Enhancing the Plausibility of Law Equation Discovery through Cross Check among Multiple Scale-Type Based Models

Takashi Washio and Hiroshi Motoda
washio@ar.sanken.osaka-u.ac.jp, motoda@ar.sanken.osaka-u.ac.jp
I.S.I.R., Osaka University,
8-1 Mihogaoka, Ibarakishi, Osaka 567-0047 Japan
Yuji Niwa
niwa@inss.co.jp
I.N.S.S. Inc.
64 Sata, Mihamacho, Mikatagun, Fukui 919-1205 Japan

Abstract

The study in the field of scientific discovery from data has been directed to the discovery of plausible law equations representing the first principles underlying objective systems. In this paper, a novel principle and an algorithm to predictively discover new scientific law equation formulae consisting of newly given quantities are proposed based on the candidate law equations governing the other quantities under current observation. The first principle based scientific law equation formulae must follow some mathematical admissibility and consistency. These conditions enable efficient reasoning of the law equation formulae in the prediction process. The soundness and the reproducibility of the equation prediction by this approach have been tested through numerical simulations of physical examples, and moreover its practicality has been confirmed through a real socio-psychological analysis. The approach can discover a set of scientific law equations representing common first principles under different set of quantities, and enables to capture general scietific features of the objective system under analysis.

1 Introduction

Many of the conventional approaches to identify numerical models from measurement data, *e.g.*, 'system identification theory' (Ljung 1987) and 'artificial neural network' (Wasserman 1989), derive an *asymptotic model* of an objective system over a narrow range of its state. Their plausibility is based on the assumption that the characteristics of the objective system over the state range can be sufficiently well captured by the presumed structure of the adopted equations such as linear and/or sigmoid formulae. However, this assumption usually does not hold over a wide range of states in the objective system because the presumed structure is merely an approximation within the narrow range. Accordingly the conventional approaches usually do not identify the law equations to represent the first principle governing the objective system over a wide state range.

In contrast, the main goal of scientific law equation discovery is to discover the first principle based law equations from measurement data. The most well known pioneering system to discover scientific law equations is 'BACON' (Langley et al. 1987). This system tries to figure out an invariant and its associated relation between two quantities over a wide state range by *bi-variate fitting* under a given laboratory experiment where some quantities are actively controlled. The found bi-variate relations are successively composed with the other relations, and finally equations representing the multiple measurement quantities are resulted. 'FAHRENHEIT' (Koehn and Zytkow, 1986) and 'ABACUS' (Falkenhainer and Michalski 1986) are successors that basically use similar algorithms to BACON to discover law equations. 'LAGRANGE' and 'LAGRAMGE' (Dzeroski and Todorovski 1995, Todorovski and Dzeroski 1997) are another type of scientific law equation discovery systems based on the ILP-like generate and test reasoning to discover equations representing the dynamics of the objects.

To reduce the ambiguity in their results under noisy measurements and the high computational cost of their algorithms, some subsequent discovery systems, *e.g.*, 'FAHRENHEIT', 'ABACUS' and 'COPER' (Kokar 1986), introduced the use of the unit dimension of physical quantities to prune the meaningless solutions. A problem of this approach is its narrow applicability only to the quantities whose units are clearly known. Our system named 'SDS' has overcome these difficulties (Washio and Motoda 1997, 1998). It discovers scientific law equations by limiting its search space to mathematically admissible equations in terms of the constraints of *scale-type* and *identity*. These constraints come from the basic characteristics of the quantities' definitions and the relations necessarily standing in the objective systems. The admissible equations discovered by SDS are considered to represent plausible relations among quantities reflecting the fundamental mechanisms governing the objective system. Since the knowledge of scale-types is widely obtained in various domains, SDS is applicable to non-physical domains including biology, sociology, and economics.

The framework has been further extended to the passively observed data where any active control on quantities are not admitted. Aforementioned FAHRENHEIT has a function to discover law equations from the observed data. LAGRANGE and LAGRAMGE can handle this task in their generate and test framework. More recently, SDS has been also extended to this task (Washio et al. 1999). Its excellent features of the robustness against observation noise and the limited computational complexity have been demonstrated.

