A rating system for AEC e-bidding that accounts for rater credibility

Martin A EKSTRÖM
M.Sc., Ph.D. Candidate
Mekstrom@stanford.edu
Hans C BJÖRNSSON
Professor
hansbj@stanford.edu
CIFE, Stanford University
Terman Eng Cntr
Stanford, CA 94305-4020

Martin Ekström is s Ph.D. candidate at the Center for Integrated Facility Engineering (CIFE), Stanford University, whose research interests include rating systems and e-bidding for construction.

Hans Björnsson is a professor at CIFE whose research interest lies in the application of information technologies to construction industry process and business problems.

Summary

We believe that a rating system can substantially improve the effectiveness of an AEC ecommerce market since it can provide a means for market participants to share information about performance. One critical issue in the design of rating system is how to weight the ratings depending on who is the rater. Source credibility theory from the area of communication science offers a solution to this problem since it provides a means for user's to explicitly rate the reliability of a given rater. We also claim that it is important what type information that is available for assessing the raters' credibility. In this paper we propose an algorithm to calculate ratings of AEC subcontractors based on the user's different types of information about the credibility of the raters. Finally, we describe an experiment in progress to investigate the use of source credibility as a basis for a functioning AEC rating system.

Keywords

Rating system, e-commerce, bidding, construction industry, source credibility, trust, pairwise

1. Introduction

Since the Internet offers the opportunity for buyers and sellers to find new market partners at a low cost [4], the large and fragmented Architecture Engineering Construction (AEC) industry offers a potentially excellent arena for e-commerce. In April 2000 approximately $1 Billion had been invested in around 200 AEC e-commerce start-ups [5]. Still (January 2001) the amount of transactions conducted in these new e-market places remains very low. We argue that one reason for the slow adoption is that, if e-commerce is to be used for finding new market partners, parties must be certain that they can trust the participating organizations. AEC General Contractors, subcontractors, and suppliers are unlikely to do business with firms about which they do not know anything

The current practice among US general contractors to evaluate the potential bidders on a project is what one chief estimator described as a "vague" process. The estimator uses his judgment in combination with information from a variety of sources to evaluate the quality of potential subcontractors. Efforts to create web based formalized prequalification systems have been undertaken in, for example, both Norway [10] and Finland [25]. A prequalification system warrants that a market participant's goods or services are above a certain minimum standard but do not provide an accurate measure of their quality.

E-commerce has in general had a faster penetration into the consumer market than into the Business to Business (B2B) field. It is therefore natural that the earliest and most adopted commercial rating systems can be found in consumer-to-consumer (C2C) e-commerce. eBay [9] is the largest and most successful Internet Auction, enabling transactions between private parties. It currently (Jan 2001) has over 18.9 million registered users. These users buy and sell items that are varied, often difficult to describe, and cannot be easily evaluated, and trust is therefore a prerequisite for the transactions to take place. To foster trust, eBay has created a system by which the market participants rate each other after each transaction. The buyer will give the seller a good rating if he or she received the items in good condition [20]. A seller's aggregated positive and negative ratings are then added and displayed on the eBay web site. The advantage of eBay's rating system is that it is simple and intuitive to use. The problem is that all ratings weigh the same. It assumes that all raters are equally credible and trustworthy. This assumption may be valid in (C2C) auctions but we doubt that it will be for complex B2B transactions, such as AEC bidding. More sophisticated rating solutions have been presented in the Business to Consumer market comprising ratings of multiple criteria (e.g., Bizrate), and weighting of ratings depending on network of trust (e.g., epinions.com). In the B2B area openratings.com have presented a solution that relies on sophisticated statistical analysis of past performance and rating behavior to calculate an overall rating.[19]

