On the Edge of Human Computation: Interactive Parametric Estimating Paradigm

Jing Du[1] and Rui Liu[2]

[1] Assistant Professor, Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3053; email: jing.du@utsa.edu
[2] Assistant Professor, Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3054; email: rui.liu@utsa.edu

ABSTRACT

Conceptual estimating in the proposal development is nontrivial and continues to be challenged by quality and efficiency related issues. As a remedy, parametric estimating is widely used. Despite the merits obtained, human involvement is always missing from the decision-making process. Information can hardly flow between different sources of opinions. This paper introduces an innovative paradigm that combines the intelligence of human beings and computational ability of computers. It smoothing the information flow in the mixture of human-network and computer-network, solicits and conserves knowledge and experience of estimators in the proposal development. A tool based on the proposed paradigm is developed named iPET (Interactive Parametric Estimating Tool). iPET was tested in the proposal development of a power plant project. Result shows that the proposed paradigm enables fast and intelligent collaborative solution seeking between estimators and computers while highlighting the role of estimators. This work offers a unique perspective to the estimators to envision and refine their work.

INTRODUCTION

Conceptual estimating is a nontrivial task in the proposal development. Estimators need to yield the estimates under extreme time pressure and with very limited information. It always ends up with inaccurate estimates and unqualified budget which further increase the financial risk in construction execution phase (Carr, 1989). Since 1936, the use of parametric methods in conceptual estimating has become popular in order to enable faster and more accurate proposal development (Kwak and Watson, 2005). Methods or algorithms commonly used include regression analysis (Kouskoulas and Koehn, 1974) and neural network (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 (Du and Bormann, 2012). The feasibility and validity of these methods have been well proven by many case studies (Kim et al., 2004).

A closer examination, however, unveils certain essential problems with conceptual estimating, which can hardly be solved by current parametric estimating methods or CBR. First, human inference is missing from the implementation of most parametric estimating or CBR frameworks. Most parametric estimating approaches require advanced statistical or modeling knowledge in order to enable better handling of the models and proper interpretation of the results. 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 controlling the estimating procedure. This partially explains why many estimators hesitate to use the latest statistical approaches. Second, different algorithms or approaches are disconnected from each other. 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.

Regarding the estimating system as an information system, under current methodology it is difficult to integrate different algorithms or approaches, to realize smooth information flow between estimators and statistical models, and to conserve knowledge and experience of estimators which allows sustainable organizational learning. 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 (Ahn, 2005) paradigm provides a possible solution. This paper aims to develop a paradigm that combines the intelligence of human beings and computational capacity of computers, and smoothes the information flow in the mixture of human-network and computer-network in proposal development phase.

LITERATURE REVIEW

Human Dimension in Cost Estimating. Altayeb (1997) raised a reflective question at a time when computer models was becoming increasingly popular in cost estimating: *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, Altayeb found the additional efficiency from computers can be gained only if human beings are able to control the entire process (Altayeb, 1997). 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 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 and 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 and Liu, 2000). Especially when construction systems are regarded as socio-technical systems where human and organizational factors exert strong influences on the decision-makings (Du and El-Gafy, 2012), 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.

Human Computation. A widely accepted definition of Human Computation is given by Ahn (2005), which reads "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 (Ahn, 2005), language understanding (Bernstein et al., 2010), and human communication (Chen et al., 1999) etc. For example, in the reCAPTCHA project, 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 for humans to decipher, as part of normal website validation procedures (i.e., the validation words; see Figure 1). Then the results are returned to the reCAPTCHA service, which are integrated to the digitization projects.

Figure 1. The application of reCAPTCHA; retrieved from Google email registration webpage: http://mail.google.com (February 2014)

Scholars found roots of human computation in several existing theories including crowed sourcing (Howe, 2008), social computing, and collective intelligence (Malone et al., 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 and Bederson, 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 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 et al., 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.

THE PARADIGM: INTERACTIVE PARAMETRIC ESTIMATING

Architecture. Building on human computation theory, this paper proposes a paradigm for conceptual estimating and relevant organizational learning as follows. 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 conservation of knowledge. There are five key modules as shown in Figure 2: Human Skills, Cost Estimating Relationships (CER), Validation, Optimization, and Learning. 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 conserve 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 and 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 and 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.