In spite of these efforts, the state of the art is that the techniques have only succeeded in discovering the plausible candidates of law equations in a weak sense that the soundness, the reproducibility and the mathematical admissibility of the candidates hold within a given experimental environment. However, a law equation should hold over various objects and/or measurements sharing the common first principle. Accordingly, 'generality' of the candidate equations should be assessed under various environments so as to retain only highly plausible law equations in a strong sense.

More strictly speaking, two types of generality should be considered. One is the generality of the law equation over multiple objects sharing some identical first principle. Another is the generality over multiple combinations of measurement quantities. The first generality is explained through the following example of Kepler's third law,

$$\frac{T^2}{a^3} = C, \tag{1}$$

where T is the period of revolution, a the major radius of elliptic orbit, and C a constant. This relation between T and a holds not only for all planets in the solar system but also for every pair of mass points in a free space. Thus, it has the generality over multiple planets. Given an experimental or observation environment, the current law equation discovery systems can figure out the generality of this equation only limited to the given objects such as the planets in the solar system. The automatic generalization of the equation for wider domains such as every pair of mass points in a free space is beyond the scope of the current research field.

The second generality is demonstrated through the following irregular expression of Kepler's laws for any elliptic orbits having a certain eccentricity,

$$T^4 \dot{\omega}_a^3 = C', \tag{2}$$

where $\dot{\omega}_a$ is the angular velocity of the planet on the major axis of the elliptic orbit and C' a constant. Kepler initially discovered this type of law equations because the distance information on the planet orbits was hardly obtained. Later, he generalized Eq.(2) into Eq.(1) in concert with the following his second law,

$$\dot{S} = \frac{r^2 \dot{\omega}}{2} = C'', \tag{3}$$

where $\dot{S}$ is the areal velocity, r the distance between the sun and the planet, $\dot{\omega}$ the angular velocity of the planet and C'' a constant. Both Eq.(1) and Eq.(2) are the relations associated with the period of revolution under different combinations of measurement quantities. Because the consistency between the two equations are ensured by Eq.(3), these two equations are considered to represent a unique first principle and show the second generality.

The objectives of this paper are:

1. to propose a principle to reason some mathematically admissible and consistent law equation formulae for a newly given set of quantities based on the candidate law equation formulae discovered from another set of quantities and their measurement data in advance,

2. to propose an algorithm to predict mathematically admissible and consistent law equation formulae by the above principle and to check the plausibility of the candidate law equations in terms of the second generality, i.e., if the predicted formulae are well supported by the measurement of the newly given set of quantities, and

3. to evaluate and demonstrate the practicality of the proposed algorithm through a real world scientific law discovery in socio-psychology.

When two candidate law equations are independently discovered by a law equation discovery system from two different sets of measurement quantities Q_s and Q_t, their second generality can be mutually assessed by checking the consistency between the two candidates based on the proposed principle. However, the algorithm proposed in this study predicts law equation formulae for Q_t from the candidate law equation for Q_s, and check if the predicted formula explains the data of Q_t. This approach has the following advantages:

A. Amount of data and reasoning cost required to check the applicability of the predicted formulae to the measurement data of Q_t are much less compared with those needed to apply the law equation discovery system to the data of Q_t.

B. The applicability checking of the predicted formulae to Q_t is more robust against the noise than the case to derive candidate law equations for Q_t by using the law equation discovery system.

C. Complex but admissible law equations which may be missed by some conventional law equation discovery systems can be discovered.

These advantages are demonstrated through the performance evaluation and the practical application of the proposed approach in this paper.

2 Scale-type Constraints

The background theory of the proposing principle is provided by the scale-types of measurement quantities and the constraints on the admissible relations of pair wise quantities associated with the scale-types. The discussion on the scale-types was given by Stevens (1946). He defined the measurement process as 'the assignment of numerals to objects or events according to some rules.' He claimed that different kinds of scale-types and different kinds of measurement are derived if numerals can be assigned under different rules, and categorized the scale-types of quantities based on the operation rule of the assignment. He mathematically characterized and categorized quantitative quantities into two major scale-types of interval scale and ratio scale. Examples of the interval scale quantities are temperature in Celsius and sound tone where the origins of their scales are not absolute, and are changeable by human's definitions. Its admissible unit conversion follows *Generic linear group*: $x' = kx + c$. Examples of the ratio scale quantities are physical mass and absolute temperature where each has an absolute zero point. Its admissible unit conversion follows *Similarity group*: $x' = kx$.

