2011

A Method to Estimate High-Dimensional Synergistic Interactions: A Case Study on Information Technology Business Value

Martin McGrane
University of Sydney, mmcgrane@it.usyd.edu.au

Simon K. Poon
University of Sydney, simon.poon@sydney.edu.au

A Method to Estimate High-Dimensional Synergistic Interactions: A Case Study on Information Technology Business Value

Martin McGrane
School of Information Technologies
University of Sydney
Sydney, Australia
Email: mmcgrane@it.usyd.edu.au

Simon K. Poon
School of Information Technologies
University of Sydney
Sydney, Australia
Email: simon.poon@sydney.edu.au

Abstract

In describing cause and effect relationships, difficulties may arise when two or more factors act as causes of a particular outcome. The complication results from the possibility that any one or more of the factors modify the extent to which others provide an effect. Such complex multi-factor interactions imply that multiple causal chains have at least some part in common; thus the evaluation of the synergistic effect of two or more causes is pertinent to the study of the causal mechanisms involved. The aim of this paper is to propose a new approach to systematically analyse combinations of interacting causal factors that might lead to good outcomes. Our approach was demonstrated on data from a highly complex field; a large dataset about Information Technology impact on business value collected by the Australian Department of Communication, IT and Art. Experimental evaluation confirms that this approach is able to statistically estimate the magnitude of higher-order interactions from multiple causal factors. Hence, synergistic interactions can be hypothesised and tested between any number of factors.

Keywords

Statistical interactions, Synergistic effect, Complementarities, High Dimensional Interactions, Statistical Testing

INTRODUCTION & MOTIVATION

The importance of quantifying interrelatedness of certain factors has been widely acknowledged in many research fields, e.g. biological systems (Moore et al., 2006), production economics (Topkis, 1978), engineering design (MacCormack et al., 2006), organizational structures (Rivkin and Siggelkow, 2007) and Management (Venkatraman, 1989). Complex systems produce their output through a combination of the individual effects of the components and the more complicated effects of interactions between components. Studying the combinations that have a stronger effect is an important step in understanding a system. Researchers in quantitative methods have long been interested in investigating synergistic (or complementary) effects among multiple interrelated factors. While this problem is being investigated in many fields, there is a need for a more efficient approach to measure and statistically validate patterns of interaction.

In such context, a group of activities (or factors) is complementary if doing more of any subset of those activities (or factors) increases the chance of achieving a goal than doing more of any subset of the remaining activities. The analysis of interactions can provide a way to capture the intuitive ideas of complementarities, i.e. "the whole is more than the sum of its parts." Techniques exist to measure the strengths of some forms of interaction but they are the subject of much debate and in some cases are incapable of measuring interaction in the presence of strong individual effects (Cordell, 2009; Vanderweele, 2010). These measures are also not designed to cope with testing for high dimensional complex interactions. There is a need for new approaches that quantify synergy statistically and do so efficiently even in high dimensional data.

Examining the effects of independent factors can be achieved by examining the effect in isolation. This is common practice as by restricting the variables examined to just one, one may be sure that variation is due to that factor. Examination of interaction between multiple factors demands the presence of those multiple factors by its very nature. The need to consider multiple factors can complicate analysis of data. A two-way interaction in a system of two variables is more unwieldy than a single variable but is likely to still be quite manageable. However, the addition of more variables rapidly increases the complexity of an analysis to impractical levels.

The challenge to studying interaction effects is first that more data are needed than for simple marginal effects, as the degrees of freedom increase with each level of interaction added (Cordell, 2009). Second, if one is looking to find an unknown interaction, all possible interactions must be examined (Freitas, 2001). It is the nature of interactions that they cannot be predicted from their parts or from more complex combinations. Therefore, the commonly used stepwise statistical approach to model interactions is inadequate to handle complexity as number of dimensions increase.

Analysing these interactions is an important step in quantitatively analysing the effects of causal factors on outcomes. Furthermore, the concept of interactions has not been adequately formulated in analysing the effects due to a combination of the ambiguity of "interaction" and the need to go beyond the traditional quantitative methods for analysing data. One critical research challenge in studying interactions is the lack of analytical tools to model interactions amongst multiple causal factors on outcomes. It requires not only systematic identification of potential interacting factors from data, but also the statistical interpretations of interactions amongst multiple factors.

In this paper, an analytical method of interaction is presented that is useful for analysing interactions between any number of factors. A statistical test is developed to measure the reliability of the patterns of interactions hypothesised. The method presented is sufficient to handle high dimensional interactions.

LITERATURE

The study of interactions is by no means a new endeavour. Hayman and Mather (1955) commented on the importance of interaction between genotype and environment. The father of modern statistical science, Fisher (1942) published work on the design of experiments to investigate simple interactions between two study factors. Interactions between drugs have been relied on in medicine for even longer. The Huangdi Neijing text, first developed around 320BCE, described treatments that included the use of combinations of drugs that relied on their interactions for pharmaceutical effect (Qibo, 320BC).