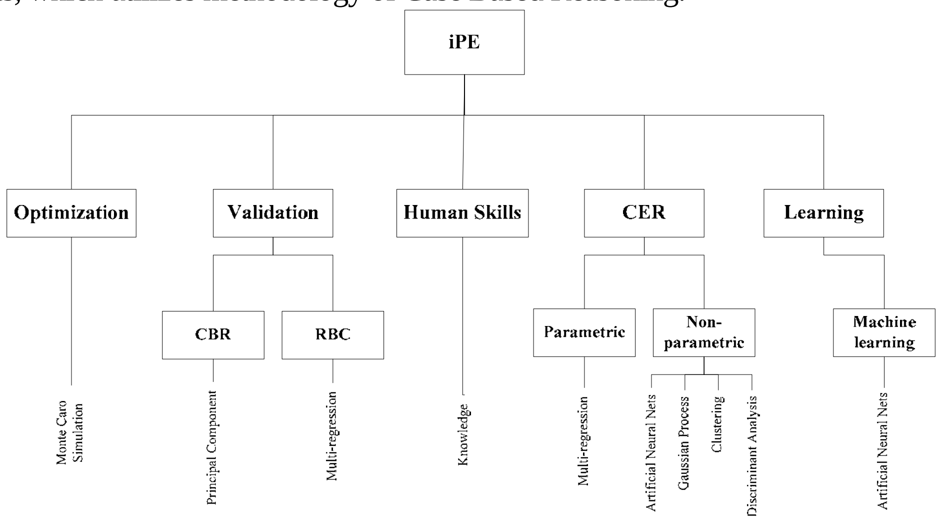
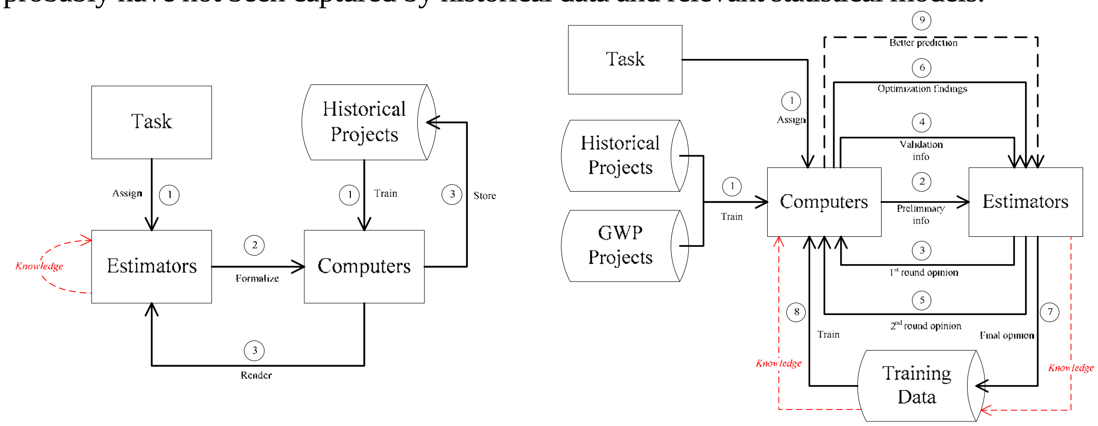


Figure 2. Architecture of proposed paradigm

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. In optimization module, Monte Carlo method is employed to discover the distribution of ultimate TIC with an overall consideration of the material price escalation, risk mitigation, probability distribution of productivity 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 conserve 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. 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 conserved for organizational learning purpose.

Analysis Flow. In a common parametric estimating practice (Figure 3 (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 et al., 2004). 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.