2. Theoretical Foundations of the Study

There are several candidate methods that can be used for synthesizing uncertain information from sources of varying reliability. Zacharia et al [31]have proposed a complex rating mechanism (partly commercialized as Openratings.com described above) where the weight of the ratings is dependant on to what extent the user trusts the rater along with previous rating behavior. It does however equate trust in a business partner with trust in a rater. It is far from certain that just because a general contractor is be regarded as a trustworthy employer of subcontractors they are therefore also a credible rater. Fuzzy set theory offers a formalized manner with uncertain information and Zimmerman [32] and others [8] [27]) has applied fuzzy set theory to the credit rating processes conducted by banks, a process similar to an AEC rating problem. Another potential solution is subjective probabilities where the reliability of the sources can be modeled by adjusting the variance of the output function. Such a model could be operationalized using for example Howard's 5-step interview process [17]. A common feature of all the methodologies discussed above is that they work very well once you have the right input information. If you model and accurately measure the reliability of a source and how it interacts with the message (rating) the synthesizing of the information can be done in a consistent and unarbitrary manner. However, without the correct input they provide little value in themselves. We argue that the major problem in creating a functioning rating mechanism instead seem to be to accurately model and measure the reliability of the sources. Unless this can be done in a non-arbitrary manner the algorithms serving to calculate the overall rating add little value in a real world situation. One way to approach this problem is to take recourse in communication where source credibility theory has been developed to explicitly judge the credibility of media sources.

2.1 Source Credibility theory

The foundations to source credibility theory were laid by Hovland, Janis and Kelly [15], who identified the perceived trustworthiness and the perceived expertise as the main dimensions of a source's credibility. The higher the trustworthiness and expertise a source are perceived to have, the higher will be the importance given to information coming from that source. Early work in source credibility applied source credibility in the context of public opinion [15, 16, 6] and interpersonal

communication. [6]. Later studies has shown that source credibility applies also applies to commercial settings where the receiver of information is evaluating possible transactions. Examples can be found in areas such as advertising [30], evaluations of used cars [7], job offer acceptance [11], buy or lease decisions[13], job performance -evaluations [2], admitting students to business schools [14]and price-perceived risk relationships [12]. Our conclusion is that the task of a contractor evaluating the performance of a subcontractor share similar characteristics with the situations listed above, and we therefore hypothesize that source credibility can also be used to assign weights to different rating systems for AEC e-bidding.

2.2 Types of information available to judge rater credibility

Given that source credibility is used as the basis of a rating application the first step is to assess the credibility of the source, the rater. Based on interviews with potential users (AEC project managers and estimators) and previous research we have identified the following types of information that can be used to deduct rater credibility.

First Hand Knowledge- If the user personally know the rater they can themselves directly asses the credibility of this person.

2nd Hand Knowledge – If the user knows a credible person who states that the rater is credible, this is likely to positively influence the user's assessment of the rater's credibility [31, 26]. However, one cannot assume that trust is perfectly transitive (that you trust a friend of a friend as much as a friend) [1].

Credibility measure inferred from network– Another interesting measure of credibility of a rater is to measure how many other users find particular rater credible. Similar models have been proposed by other researchers [14, 31, 26].

Credibility of the Organization - Newhagen and Nass [23] point out the importance of differentiating between an organization and an individual when evaluating the credibility of a message source. Our interviews with industry practitioners show that this is certainly true for the construction industry. An estimator may regard a GC as an organization that is the definition of dishonesty and corrupt behavior but still trust those of the GC's employees that he/she knows on a personal basis.

Past ratings – Statistical analysis of past ratings can provide a measure of the expertise and trustworthiness of raters in the AEC e-commerce market. This solution has been proposed by Avery et al. [3] for an Internet market for consumer evaluations. They suggest that a statistical analysis will provide a means of monitoring the quality of subjective information such as ratings. The idea is that someone whose ratings deviate substantially from those of the rest of the market may not be honest or do not posses the expertise to rate properly. However, as Avery et al. point out, if you only penalize raters that are too far from the mean, this may discourage actors from posting idiosyncratic opinions and result in a universal "average" rating. In their article they suggest that raters should be penalized both for staying too close to and too far from the average. Furthermore, Zacharia [31] argues that it is reasonable that new users in the system should be assigned a low credibility/trustworthiness (they do not differentiate between the two factors).