Interaction has been defined in various ways. At root, all definitions describe a difference from the expected value due to a combination of parts, given knowledge only of the parts. Beyond this, the various definitions can differ greatly. Here we shall describe the general understanding of how to measure interaction, and explore in more depth several of the more well used approaches and those of particular relevance to our interest in the synergistic effects from multiple causal factors.

Defining Interactions

Study of interactions in data at its simplest requires a definition of an interaction. This will vary in detail, and depending on what is being measured, but will depend in some way on a difference from the expected independent effect. Cole and Frangakis (2009) argued that consistency was overlooked by this definition and explored the importance of consistency to causal inference. The complexity of these factors can easily cloud the existence of interaction, as other factors confound identification of weaker effects, and even strong effects when the data is sufficiently unclear (Greenland, 2009). An important concept in understanding interaction is the difference between statistical interaction and causal interaction. This difference has been carefully studied in the area of clinical epidemiology since 1974 (Rothman, 1974; Rothman et al., 2008). A statistical interaction is an interaction found in data using a statistical model. These interactions are not necessarily causal, and may even be specific to a particular statistical test. A causal interaction is a direct physical interaction between effects. This is usually referred to as biological interaction in epidemiology. Analysis of data for interaction will find statistical interactions only, unless the data collection methodology was carefully designed with causal interactions in mind.

Attempts to identify causal interactions are hindered by the need to also consider the nature of the interaction, its strength, and the parameters measured. Interactions may operate indirectly as well as directly. For statistical interaction it is enough to say that the presence of two factors leads to an increased result but the causal interaction may involve a complicated chain of factors, none of which have been measured in the known data. The concepts of moderators and mediators were developed to better understand the possible mechanisms of an interaction Baron and Kenny (1986). A moderator is a factor that affects the strength of the relation between other factors. A moderator is not required for the interaction between the other factors to occur. A mediator factor is required, it can explain the interaction. Venkatraman extended these concepts for use in strategy research Venkatraman (1989) Here systems were represented using the concept of "fit". This describes moderation as an interaction between two factors that predict a third while mediation involves a key intervening factor. These are extended with four more categories, matching, gestalt, profile deviation, and co-variation.

The concept of sufficient cause was developed to better understand the division between statistical and causal interactions. As explained by VanderWeele (2009) it is possible to infer sufficient cause interactions from measured statistical interactions in some cases. This approach seeks to identify factors that are sufficient to

achieve the desired outcome. This does not presume causal effect, but rather shows that given the data analysed the present of some factors is sufficient for the outcome to occur. Braumoeller and Goertz (2000) explored a similar concept with their study of necessary conditions. Rather than simple sufficiency, this method finds the factors that are required, i.e. without these factors the outcome is not observed.

The simultaneous research into interactions in multiple fields has led to debate over terminology, as well as differing methodologies. The term "epistasis" is widely used to describe interactions, but the meaning varies between fields (reviewed in depth by Cordell, 2002; Phillips, 2008). In the epidemiological field it refers to a statistical departure from independence in the data. In contrast, in genetics it refers to a modification the phenotype expressed due to the effects on one gene by other genes (Moore, 2003). The precise meaning can vary as much as that of "interaction" since the test used for departure for independence will vary. Epistasis is, in effect, equivalent to interaction but the lack of agreement in current usage confuses the issue as much as it clarifies. Hence, in this work the terms statistical interaction and causal interaction are used rather than epistasis. It can only be hoped that the enthusiasm of the proponents of epistasis leads to a greater acceptance of more precise definitions.

A further complication to defining interaction is that it refers to any deviation from independence. If one is interested in a particular subset then a more precise definition is needed. In this work the focus is on Synergism which refers to the combined effect is greater than the sum of the parts. While seemingly clear on the surface, there is still much variety in the application of the term "interactions" in the literature. Rothman (1974) used a precisely defined definition based on an interaction measure called Interaction Contrast (IC). However, it is also common to refer to Synergy as an increase over the largest part rather than larger than the combination of the parts (Dunne, 2010), and Blot and Day (1979) argued that it may be useful to use both definitions; to use the definitions of Synergy for any increase over the parts; and for the statistical interaction, an increase over that expected from the combination, depending on the context.

The theory of Complementarities is an example of the simpler measures of interaction. The original concept of complementarities was first developed by Edgeworth in 1881 (Edgeworth, 1881). It defines activities as complements if doing any one of them increases the returns of doing the others. This was later explored in greater depth by Bulow et al. (1985). Bulow extended the work to demonstrate the interactions between markets as well as between products within a market. The notion of complementarities discussed here can also be represented by the "supermodularity" of a function with respect to two or more complementary variables (Topkis, 1998). Supermodularity dictates that the sum of the increases in the value of a function when the levels of the complements are changed one at a time would be less than the increase in the function's value when the levels are changed simultaneously. Supermodularity is defined as existing when $f(x \wedge y) + f(x \vee y) \geq f(x) + f(y)$, i.e. the sum of applying a function to the component-wise maximum and component-wise minimum is greater than the sum of the parts. For linear functions this can clearly be mapped directly to the concepts of synergy and antagonism. The three concepts of interaction contrast, complementarities, and supermodularity are thus related, despite their different backgrounds.