(a) Traditional Parametric Estimating (b) Interactive Parametric Estimating

Figure 3. The flowcharts of traditional and interactive parametric cost estimating

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. Figure 3 (b) describes the analysis flow of iPE:

(1) A variety of statistical models are built based on historical data and RBCs;
(2) Opinions from all the statistical models are demonstrated in one graph so the differences can be perceived by the estimators;
(3) Based on the preliminary opinions provided by statistical models, estimators make the first round decisions;
(4) The first round decisions 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 decisions based on the information obtained in step 4;
(6) The second round decisions from different disciplines are fed into the optimization module, where Monte Carlo simulation is conducted to investigate the optimum numbers for each craft;
(7) Estimators make their final decisions (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 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 decision and train the models for future use. A tool named iPET (interactive parametric estimating tool) is developed following the analysis flow of the proposed paradigm, and is tested in a power plant project proposal development. Limited by the space, details of iPET are not covered in this paper.

The Implementation of iPET. A prototype of iPET has been developed based on actual data collected from a general contractor Z. This tool then was applied to assist the proposal development of a power plant project. From a case study we noticed the difference between the final decision and the preliminary results of statistical models. Final estimating results capture both historical pattern and estimators' perception to future conditions which were unique to this new project. Most estimators agreed that final numbers were more reasonable for them considering many implicit and unknown contingencies. 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 that allow a consideration of probability and confidence intervals. Results of multiple statistical models were demonstrated in a single chart, so estimators could easily adjust their opinions with an overall consideration. The estimates were also compared to historical data to validate the reliability of estimating decisions. Before final decisions, opinions from different disciplines/crafts were integrated and optimized using Monte Carlo 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 system is able to remember the opinions of estimators, and evaluate and document 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.

CONCLUSION

Parametric cost estimating is a widely applied 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 projects. However, each project is unique in nature and future conditions are sometimes unexpected. Experience and knowledge of estimators are important for conceptual estimating. Traditional parametric methodology isolates 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. 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. Multiple opinions are considered and rendered to estimators, who future can revise the decisions based on experience, historical data and global optimization strategies. The interactive estimating tool (iPET) proves the applicability of the proposed paradigm.

REFERENCES

Adeli, H. (2001). "Neural networks in civil engineering: 1989-2000. " *Computer-Aided Civil and Infrastructure Engineering, 16*(2), 126-142.

Ahn, L. v. (2005). *Human Computation.* (PhD), Cargegie Mellon University.

Altayeb, S. A. (1997). "Human factor versus the computer factor. " *Cost Engineering(Morgantown, West Virginia), 39*(11), 15-16.

Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., . . . Panovich, K. (2010). *Soylent: a word processor with a crowd inside*.

Carr, R. I. (1989). "Cost-estimating principles. " *Journal of Construction Engineering and Management, 115*(4), 545-551.

Chen, M., Hearsty, M., Hong, J., & Lin, J. (1999). *Cha-cha: A system for organizing intranet search results*.

Du, J., & Bormann, J. (2012). "Improved Similarity Measure in Case Based Reasoning with Global Sensitivity Analysis: An Example of Construction Quantity Estimating. " *Journal of Computing in Civil Engineering, In Press*.

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, 7*(3), 198-209.

Du, J., & El-Gafy, M. (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, 26*(3), 282-297.

Fellows, R., & Liu, A. M. M. (2000). "Human dimensions in modelling prices of building projects. " *Engineering Construction and Architectural Management, 7*(4), 362-372.

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

Glaser, B. G., & Strauss, A. L. (1968). *The discovery of grounded theory*: Weidenfeld and Nicolson.

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

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, 39*(10), 1235-1242.

Kouskoulas, V., & Koehn, E. (1974). "Predesign cost-estimation function for buildings. " *Journal of the Construction Division, 100*(4), 589-604.

Kwak, Y. H., & Watson, R. J. (2005). "Conceptual estimating tool for technology-driven projects: exploring parametric estimating technique. " *Technovation, 25*(12), 1430-1436.

Malone, T. W., 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]*.

Polanyi, M. (1961). "Faith and reason. " *The Journal of Religion, 41*(4), 237-247.

Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy*: Routledge & Kegan Paul London.

Quinn, A. J., & Bederson, B. B. (2011). *Human computation: a survey and taxonomy of a growing field*.

Tauber, A. I. (1997). *Science and the Quest for Reality*: New York University Press.