2.3 The importance of time

Rater credibility is not the only factor that influences the weight of a rating. Zacharia et al [31] point out that time is another important factor. Based on interviews we have found that ratings that were more than 2 years old were regarded as substantially less credible than recent ones, even though the discount factors seemed to be highly individualized.

3. A proposed rating algorithm.

Based on the theoretical discussion above we propose the following 3-lever rating model that takes into account the credibility of the rater. Let us assume that user (i) seeks to find the rating for subcontractor (j) on criteria k. The user will base his/her decision on ratings from a set of raters (indexed m). In order to assess the credibility of the raters the user has different types of information (indexed l) such as information about past ratings, direct credibility assessments etc.

3.1 Calculating the overall rating

The calculation of the overall rating of a subcontractor on a given criterion, is a straightforward weighted average (similar formulas have been context of source credibility theory [28]) formula where the ratings are weighted by the credibility of the rater as well as time. Weighted averages have been

$$r_{ijk} = \sum r_{jkm} c_{im} \lambda^{-t(r_{jkm})} / \sum c_{im} \lambda_i^{-t(r_{jkm})} \quad (1)$$

r_{ijk}: aggregated rating on criteria k of contractor j from user i' s perspective
c_{im}: credibility of rater m
λ_i: Time Constant, discounting old ratings
$t(r_{jkm})$: age of rating

3.2 Calculating the credibility of a rater

In order to find the credibility of each rater we add the credibility assessments provided from the different types of information available to assess credibility.

$$c_{im} = \sum_l (w(x)_{il} X_{ilm} + w(tw)_{il} TW_{ilm}) \quad (2)$$

c_{im}: Credibility of rater m from user i' s perspective
X_{ilm}: Expertise of m from i' s perspective based on info type l.
TW_{ilm}: Trustworthiness of m from i' s perspective based on info type l.
$w(tw)_{il}, w(x)_{il}$: Weight coefficients

3.3 Assessment of weights in credibility calculations

The individual weights $w(x)_{il}$ in $w(x)_{il}$ in (2) indicate how important different types of information about Expertise (x) and Trustworthiness (tw) to user i when assessing credibility. To estimate these weights we use methodology of pairwise comparisons that have been widely adopted in decision-making tools that apply the Analytic Hierarchy Process (AHP)[29] (including applications for building systems [24]). Pairwise comparisons is an efficient method to assess preference based on the user's judgement of the relative importance of two different criteria [24]. To assesses the weight of N different criteria requires that for each pair (p,q) of criteria the user states (a_{pq}) how many times more important is p than q. Normally in AHP, an eigen-vector method is then used to calculate the weights of the different criteria allowing for inconsistencies among the pairwise comparisons. However, one obtains similar results using a least-square method [18], which has the advantage that it can be readily adapted to account for the added complexity of our 3-level model. In our model the user will first rate the credibility of N rater's before making pair-wise comparison's of each pair of raters by answering the question: *How much more important are ratings from rater p compared with ratings from rater q?*

We can then asses the respective weights for the different types of information by solving the straight forwards constrained optimization problem below.

Observe that this methodology is robust since it allows for the user's individual interpretation of the scale

$$\text{Min } Z = \sum_{r,p} (a_{p,q} \sum_l (w(x)_{il} X_{ilp} + w(tw)_{il} TW_{ilp}) - \sum_l (w(x)_{il} X_{ilq} + w(tw)_{il} TW_{ilq}) \quad (3)$$

$$\text{s.t.} \quad \sum w(x)_{il} + w(tw)_{il} = 1$$

X_{ilq}: Expertise of m from i' s perspective based on info type l., $w(tw)_{il}, w(x)_{il}$: Weight coefficients
TW_{ilq}: Trustworthiness of m from i' s perspective based on info type l., $a_{p,q}$: User i' s assessement of the importance of rater q' s ratings compared to rater p' s

4. Research methodology

We are currently (January, 01) building a prototype application that implements the propose rating algorithm and which will support an experiment described in this section. The experiment will simulate the user interface of an AEC e-market place where the participants (industry practitioners) will estimate the expected performance of a set of subcontractors based on peer ratings. The experiment will be done using data from a real AEC project where the GC has recently awarded the winning subcontractors. Our objective is that the experiment will test the following hypotheses.