insert table 1 and 2 here

Luce (1959) claimed that the basic formula of the functional relation among quantities of ratio and interval scales can be determined by their scale-type features, if the quantities have direct dependency without being coupled through any dimensionless quantities. Under this condition, some unit dimensions of the quantities are related to each other, and consequently the unit change of a quantity affects the value of other quantity. Suppose x_i and x_j are both ratio scale quantities, and x_i is defined by x_j through a logarithmic functional relation $x_i = u(x_j)$, *i.e.*, '$x_i = \log x_j$'. We multiply a positive constant k to x_j, *i.e.*, 'a change of unit', without violating the group structure of the ratio scale quantity x_j, then this leads $u(kx_j) = \log k + \log x_j$. This fact causes the shift of the origin of x_i by $\log k$, and violates the definition of x_i which is the ratio scale quantity. Hence, the direct functional relation from x_j to x_i must not be logarithmic. As the admissible transformations of x_i and x_j in their group structures are $x'_i = Kx_i$ and $x'_j = kx_j$ respectively, the generic formula of $x_i = u(x_j)$ must satisfy the invariant condition of $x'_i = u(x'_j) \leftrightarrow Kx_i = u(kx_j)$ under the unit conversion. The factor K of the changed unit of x_i depends on k, but it shall not depend upon x_j, so we denote it by $K(k)$. Consequently, we obtain the following constraints on the continuous function $u(x_j)$,

$$u(kx_j) = K(k)u(x_j),$$

where $k > 0$ and $K(k) > 0$ as these are the factors of the unit change. The constraints for all combinations of the scale types are summarized in table 1. Luce (1959) derived each solution of $u(x_j)$ under the condition of $x_j \geq 0$ and $u(x_j) \geq 0$. We have extended his theory to cover the negative values of x and $u(x_j)$ (Washio and Motoda 1996). The generic solution of $u(x_j)$ in each case is summarized in table 2. The impossibility of the definition of a ratio scale from an interval scale is because the ratio scale involves the information of an absolute origin, but the interval scale does not. In this table, the inverse functions of the cases 2.1 and 2.2 are listed at 3.1 and 3.2 for use in the algorithm shown in the next section.

3 Principle and Algorithm

Let Q_s be a source set of measurement quantities, and $\psi_s = 0$ a source equation where all of its arguments belong to Q_s. Furthermore, let δ_{ij} be an operator to commute a quantity x_i to another x_j in a set of quantities, and $\delta_{ij}\psi = 0$ an equation where the argument x_i is changed to x_j in $\psi = 0$ by substituting the relation $x_i = u(x_j)$. Our task is to derive a set of the admissible target equation $E_t = \{\psi_{tk} = 0, k = 1, ..., m\}$ from the source equation $\psi_s = 0$ where all arguments of each $\psi_{tk} = 0$ belong to a target set of measurement quantities Q_t, and m is the number of candidate equations. Q_t is derived from Q_s by applying a set of the commutation operators $\Delta_{st} = \{\delta_{ij} | x_i \in Q_s, x_j \in Q_t\}$, and thus the cardinality of Q_t is equal to that of Q_s.

If $\psi = 0$ and $x_i = u(x_j)$ is known in a priori, then the equation formula $\delta_{ij}\psi = 0$ is easily derived. However, our interest is to derive $\delta_{ij}\psi = 0$ when $x_i = u(x_j)$ is unknown in advance. In this paper, the situation, where the following two assumptions hold, is considered.