Analysing Interactions

Statisticians have solved the problem of reliability by comparing the difference between expected and measured result to what might be expected by chance variation alone. To do this there is first a need to estimate how wide the possible range of measurements is given that there is no interaction. Given this it is now possible to compare to this range to determine how large is the deviation from expected. If it is unlikely to have occurred by chance then it can be treated as reliable, or statistically significant. This does not mean that the effect is true, but that it is likely to hold in future tests.

Several approaches to defining interactions have been taken in the past. A standard test in many areas is to fit a line through data using a parameter for each potential factor or combination of factors. The stronger an effect, the greater its influence on the line of best fit. Thus the interaction measure here is the size of a coefficient that changes the shape of the curve. A drawback of this method for an interaction test is that the results are sensitive to noise and the relative level of independent and interactive effects. Strong independent effects will be favoured over weak interactions as that will better describe the data overall. This is increasingly so for higher dimensions of interaction where the noise is increasing as the individual effects become smaller. The test may also be affected by skewed data. Clearly if the majority of the data relates to a particular attribute then the best fit for the dataset will describe that attribute rather than the attributes that occur rarely. For high dimensional datasets there is a final problem that the efficiency is too low to be practical.

In the field of Epidemiology, Rothman (1974) developed the concept of Interaction Contrast (IC) similar to the supermodular function described in Topkis (1998) in measuring interactions between risk factors in public health. This measure contrasts the effect of a factor in isolation to that when the factor is combined with another. Synergy is defined as positive values of IC.

$$IC = (R_{11} - R_{01}) - (R_{10} - R_{00}) \qquad (1)$$

This approach simplifies interactions between probabilities to the main factors. While effective for identifying strong effects it is increasingly inaccurate as the probabilities increase. This measure can easily be extended to higher dimensions, the simplification at its heart causes this measure to be increasingly inaccurate. Rothman (1976) formalised arguments in the epidemiology field on interaction by defining the terms synergy and antagonism using the concept of the Synergy Index (SI). This measure approximates the ratio of the measured combined risk to the predicted combined risk, as shown in Equation 2.

$$Synergy\ Index = \frac{R_{11} - R_{00}}{R_{01} - R_{00} + R_{10} - R_{00}} \qquad (2)$$

The Synergy Index (SI) compared the combined effect of two factors to that of the parts. A ratio of exactly 1 is seen when the factors are independent. A ratio of greater than 1 is synergistic. A ratio of less than 1 is antagonistic. This work also provided a method to approximate variances and co-variances of interactions. Foraita (2009) later extended Rothman's concept of a synergy index for use in gene-gene interactions in genetic studies. While sufficiently accurate for its use in epidemiology with low probabilities, it is not appropriate for high probabilities or sufficiently large studies. Hogan (1978) argued that this simplification of the probability model is too imprecise as it discarded the combined effect of the probabilities. Hogan (1978) argued that analysis using SI is simple for two way interactions but is less meaningful at higher dimensions and statistical analysis is increasingly complex.

The most commonly used statistical model fitting approach is Analysis of Variance (AoV or ANOVA). This technique can be constructed with coefficients for each factor and combination of factors. The coefficients are adjusted to reduce the total sum of squares of the errors for each point. While this model is quite flexible and can be adjusted to consider complex functions, for the study of interactions it is prone to errors due to noise that outweighs the interaction (Ganzach, 1998).

Jakulin and Bratko (2003) explored a related measure based on information gain to estimate the presence of interaction between two factors. The "interaction gain" model is designed to be easily applied in machine learning contexts. Rather than estimating interaction directly it uses the decrease in entropy to measure whether factors are interacting - lower entropy due to combining factors suggests an interaction effect. This approach lends itself to graphical depiction of interaction effects but is problematic in higher dimensions. Although the entropy measure is useful in itself for illustrating the expected interactions present in data, but it is limited to low dimensions and the interactions shown are limited by the generality of entropy measures.

A METHODOLOGY TO ANALYSE HIGH-DIMENSIONAL INTERACTIONS

Recall that the IC (1) measure contrasts the effect of a factor in isolation to that when the factor is combined with another. Synergy is defined as positive values of IC. This approach simplifies interactions between probabilities to the main factors. While effective for identifying strong effects it is increasingly inaccurate as the probabilities increase. Rothman's Synergy Index (SI) instead approximates the ratio of the measured combined risk to the predicted combined risk (2). Analysis using SI is simple for two way interactions but is less meaningful at higher dimensions (Hogan et al., 1978).