- H1: Source Credibility can be used to assess the weights of ratings from different raters in the context of AEC-bidding
- H2: The assessed credibility of a rater will be positively correlated to the predictive accuracy of the rater's ratings.
- H3: The ratings calculated using weights based on source credibility leads to superior predictions compared to ratings where all ratings weigh the same (compare eBay).
- H4: The use of a rating tool using source credibility leads to superior predictions compared to the use of no tool in a time constrained setting.
- H5: The users will have more trust in ratings calculated with a rating tool using source credibility than their assessments using no tool.
- H6: The users will have more trust in ratings calculated with a rating tool using source credibility than their assessments using a tool where all ratings have the same weight.

By testing the validity of the first hypothesis (H1) we investigate whether source credibility can serve as basis weight calculation. H1 therefore serve to test the fundamental question whether source credibility is valid basis for an AEC rating system. The purpose of H2 is to see whether there are rational factors that motivate the difference in credibility judgments. Does the fact that someone is assessed to have a low credibility mean that that person is also a poor rater? The next two hypothesis (H3, H4) tests whether the proposed rating tool is superior when it comes to predicting the performance of subcontractors. The last two hypotheses (H5, H6) investigates whether the fact that a user's own preferences in terms of expected credibility have been used to weigh the ratings makes the ratings more reliable from the user's perspective. Observe that a rating tool can still be useful, even if it does not lead to superior predictions, as long as the user trusts these predictions enough to influence their decisions in positive manner. We hope to be able to deliver some preliminary results from this study at the ECCE-conference in June, 2001.

5. Discussion

The construction of an AEC rating system is a very complex task and we have therefore by focusing the scope of this research project left some interesting left some interesting issues to future research. One very important issue is how to provide incentives for raters contribute accurate ratings for the system. The proposed mechanisms serve as a filter against dishonest ratings but it does not in itself provide incentives for the raters to be honest. Another interesting question regards how to integrate ratings on multiple different criteria (e.g., how do you compare a credit rating with a rating for collaboration?). For the former issue research in game theory [21] can provide guidance whereas for the latter research in multi-criteria decision making treat similar problems (e.g. [17], [22]).

To conclude- we argue that this research project will contribute to research by 1) demonstrating to what extent source credibility theory can improve the rating process for AEC e-bidding, 2) providing a methodology to calculate ratings based on source credibility, and 3) formalising a way in which decision-makers can rate subcontractors in the AEC industry.

We also claim that this research project can provide contributions to the AEC industry. Firstly, because ii will give insight into how to construct weights of an industry rating system. Secondly, a

functioning rating system can create increased trust and therefore also more transaction in an e-market places

6. Bibliography

[1] A. Abdul-Rahman and S. Hailes, *Supporting Trust in Virtual Communities*, , *Hawaii's International Conference on Systems Sciences*, IEEE, Maui, Hawaii, 2000.

[2] M. D. Albright and P. E. Levy, *The Effects of Source Credibility and Performance Rating Disrepancy on Reactions to MultipleRaters*, Journal of Applid Social Psychology, 25 (1995), pp. 557-600.

[3] C. Avery, Reznick, P., Zeckhauser, R., *The Market for Evaluations*, American Economic Review, 89 (1999), pp. 564-584.

[4] Y. J. Bakos, *Reducing Buyer Search Costs: Implications for Electronic Commerce*, Managment Science, 43 (1997), pp. 1676-1992.

[5] C. Bazz, *Speach At CEE 320, Stanford University*, , 2000.

[6] D. K. Berlo, J. B. Lemert and R. Mertz, *Dimensions for evaluating the acceptability of message sources*, Public Opinion Quarterly, 33 (1969), pp. 536-576.

[7] M. H. Birnhaum and S. E. Stegner, *Source Credibility in Social Judgment: Bias, Expertise and the Judge's point of view*, Journal of Personality and Social Psychology, 37 (1979), pp. 48-74.