Assumption 1 *The scale types of the quantities for commutations are known.*

Assumption 2 *The quantities x_i and x_j have direct dependency without being coupled through any dimensionless quantities.*

The first assumption does not yield any strong limitations since the scale-types of measurement quantities are widely known (Washio and Motoda 1997). The second assumption holds, when x_i and x_j are the quantities to measure an identical feature through different processes and/or when they are known to have direct dependency based on the background knowledge in the domain as for the case of Kepler's laws. This type of quantity pairs are widely seen in various domains as shown later. When the two assumptions hold, some unit dimensions are shared by x_i and x_j, and thus the scale-type constraints indicated in table 2 can be applied. Starting from the source equation $\psi_s = 0$, the application of all operators $\delta_{ij} \in \Delta_{st}$ derives the target equation $\psi_{tk} = 0$. In each application of δ_{ij}, $x_i = u(x_j)$ is selected from table 2 based on the scale-types of x_i and x_j. Multiple solutions of $\psi_{tk} = 0$ may be derived since both two candidates of $x_i = u(x_j)$ are applied in case that x_i and x_j are the pair of interval and ratio scale quantities. Accordingly, these commutation operations may result in a set of target equations E_t. In case of a pair of interval and ratio scale quantities, the relations of 2.1 and 2.2 in table 2 must be used since the interval scale quantity is always defined by the ratio scale quantity. The relations 3.1 and 3.2 which is the inverse of 2.1 and 2.2, must be applied in case to commute a ratio scale quantity to an interval scale quantity.

insert table 3 here

By using this principle, the algorithm shown in table. 3 tries to discover the law equation formulae $\psi_{tk} = 0$ for the target set of measurement quantities Q_t, and confirms the second generality of the candidate law equations if $\psi_{tk} = 0$ is checked to be consistent with the measurement data of Q_t. For a comprehensive explanation, this algorithm is demonstrated through a simple example shown by the aforementioned Kepler's third law. Let Eq.(2) be a source equation $\psi_s = 0$, and consider the case to commute the angular velocity $\dot{\omega}_a$ to the major radius a where both are ratio scale. Hence, $Q_s = \{T, \dot{\omega}_a\}$ and $Q_t = \{T, a\}$. In the step (S1), the candidate law equations under Q_s are discovered by a law equation discovery system such as SDS. In this example, Kepler discovered the candidate law equation Eq.(2).

In the step (S2), the target equation formulae under Q_t are reasoned through the procedure $REASONING$. In $REASONING$, when Δ is not empty, a commutation operator δ_{ij} is popped from Δ, and x_i in ψ is commuted to x_j by the operator. If $x_i = u(x_j)$ is one of the cases 2.1, 2.2, 3.1 and 3.2 in table 2, two candidate formulae are derived, i.e., $h = 1, 2$, otherwise a unique candidate formula is derived. This procedure is recursively applied to each candidate formula until Δ becomes empty. In the example of Eq.(2), $\dot{\omega}_a = \alpha_* |a|^\beta$ is selected from table 2 for the commutation in $REASONING$ of the step (S2). By substituting this relation to Eq.(2), the following equation formula is predicted,

$$T^4 |a|^{3\beta} = C' \alpha_*^{-3}. \tag{4}$$

Without using the measurement data on a, the shape of Kepler's third law is obtained.

Finally, in the step (S3), the least square fitting of the predicted target equations to the measurement data of Q_t is conducted, and their consistency with the data is assessed. The following statistical F-test is used to judge if a target equation shows the consistency with the data of Q_t. This is the standard F-test to judge if the data fitting of an equation is acceptable in statistical sense.

$$\begin{aligned} &\text{If } F_0 > F(d-1, n-d, \alpha) \\ &\text{then the fitting is acceptable, else unacceptable,} \\ &\text{where } V_R = S_R/(d-1), V_e = S_e/(n-d) \text{ and } F_0 = V_R/V_e. \end{aligned} \tag{5}$$

Here, S_R is the regressive component, S_e the residual error component, d the number of measurement quantities in the equation, n the total number of measurement data used for the fitting and $F(d-1, n-d, \alpha)$ the lower bound of F value under the degree of freedom $(d-1, n-d)$ and α a risk rate. α is set to be 0.05 throughout this paper. When the target equation $\psi_{tk} = 0$ is accepted, both $\psi_s = 0$ and $\psi_{tk} = 0$ are considered to have the second type of generality. In the example, the formula Eq.(4) is adopted to the equation fitting on the measurement data $Q_t = \{T, a\}$ of the planet orbits, and the value of β becomes known to be -2. The resultant equation involves an absolute value operator $|\bullet|$ on a, however, this does not have any essential effect on the relation since the major radius a is always positive. Thus, Eq.(1) is obtained, and the mutual generality of Eq.(1) and Eq.(2) has been confirmed. This algorithm can be further applied, if the measurement data of other Q_t on the same objective system are available.