While these groups of interaction measures can be extended to higher dimensions, they suffer from an increasingly inaccurate result as either noise outweighs the interaction (AoV), or the simplifications of the model are exacerbated with each added dimension(IC & SI). Further, the existing models do not allow for the effect of the interactions between components. For the model explored in this thesis, the aim is not to find the best-fitting model for a dataset, but to find the interactions. It is more useful to apply a specific model, as for IC and SI, and test if the data fits. Testing the fit of a single model increases the power of our analysis at the cost of being focussed only on interaction, an acceptable cost give that that is our focus. Rather than describe the data generally the model attempts to answer the specific question, "is this an interaction?" IC and SI meet our need for a specific model, but do not accurately describe an interaction.

These measures are not adequate to the task of measuring interactions in arbitrary dimensions, but they do provide the starting point for the statistical method explained here. The new measure will meet the requirements of specificity to interaction, be unaffected by skewed data, and will not be prone to type 2 errors due to noise.

Defining High Dimensional Interactions

The basis of the IC formula can be derived from the inherent probabilities in binary data:

$$R_{00} = P_{00} \qquad (3)$$

$$R_{10} = 1 - (1 - P_{00})(1 - P_{10}) \qquad (4)$$

$$R_{01} = 1 - (1 - P_{00})(1 - P_{01}) \tag{5}$$

$$R_{11} = 1 - (1 - P_{00})(1 - P_{10})(1 - P_{10}) \tag{6}$$

The presence or absence of an attribute is denoted by 1 or 0, respectively. Similarly, a 'good' outcome is denoted and by 1 and 'bad' by 0. Thus the probability of a good outcome due to attribute 1 being present and attribute 2 being absent can be represented as $P(outcome = 1 | a_1 = 0, a_2 = 0) = P_{00}$.

Measured risks will always include the background probability, that present regardless of the presence of the attributes. Thus we use the concept of 'risks' to describe this separately to probability, i.e. $R(outcome = 1 | a_1 = 0, a_2 = 0) = R_{00}$.

Equation 6 can be restated in terms of R_{01}, R_{10}, and R_{00}. It can be seen that this is now approximately equivalent to the IC model already described.

$$R_{11} = R_{10} + R_{01} - R_{00} + \frac{(R_{10} - R_{00})(R_{01} - R_{00})}{(1 - R_{00})} \tag{7}$$

$$IC \approx R_{11} - R_{10} - R_{01} + R_{00} \tag{8}$$

As discussed earlier, extending the IC equation to higher dimensions is problematic as the simplifications increase in error with each dimension added. Here the errors introduced by simplification are avoided by allowing for the use of the model in any dimension. In our new model, the basic probability model is extended, taking into account interactions at each level of dimensionality. The final measurement is then the effective change in probability due to that combination of factors. This removes the effect of errors due to simplification. If we now rearrange this in terms of the change in probability due to the synergistic effect, we find Equation 9:

$$P_{11} = 1 - \frac{(1 - R_{11})(1 - R_{00})}{(1 - R_{01})(1 - R_{10})} \tag{9}$$

At this point the advantage of the new model is that it is an exact test of the behaviour it seeks to measure and is not affected by skewed data. This contrasts with tests based on analysis of variance which attempts to find the most significant effect, thus is affected by both the size of individual effects, and does so using a test that is affected by skewed data.

While interactions between two factors are important, they are not the only interactions possible. To be truly representative of interactions the test must be capable of analysing interactions involving any number of factors. Thus the measure presented must now be extended to handle interactions of any dimensionality. In higher dimensions, a combination of factors is affected not just by those factors but by interactions between sub-combinations within. If the effect of combining factors A, B, and C is measured, then the measurement is also affected by the effects of combining A and B, A and C, B and C.

To measure the interaction solely due to a particular combination the effects of interactions in lower dimensions must be accounted for. The 3-dimensional form is thus not just a measure of the 3-dimensional interaction but that minus the effects of two dimensional interactions. To model this, the changes in probability due to each sub-combination are included. Each combination is represented with a probability value in the same way as the individual factors are. To represent this effect the notation P_{011} is used for the effect of the interaction between two of the attributes.

