'.

Parameswaran, M & Whinston, AB 2007, 'Social computing: An overview', *Communications of the Association for Information Systems*, vol. 19, no. 1, p. 37.

Perera, S & Watson, I 1998, 'Collaborative case-based estimating and design', *Advances in Engineering Software*, vol. 29, no. 10, pp. 801-808.

Polanyi, M 1961, 'Faith and reason', *The Journal of Religion*, vol. 41, no. 4, pp. 237-247.

—— 1962, *Personal knowledge: Towards a post-critical philosophy*, Routledge & Kegan Paul London.

Pyra, J & Trask, J 2002, 'Risk management post analysis: gauging the success of a simple strategy in a complex project', *Project Management Journal*, vol. 33, no. 2, pp. 41-48.

Quinn, AJ & Bederson, BB 2011, 'Human computation: a survey and taxonomy of a growing field', in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1403-1412.

Quinn, AJ & Bederson, BB 2011, 'Human computation: a survey and taxonomy of a growing field', in pp. 1403-1412.

Rush, C & Roy, R 2000, 'Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle', in *7th ISPE International Conference on Concurrent Engineering: Research and Applications, Lyon, France, July 17th-20th, Technomic Inc., Pennsylvania USA*, pp. 58-67.

Schexnayder, C, Knutson, K & Fente, J 2005, 'Describing a beta probability distribution function for construction simulation', *Journal of Construction Engineering and Management*, vol. 131, p. 221.

Shtub, A & Versano, R 1999, 'Estimating the cost of steel pipe bending, a comparison between neural networks and regression analysis', *International Journal of Production Economics*, vol. 62, no. 3, pp. 201-207.

Smith, AE & Mason, AK 1997, 'Cost estimation predictive modeling: Regression versus neural network', *Engineering economist*, vol. 42, pp. 137-162.

Steffey, RW & Anantatmula, VS 2011, 'International projects proposal analysis: risk assessment using radial maps', *Project Management Journal*, vol. 42, no. 3, pp. 62-74.

Subashini, S & Kavitha, V 2011, 'A survey on security issues in service delivery models of cloud computing', *Journal of Network and Computer Applications*, vol. 34, no. 1, pp. 1-11.

Tauber, AI 1997, *Science and the Quest for Reality*, New York University Press.

Trost, SM & Oberlender, GD 2003, 'Predicting accuracy of early cost estimates using factor analysis and multivariate regression'.

Von Ahn, L, Blum, M, Hopper, N & Langford, J 2003, 'CAPTCHA: Using hard AI problems for security', *Advances in Cryptology—EUROCRYPT 2003*, pp. 646-646.

Von Ahn, L, Maurer, B, McMillen, C, Abraham, D & Blum, M 2008, 'recaptcha: Human-based character recognition via web security measures', *Science*, vol. 321, no. 5895, p. 1465.

Walton, JR & Stevens, JD 1997, 'Improving conceptual estimating methods using historical cost data', *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1575, no. 1, pp. 127-131.

Wright, TP 1936, 'Learning curve', *Journal of the Aeronautical Sciences*, vol. 3, no. 1, pp. 122-128.