Assumption 2 is a sufficient condition that x_i and x_j have the relations represented in table 2, i.e., if they have a direct dependency, then the target law equation formula $\psi_{thk} = 0$ holds. In other words, if the measurement data of Q_t do not follow any target law equation formulae, the strong evidence that x_i and x_j do not have any direct dependency is provided. Otherwise, $\psi_{thk} = 0$ can be accepted as a law equation formula for Q_t as far as it well fits to the measurement data of Q_t.

4 Performance Evaluation

The basic performance of the proposed method has been evaluated through simulation examples. One of the major issues of the performance is the noise robustness of the equation fitting and the statistical F-test to judge if the derived target equation shows the consistency with the data of Q_t. The second important issue is the performance to identify a correct target equation from the multiple candidate target equations in case that the commutations between the interval and ratio scale quantities are involved. The third important issue is the performance to judge if the commuted quantity x_i has the direct dependency with x_j.

insert table 4 here

Table 4 shows the evaluation result for the four artificial simulation examples. The second column indicates the original candidate law equations which have been discovered in the step (S1) in table 3. SDS has been used to discover these equations since the scale-types of the quantities are all known in these examples, and the performance of SDS is known to be high from the past experiments (Washio and Motoda 1997, 1998). The third column shows the true formulae of the target equations used in the simulations. The fourth indicates the candidate target equations deduced in the step (S2). The symbols of the constants appearing in table 2 are retained in these expressions. The fifth shows the equations resulted in the least square fitting in the step (S3). The data of Q_t was generated through the simulation using the true target equations, and the nonlinear least squares fitting method of Levenberg-Marquardt has been applied (More 1977). Some constants in the candidate equations have been put together into a smaller number here. They are represented, only when the equations are accepted by F-test in the majority of trials. The rest of the columns shows the percentage of the accepted cases for each candidate target equation under 50 measurement data of Q_t with the noise level of 0%, 5% and 20%. The aforementioned fifth column indicates the equations identified under the 5% noise level. The noise level stands for the standard deviation of the Gausian random noise relative to the absolute value of each quantity. Totally, 100 trials were conducted for each candidate, and the percentage of the acceptance was calculated.

The first example is the case of the aforementioned Kepler's third law. The candidate target equation has been successfully accepted even under the large noise of 20%. The second is an example of heat transfer across a surface between materials of temperature T_{c1} and T_{c2} in Celsius unit which are interval scale. K is the heat transfer coefficient, and $\dot{H}$ is the heat transfer rate. In this example, the temperature is commuted to the absolute temperature T_{a1} and T_{a2} in Kelvin unit which are ratio scale. Because of the two candidate relations of 2.1 and 2.2 in table 2 for each commutation, totally four candidate target equations are obtained. The second candidate is the correct one, and it is perfectly accepted by F-test under any noise levels, while the others are mostly rejected. The target equation has been successfully reconstructed in the equation formula shown in the fifth column. The third example is the relation between the input and the output voltage differences V_i and V_o of the electric emitter follower amplifier depicted in figure 1 consisting of a transistor and a resistance. R_{BE} is the resistance between the base and the emitter of the transistor and h_{fe} the amplification ratio between the base and the collector electric currents. V_i and V_o which are ratio scale are represented in form of the logarithmic intensity, A_i and A_o in dB in the target equation. Since A_i and A_o are interval scale, again four candidate target equations are deduced, where the first candidate is correct. As shown in the columns of F-test, only the first is accepted in the majority of the trials. The robustness against noise is slightly degraded since the candidate target equations are quite complex for the data fitting. The identified equation in the fifth column shows the almost perfect reconstruction of the true target.

insert figure 1 here

The fourth example is the relation among the velocity of a pendulum $\dot{x}$, the elapsed time t, the oscillation angle velocity ω and the oscillation amplitude A. The elapsed time t which is a ratio scale quantity is commuted to the position of a pendulum x which is another ratio scale quantity. In this case, the true target equation formula does not match to the candidate since t and x have an indirect relation $\omega t = \arcsin(x/A)$ where they are coupled by the dimensionless quantities ωt and x/A. In fact, the candidate equation formula was rejected in all F-tests.