The synergistic effect for a three way interaction (P_{011}) can thus be shown as:

$$P_{111} = 1 - \frac{(1 - R_{111})(1 - R_{001})(1 - R_{010})(1 - R_{100})}{(1 - R_{011})(1 - R_{101})(1 - R_{110})(1 - R_{000})} \tag{10}$$

To be able to find interactions in arbitrary dimensions, a general form of the synergistic measure is needed. The basis of the measure is the product of all the individual probabilities and the interaction probabilities. To be usable, it must be possible to rewrite this using the observed risks rather than underlying, and unmeasurable, probabilities. It can be seen that the two and three way forms follow a similar pattern. It can be predicted that the general form will follow the same pattern.

$$P_n = 1 - \prod_{i=0}^{n} \prod_{j=0}^{i} (1 - R_{a_{j,2_i,2_n}})^{(-1)^i} \tag{11}$$

where $a_{i,n} = \{a_1, a_2, ..., a_n | a_k = \{0,1\}; \sum a_k = i\}$, $a_n = a_{n,n}$, $a_{j,i,n} = element\ j\ of\ a_{i,n}$, and $P_n = P_{a_n}$

Test for Statistical Significance

The data used for an analysis can vary in quality. For multi-dimensional analysis a limiting factor is often that of sample size. This affects analysis since for the same number of rows in the data, as the dimensionality increases, the size of each cell (or stratum) decreases. As the number of rows in each cell decreases, the probability becomes less and less certain. Chance variation is increasingly powerful as there are fewer samples to average over, i.e. using this data results in over-fitting and loss of statistical power.

A statistical test is applied to the interaction rules to determine the likelihood that a measured outcome is the result of random variation or represents a synergistic interaction. A statistical test has the advantage of being less arbitrary than a cut-off based on value and will by its nature adjust for low quality data.

A Taylor series approximation is used to approximate the variance at any given point. Note that the general form of the Taylor series would also include correlation between the factors, but for the purposes of this method, independence is assumed. The test begins with independent factors then tests if that assumption holds.

The resulting generalised variance formula is thus:

$$V_n = \sum_{i=1}^{n} \sum_{j=0}^{i} V_{a_j,i,n} \left(\frac{1 - R_n}{1 - R_{a_j,i,n}} \right)^2 \tag{12}$$

The variance of the strong antagonism formula is a much simpler comparison of two means with the variance calculated as before. The final z value is then calculated and tested against the normal distribution.

Note that there are other tests may in future be shown to more accurately model the data. This model was chosen as it is sufficiently accurate. It is important to note the effect of false positives. A simple test of probability is measuring how often an effect will occur by chance. In a dataset with no interactions a p-value of 0.05 will return a false positive for every twenty possibilities. In a 128 column dataset there are ten million 4-way combinations. Two hundred thousand of these will be false positives. The α-value can be adjusted to reduce the amount of false positives but without knowing the number of interactions it cannot be known in advance what value will give a desired power. Also, power of the sample can also be calculated based in the number of interacting variables to be tested in the model. The choice of p-value does not affect the implementation or operation of this model, but must be carefully considered by the user when interpreting the results.

Data to be analysed consists of multiple samples, each recording the presence or absence of the factors to be tested and an outcome. For example, if the outcome was whether or not a combination of study factors was effective, the factors would be the presence or absence of each factor tested. It is possible to analyse more complex data but it must first be divided into an appropriate binary form for each factor and outcome. For example, ordinal (likert-scale) data can be converted into a number of binary variables.

The risks needed for the model's calculations are calculated from the row counts, e.g.

$$R_{01} = \frac{Number\ of\ rows\ with\ a_1 = 0,\ a_2 = 1\ and\ outcome = 1}{Total\ number\ of\ rows\ with\ a_1 = 0,\ a_2 = 1} \tag{13}$$

The risks determined from the data can simply be used in the formula derived above. In this work synergy is defined using the change in probability between measured outcome and predicted outcome. Synergistic interactions are positive while antagonistic are negative.

To evaluate the statistical method we first test against simulated data with known properties. Interactions are inserted into a random dataset then the algorithm is tested on how reliably it detects those interactions, and how they compare to the chance effects of the individual factors. The effect of varying interaction strength and individual factor strength is also evaluated. After establishing the utility of the algorithm, we move on to real world datasets to look for significant interactions. We next analyse a dataset evaluating the influence of information technology business strategies on business value.

CASE STUDY OF INFORMATION TECHNOLOGY BUSINESS VALUE DATA

In this section, we apply our method to analyse the interactions amongst multiple organisational practices on organisational business value. The main goal is to discover a set of fundamental issues that are critical to understanding the mechanisms by which organisational practices interact with each other in business value studies. In this context, the role of organisational practices is closely related to the knowledge capital of a firm and that would enhance organisational performance. In the field of economics, Milgrom and Roberts (1995) proposed the concept of a "web of complementarities" which marked a paradigmatic shift in conceptualising complex dynamics among organisational practices. The aim is to highlight a set of fundamental issues that are critical to understanding the mechanisms by which organisational practices interact with each other. Our approach can be shown to be appropriate to deal with the interlocking nature of organisational practices.

Data Description

The dataset used was originally collected by the Australian Department of Communication, IT and Arts in 2004 (Gregor et al., 2004). It is based on a questionnaire answered by 1050Australian organisations from different industry sectors and of different organisation sizes. It provides information about organisational practices firms used in the last 18 months as well as the benefits they gained. The firm-level questionnaire was developed based on a collection of previous research with a focus on organisational transformation and IT investments. The firms questioned rated their use of 11 organisational practices (listed in Table 1), and estimated the benefits they had gained from those practices overall (listed in Table 2). After data cleaning by removing records with incomplete data and IT resources with large missing values, a sample of 558 organisations was subject to analysis. In this analysis, we only focus on the business value was measured in terms of the informational benefits.