[8] L.-H. Chen and T.-W. Chiou, *A fuzzy credit-rating approach for commercial loans: a Taiwan Case*, Omega, International Journal of Management, 27 (1999), pp. 407-419.

[9] eBay, *About eBay*, , 2000.

[10] S. B. etat, *Godkjenningskatalogen - Godkjente foretak*, , Statens Bygningstekniske etat, 2000.

[11] C. D. Fisher, D. R. Ilgen and W. D. Hover, *Source Credibility, information favorability, and job offer acceptance*, Academy of Management Journal, 22 (1979), pp. 94-103.

[12] D. Grewal, J. Gotlieb and H. Marmorstein, *The Moderating Effects of Message Framing and Source Credibility On the Price-Perceived Risk Relationship*, Journal of Consumer Research, 21 (1994), pp. 145-153.

[13] R. R. Harmon and K. A. Coney, *The persuasive effects of source credibility in buy and lease situations*, Journal of Marketing Research, 19 (1982), pp. 255-260.

[14] M. Higgins, *Meta-information ; and time: Factors in human decision making*, Journal of the American Society For Information Science, 50 (1999), pp. 132-139.

[15] C. I. Hovland, I. L. Janis and H. H. Kelley, *Communication and Persuasion*, Yale University Press, New Haven, 1953.

[16] C. I. Hovland and W. Weiss, *The Influence of Source Credibility on Communication Effectiveness*, Public Opinion Quarterly (1951).

[17] R. A. a. J. E. M. Howard, ed., *Readings on the Principles and Applications of Decision Analysis*, Strategic Decisions Group, Menlo Park, 1984.

[18] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making - Methods and Applications*, Springer-Verlag, New York, 1987.

[19] isource, ***Are They Trustworthy?***, , *isourceonline*, 2000.

[20] L. F. Kaiser and M. Kaiser, *The Official eBay Guide*, Simon & Schuster, New York, 1999.

[21] J. Laffont, Tirole, J, *A theory of incentives in procurement and regulation*, MIT Press, Cambridge, 1993.

[22] F. A. Lootsma, *Multi-Criteria Decision Analysis via ratio and difference judgement*, Kluwer Academic Publishers, Boston, 1999.

[23] J. Newhagen and C. Nass, *Differential Criteria for Evaluating Credibility of Newspapers and TV News*, Journalism Quarterly, 66 (1989), pp. 277-284.

[24] G. A. Norris and H. E. Marshall, *Multiattribute Decision Analysis Method for Evaluating Buildings and Building Systems*, (1995).

[25] Ralacon, *Rakentamisen Laatu ry (RALA) : QUALIFICATION OF CONSTRUCTION ENTERPRISES IN FINLAND IN 2000*, , Ralacon, 2000.

[26] *A. R. Rohit Khare*, *Weaving the Web of Trust*, , Computer Science, California Institute of Technology, Pasadena, 1997.

[27] S. G. Romaniuk and L. O. Hall, *Decision making on creditworthiness using a fuzzy connectionist model*, Fuzzy Sets and Systems, 15-22 (1992).

[28] M. E. Rosenbaum and I. P. Levin, *Impression formation as a function of source credibility and order of presentation of contradictory information*, Journal of Personality and Social Psychology, 10 (1968), pp. 167-174.

[29] T. L. Saaty, *The Analytic Hierarchy Process, Planning, Priority, and Resource Allocation*, McGraw-Hill, New York, 1980.

[30] Vandenbergh, Soley and Reid, *Factor Study of Advertiser credibility*, Journalism Quarterly, 58 (1981).

[31] G. Zacharia, A. Moukas and P. Maes, *Collaborative Reputation Mechanisms in Electronic Marketplaces*, , *Thirty-second Annual Hawaii International Conference on System Sciences (HICSS-32)*, Wailea, Hawaii, 1999.

[32] H.-J. Zimmerman and P. Zysno, *Decision Evaluations by hierarchical aggregation of information*, Fuzzy Sets and Systems, 10 (1983), pp. 243-260.