insert table 5 here

Table 5 indicates some advantages of the proposed method in comparison with the case to discover target equations by a law equation discovery system and check its mathematical consistency with the source equation. SDS is used for the discovery of the target equations. The performances of the proposed method and SDS have been evaluated for the three aforementioned examples under the 5% noise level. The task of the nonlinear least squares fitting occupies the major portion of the computation time of the proposed method because the search space of the reasoning needed to derive the candidate equations is quite limited. The complexity of the nonlinear least squares fitting is $O(m^2) \sim O(m^{2.5})$, where m is the number of quantities involved in the target equations. This is almost comparable with $O(m^2) \sim O(m^3)$ of SDS. However, the actual computation time of the proposed method indicated for the 50 samples case in the third column is far smaller than that of SDS, since the task of the data fitting is the heaviest process, and the required number of the data fitting in SDS is proportional to $O(m^2)$, whereas the proposed method performs only once. The fourth and the fifth columns in the table show the error percentage averaged over the coefficient errors relative to the absolute values of the coefficients in each equation. The proposed method shows very strong noise robustness in case of the larger number of quantities and the small sample data. This is because the noise involved in the data does not affect the reasoning to derive candidate equations. Only the F-test at the final step can be distorted by the noise. In contrast, the reasoning of SDS can be statistically affected by the noise since the least squares fitting is essentially involved in the reasoning mechanism of the equation formulae. This is a common feature of the conventional law equation discovery systems.

5 Application to a Practical Problem

The power of the proposed method is demonstrated through a real world problem in the socio-psychological domain. The objective of the application is to enhance the plausibility of the candidate law equations governing the mental preference of people on their houses subject to the cost to buy the house and the earthquake risk at the place of the house.

In the step (S1) of table. 3, we designed a questionnaire sheet to ask the preference of the house in the trade off between the frequency of huge earthquakes, x_1 (earthquake/year), and the cost to buy, x_2 (Yen). In the questionnaire, 9 cases of the combinations of the cost and the earthquake frequency are presented, and each person chooses his/her preference from the 7 levels for each case. We distributed this questionnaire sheet to the people owning their houses in the suburb area of Tokyo, and totally 400 answer sheets were collected back. The answer data has been processed by following the method of successive categories which is widely used in the experimental psychology to compose an interval scale preference index y_I (Torgerson 1958). The basic principle of this method is to evaluate the quantitative interval distances among the categorical preference levels based on the answer distributions on the categorical levels. The answers on the 7 levels have been transformed to the range of $[-1.37, 2.04]$ on the interval scale. Hence, a set of observed data $OBS_I = \{X_1, X_2, ..., X_{400}\}$ where $X_i = [x_{1i}, x_{2i}, y_{Ii}]$ is obtained. Because this is a passively observed data set, the original SDS is not applicable. Accordingly, we adopted the extended SDS which can discover law equations from passively observed data (Washio et al. 1999). The extended SDS seeks law equations of the form $y_I = f(x_1, x_2)$, where x_1 and x_2 are ratio scale quantities. The discovered candidate law equations are the following two,

$$y_I = 0.63 \log x_1 + 0.34 \log x_2 - 2.9, \tag{6}$$

$$y_I = -7.9 x_1^{-0.23} x_2^{-0.11} + 3.5. \tag{7}$$

The plot of Eq.(6) is depicted in figure 2 and that of Eq.(7) figure 3. Each black dot in the plots stands for the average point of all cases subject to the cost and the earthquake frequency. Both of Eq.(6) and Eq.(7) fit nicely to the data, and show the monotonic relations among the three quantities.