Table 1. Organisational Practices (as IT complementary resources)

Practices	Descriptions	Mean (sd)
Formal Project Management (X1)	The frequency of applying formal project management methodology	0.52(0.035)
Business Case (X2)	The frequency of developing business case	0.48(0.035)
Post Implementation Review (X3)	The frequency of having post implementation reviews	0.53(0.034)
Change Management (X4)	The frequency of employing external change management specialists	0.50(0.022)
ICT opportunism (X5)	The frequency of recognising and achieving significant additional benefits which were initially unanticipated	0.61(0.029)
ICT Skill Level (X6)	The frequency of achieving valuable increases in ICT skill level within the organisation	0.57(0.026)
Business Strategy Planning (X7)	The frequency of engaging in formal business strategic planning	0.54(0.025)
ICT Strategic Planning (X8)	The frequency of engaging in ICT strategic planning	0.56(0.029)
Industry Leadership (X9)	The frequency of seeking to be an industry leader in adopting new ICT	0.59(0.031)
Formal Contracting (X10)	The frequency of establishing formal contractual arrangements for ICT investments	0.53(0.031)
ICT Integration (X11)	The frequency of integrating new ICT into existing business processes across key functional areas	0.54(0.024)

Table 2. IT Business Value Dimensions

Strategic IT Business Value (Y1)	Strategic benefits include the ability to create competitive advantage, align business strategies to directly support organisational goals, provide new products or services, and improve relationships with customers
Informational IT Business Value (Y2)	Informational benefits include faster and easier access to internal and external information, more useful, accurate and reliable information, and increased flexibility for manipulation of content and format of information
Transactional IT Business Value (Y3)	Transactional benefits include operational and cost savings, supply chain management savings, staff cost savings, and improved business efficiency of employees, business processes and financial resources.

The 11 organisational practices were rated by management executives based on how often their organisation performed those practices, in the range of 1 (never) to 5 (always); these values were dichotomised as follows: 1; 2 → 0 and 3; 4; 5 → 1. The business value of the resources chosen by an organisation was rated on a scale of 1 (never achieving business value) to 10 (always achieving business value). These variables are to measure the level of use in each practice for their business. We dichotomised the values of each organisational variable into

binary values. These were dichotomised as 1-6 → 0 and 7-10 → 1. This is similar to several earlier business value impact studies (see Banker and Johnston, 1995) that have bypassed specifying scale measures and modelled the "use" versus "non-use" of an input as a predictor of firm performance. The justification of using a coarser measure is to broaden the impact of use to particularly reflect the intangible enabling potentials of inputs.

Analytical Results

In the 2-way analysis, 10 synergistic interactions were found statistical significant at 0.05 level. In 3-way analysis, 40 synergistic interactions were found statistical significant using p-value 0.05 level. In 4-way analysis, 13 synergistic interactions were found statistical significant using p-value 0.05 level. There is no statistical significant 5-way interactions found in the dataset. Due to the space limitation, only the results of 2-way and 4-way synergistic interactions are shown in Table 3.

Table 3. Results from 2-way and 4-way significant interactions

2-way interactions	Synergy	p-value	4-way interactions	Synergy	p-value
it,itskilldev	0.471	0.022	it,projman,buscase,industlead	0.807	0.029
projman,itskilldev	0.622	0.000	it,projman,industlead,formalcon	0.845	0.002
projman,busstrat	0.557	0.000	it,buscase,itstrat,industlead	0.928	0.000
buscase,itskilldev	0.542	0.001	it,buscase,industlead,formalcon	0.815	0.010
postimp,itskilldev	0.645	0.000	it,postimp,itstrat,industlead	0.931	0.000
itoppt,itskilldev	0.457	0.024	projman,buscase,busstrat,formalcon	0.957	0.000
itoppt,itstrat	0.414	0.024	projman,buscase,itstrat,formalcon	0.895	0.000
itoppt,itstrat	0.414	0.024	projman,buscase,itstrat,formalcon	0.895	0.000
itoppt,itintegrate	0.526	0.003	projman,postimp,itoppt,industlead	0.949	0.000
itskilldev,formalcon	0.693	0.000	projman,postimp,itstrat,formalcon	0.805	0.037
busstrat,itstrat	0.515	0.009	buscase,postimp,itstrat,formalcon	0.961	0.000
			buscase,itoppt,itstrat,formalcon	0.969	0.000
			postimp,itoppt,itstrat,formalcon	0.870	0.001
			itoppt,itstrat,industlead,formalcon	0.843	0.001

Comparison with interactions results from different dimensions show commonality in factors. These would be interesting points to analyse to discover the reasons for the shared response. In the 3-way and 4-way analyses we can now see strong synergistic interactions that were not detectable in the 2-way analysis. It is because the same factors may cause an antagonistic response between two factors that now take part in a synergistic 3-way or 4-way interaction. This illustrates the importance of higher order interactions to understanding a system. A simple low order model would not detect the strong synergies found in higher dimensions. Furthermore, the synergistic effect from each factor can be presented in an interaction graph. One can identify the importance of each factor based on the vertex degree. An example based on our results is illustrated in Figure 1.