insert figure 2 and 3 here

In the step (S2), we designed another style of the questionnaire sheet to ask the identical contents to the same people. In this questionnaire, the preference was asked in form of the paired comparison among the 9 cases of the combinations of the cost and the earthquake frequency. Each person compares two cases at a time, and chooses its relative preference from the 7 levels in each comparison. The answer data have been processed by following the constant-sum method which is also widely used in the experimental psychology to compose a ratio scale preference index y_R (Comrey 1950). The basic principle of this method is to evaluate the quantitative ratios among the categorical relative preference levels based on the statistical expectations. The answers have been transformed to the range of $[0.04, 12.06]$ on the ratio scale. Through this process, a set of observed data $OBS_R = \{X_1, X_2, ..., X_{400}\}$ where $X_j = [x_{1j}, x_{2j}, y_{Rj}]$ is obtained. Because both y_I and y_R measure the identical psychological feature, they are considered to have the direct dependency. Thus, the commutation of y_I to y_R based on the scale-types is applied to both candidates of Eq.(6) and Eq.(7). By substituting 2.1 and 2.2 in table 2 to the equations, the following four candidate target equations are deduced,

$$y_R = e^{-\frac{\beta_* + 2.9}{\alpha}} {x_1}^{0.63/\alpha} {x_2}^{0.34/\alpha} \text{ from Eq.(6),} \tag{8}$$

$$y_R = \left(\log {x_1}^{0.63/\alpha} {x_2}^{0.34/\alpha} - \frac{\delta + 2.9}{\alpha} \right)^{1/\beta} \text{ from Eq.(6),} \tag{9}$$

$$y_R = e^{\frac{-\beta_* + 3.5}{\alpha}} e^{-\frac{7.9}{\alpha} {x_1}^{-0.23} {x_2}^{-0.11}} \text{ from Eq.(7)}, \quad (10)$$

$$y_R = \left(-\frac{7.9}{\alpha_*} {x_1}^{-0.23} {x_2}^{-0.11} + \frac{-\delta + 3.5}{\alpha_*}\right)^{1/\beta} \text{ from Eq.(7)}. \quad (11)$$

In the step (S3), these equations are subject to the least square fitting and F-test under the data OBS_R. Only Eq.(8) and Eq.(11) have been accepted by F-test, and their resultant equation formulae are as follows,

$$y_R = 0.081 {x_1}^{0.438} {x_2}^{0.236} \text{ from Eq.(8)}, \quad (12)$$

$$y_R = (1.27 {x_1}^{-0.23} {x_2}^{-0.11} + 2.46)^{-1.92} \text{ from Eq.(11)}. \quad (13)$$

This fact indicates that both Eq.(6) and Eq.(7) are plausible in terms of the generality over the two questionnaire investigations.

6 Discussion and Related Work

The extended SDS has also been applied to the data OBS_R obtained from the second questionnaire investigation. The following unique candidate law equation has been discovered by the extended SDS based on the data.

$$y_R = 0.146 {x_1}^{0.449} {x_2}^{0.207}. \quad (14)$$

The structure of the equation is identical with Eq.(12). Furthermore, their power coefficients are almost the same to each other. This evidence supports the high plausibility of the equations of Eq.(6), Eq.(12) and/or Eq.(14). However, the equation similar to the more complex Eq.(13) has not been discovered by the extended SDS. This may be because the basic algorithm of SDS seeks the law equations starting from the simpler formulae in a bottom up manner. Many of the conventional law equation discovery systems apply similar search strategies taking into account the principle of parsimony. Though this is one of the most important criteria of the first principle law equation, the extra equations meeting with the other important criteria such as the mathematical admissibility and the statistical goodness of fitting are considered to be also plausible, and should be retained in the candidate law equations. In this sense, Eq.(13) should not be missed.

insert figure 4 here

The plot of Eq.(14) depicted in figure 4 shows some fitting error. Eq.(12) and Eq.(13) have the same tendency. The cause of this error is considered to be the higher work load of the people to answer the pair wise comparisons in the questionnaire sheet. Because the pair wise comparison requires to judge over multiple cases and to answer many questions, the answers of some people can become inconsistent due to the workload.

The generality over the multiple types of measurement conditions has hardly been discussed in the past study of the scientific discovery. In the field of qualitative reasoning, the modeling of physical process based on multiple views has been actively studied (Forbus 1984), and some automated systems such as SIMGEN have been developed to build the model of a given object and conduct the simulation while taking into account the consistency among multiple domain descriptions (Forbus and Falkenhainer 1990). In more recent study, the hybrid modeling approach have been proposed to apply quantitative and qualitative descriptions to the process involving discontinuous phenomena (Mosterman and Biswas 1996). In the research field of multi agent learning and organizational learning, many studies have investigated the learning of generic knowledge in the space where the agents explore (Weis 1996). Some of them addressed the use of heterogeneous agents having mutually different combinations of sensors, and reported the increase of efficiency and/or quality of the learning under certain conditions. However, most of the past studies have not paid much attention on the contents of the required axioms which enable mutual translation and sharing the knowledge among the models.