The presence of a factor in multiple results is interesting as it suggests a common cause of synergy between the factors. This may suggest that some organisational factors in complex organisational configurations may play different roles in affecting organisational performance. Such organisational practices can be overlooked analytically but often play a non-trivial role in the organisational processes.

DISCUSSION & CONCLUSION

Complex systems are defined by the complex interactions between their components. Understanding the interactions is the key to understanding the system. It cannot be hoped to fully model a system without the ability to adequately analyse the interactions within. However, currently this ability is not sufficiently developed to cope with any but the simplest of interactions. While this black box approach can produce useful results, it fails to enrich our understanding of the workings of the system. Examining individual components can be rewarding, but once these low-hanging fruit have been picked a far more rigorous model to analyse data is needed.

(a) Frequency of participation (Size of node) (b) Magnitude of synergy (thickness of edge)
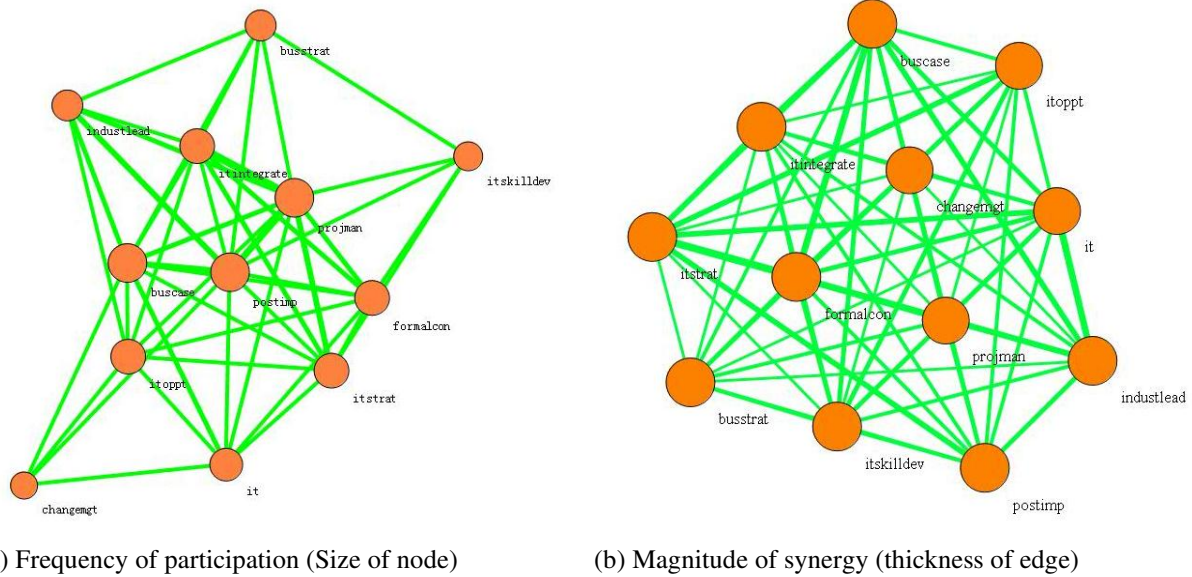
Figure 1. Interaction Graphs

Deeper analysis of interactions will aid a much richer understanding of complex systems in many and diverse fields. Understanding these complex systems is increasingly important as the simple discoveries based on individual factors are exhausted and scientific endeavour must delve deeper into ever more complex areas to gain more knowledge.

The method presented in this paper addresses these challenges to provide a useful measure of interaction effect. A clear definition has been laid out of the interactions the model measures in terms of changes in probability. How to determine which of the interactions in our data are reliable, thus more likely to represent 'true' causal interactions, has been described. The model designed in this paper is able to test for statistically significant interactions of any dimensionality.

The major limitation of this method is the reliance on dichotomous attributes. This limits the analysis of non-linear effects. If the interaction cannot be reduced to present/not-present then this model may itself be inadequate. The practical limitations of analysing such interaction are even greater, however. If the levels of each factor are represented by continuous curves, sufficient data for statistical significance must be present over enough of the curve to capture the interaction at every inflection point – without knowing where or how many of these points there are. Fortunately many systems can be reduced to dichotomous attributes. Medications may have complex non-linear interactions, yet in practice only be used at levels high enough to ensure an effect in most patients. The practical limitations thus benefit data analysis.

REFERENCES

Baron, R. and Kenny, D. 1986. "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *Journal of personality and social psychology* (51:6), pp. 1173–1182.

Blot, W.J. and Day, N.E. 1979. "Synergism and interaction: Are they equivalent?" *American Journal of Epidemiology* (110:1), pp. 99–100.

Braumoeller, B.F. and Goertz, G. 2000. "The Methodology of Necessary Conditions," *American Journal of Political Science* (44:4), pp. 844 – 858.

Bulow, J.I., Geanakoplos, J.D. and Klemperer, P.D. 1985. "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy* (93:3), pp. 488–511.

Cole, S.R. and Frangakis, C.E. 2009. "The Consistency Statement in Causal Inference: A Definition or an Assumption?" *Epidemiology* (20:1), pp. 132-148.

Cordell, H. J. 2002. "Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans," *Human Molecular Genetics* (11:20), pp. 2463–2468.

Cordell, H. J. 2009. "Detecting gene-gene interactions that underlie human diseases," *Nature reviews. Genetics* (10:6), pp. 392–404.

Dunne, R. 2010. "Synergy or antagonism-interactions between stressors on coral reefs," *Coral Reefs* (29:1), pp. 145–152.

Edgeworth, F Y. 1881. Mathematical Psychics: An essay on the application of mathematics to the moral sciences.

Fisher, RA. 1942. "The Theory of Confounding in Factorial Experiments in Relation to the Theory of Groups," *Annals of Eugenics* (11), pp. 341–353.

Foraita, R. 2009. "A conditional synergy index to assess biological interaction." *European Journal of Epidemiology* (24:9), pp. 485–494.

Ganzach, Y. 1998. "Nonlinearity, Multicollinearity and the Probability of Type II Error in Detecting Interaction," *Journal of Management* (24:5), pp. 615–622.

Greenland, S. 2009. "Interactions in Epidemiology: Relevance, Identification, and Estimation," *Epidemiology*, (20:1).

Gregor, S., Fernandez, W., Holtham, D., Martin, M., Stern, S., Vitale, M. and Pratt, G. 2004. "Achieving Value from ICT: Key Management Strategies," *Department of Communications, IT and the Arts, ICT Research Study*, Canberra.

B. I. Hayman and K. Mather. 1955. The Description of Genic Interactions in Continuous Variation. Biometrics, 11(1):69 – 82.

Hogan, M.D., Kupper, L.L., Most, B.M., and Haseman, J.K. 1978. "Alternatives To Rothman's Approach For Assessing Synergism (Or Antagonism) In Cohort Studies," *American Journal of Epidemiology* (108:1), pp.60–67.

Jakulin, A. and Bratko, I. 2003. "Analysing Attribute Dependencies," pages 229–240. Springer, Berlin.

MacCormack, A. Rusnak, J. and Baldwin, C.Y. 2006. "Exploring the Structure of Complex Software Designs: An Empirical Study of Open Source and Proprietary Code," *Management Science* (52:7), pp. 1015–1030.

Milgrom, P. and Roberts. J. 1995. "Complementarities and fit strategy, structure, and organizational change in manufacturing," *Journal of Accounting and Economics* (19:2-3), pp. 179–208.

Moore, J.H., Gilbert, J.C., Tsai, C-T, Chiang, F-T, Holden, T., Barney, N. and White, B.C. 2006. "A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility," *Journal of Theoretical Biology* (241:2), pp. 252–261.

Phillips, P.C. 2008. "Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems," *Nature reviews. Genetics* (9:11), pp. 855–67

Qibo. 320BC. Huangdi Neijing. First edition.

Rivkin, J.W. and Siggelkow, N. 2007. "Patterned Interactions in Complex Systems: Implications for Exploration," *Management Science* (53:7), pp. 1068–1085.

Rothman, K J. 1974. "Synergy and antagonism in cause-effect relationships," *American Journal of Epidemiology* (99:6), pp. 385–388.

Rothman, K J. 1976. "The Estimation of Synergy or Antagonism," *American Journal of Epidemiology* (103:5), pp. 506–511.

Rothman, K J, Greenland, S. and Last, T. 2008. *Modern Epidemiology.* PA: Lippincott Williams & Wilkins, Philadelphia, third edition.

Topkis, D.M. 1978. "Minimizing a Submodular Function on a Lattice," *Operations Research* (26:2), pp. 305–321.

Topkis, D.M. 1978. *Supermodularity and Complmentarity*. Princeton University Press.

VanderWeele, T.J. 2009. "Sufficient Cause Interactions and Statistical Interactions," *Epidemiology* (20:1).

Vanderweele, T.J. 2010. "Epistatic Interactions," *Statistical Applications in Genetics and Molecular Biology* (9:1).

Venkatraman, N. 1989. "The Concept of fit in strategy research," *Academy of Management Review*, (14:3), pp.423-444.

COPYRIGHT