7 Conclusion

This paper pointed out the importance of the use of the second generality criterion over multiple combinations of measurement quantities to enhance the plausibility of the scientific discovery. The proposed method based on the admissible relations yielded by the scale-type constraints has the performance of the efficient reasoning, the superior noise robustness and the applicability to the small sample data. These features are highly beneficial since the data acquisition and/or sensing in high quality for the new set of measurement quantities is very expensive in many practical fields. In addition, the ability of the proposed method has been demonstrated to capture the complex but admissible law equations which may be missed by some conventional law discovery systems as demonstrated in the aforementioned application. Moreover, the ability to detect the indirect dependency between the quantities for commutation has been also demonstrated. Finally, the practicality of the proposed method has been confirmed through the real world scientific law discovery in socio-psychology.

References

Comrey, A. L., 1950, A proposed method for absolute ratio scaling, *Psychometrika*, **15**: 317–325.

Dzeroski S. and Todorovski L., 1995, Discovering dynamics: from inductive logic

programing to machine discovery, *Journal of Intelligent Information Systems*, **4**, (1): 89–108.
Falkenhainer, B. C. and Michalski, R. S., 1986, Integrating quantitative and qualitative discovery: the ABACUS system, *Machine Learning*, **1**, (4): 367–401.
Forbus, K., 1984, Qualitative process theory. *Artificial Intelligence*, **24**: 85–168.
Forbus, K. and Falkenhainer, B., 1990, Self-explanatory simulations: An integration of qualitative and quantitative knowledge, *Proceedings of Eighth National Conference on Artificial Intelligence*: 380–387.
Koehn, B. and Zytkow, J. M., 1986, Experimenting and theorizing in theory formation, *Proceedings of the International Symposium on Methodologies for Intelligent Systems*: 296–307.
Kokar, M. M., 1986, Determining arguments of invariant functional descriptions, *Machine Learning*, **1** (4): 403–422.
Langley, et al., 1987, *Scientific discovery: Computational explorations of the creative processes*, Cambridge, MA: MIT Press.
Ljung, L., 1987, *System identification theory for the user*, New Jersey: P T R Prentice-Hall.
Luce, R. D., 1959, On the possible psychological laws, *Psychological Review*, **66** (2): 81–95.
More, J. J., 1977, The Levenberg-Marquardt algorithm: Implementation and theory, In Watson, G. A. (Ed.), *Numerical analysis: Lecture notes in mathematics 630*, New York: Springer-Verlag.
Mosterman, P. J. and Biswas, G., 1996, A formal hybrid modeling scheme for handling discontinuities in physical system models, *Proceedings of Thirteenth National Conference on Artificial Intelligence*: 985–990.
Stevens, S. S., 1946, On the theory of scales of measurement. *Science*, **103** (2684): 677–680.
Todorovski, L. and Dzeroski, S., 1997, Declarative Bias in Equation Discovery, In *Proceeding of the fourteenth International Conference on Machine Learning*, San Mateo, CA, Morgan Kaufmann: 376–384.
Torgerson, W. S., 1958, *Theory and methods of scaling*, New York: John Wiley.
Washio, T. and Motoda, H., 1996, Scale-based reasoning on possible law equations, *Qualitative reasoning: The Tenth International Workshop*, AAAI Press.: 255–264.
Washio, T. and Motoda, H., 1997, Discovering admissible models of complex systems based on scale-types and identity constraints, *Proceedings of Fifteenth International Joint Conference on Artificial Intelligence*: 810–817.
Washio, T. and Motoda, H., 1998, Discovering admissible simultaneous equations of large scale systems, *Proceedings of Fifteenth National Conference on Artificial Intelligence*: 189–196.
Washio, T., Motoda, H., and Niwa, Y., 1999, Discovering admissible model equations from observed data based on scale-types and identity constraints, *Pro-*

ceedings of Sixteenth International Joint Conference on Artificial Intelligence: 772–779.

Wasserman, P. D., 1989, *Neural computing: theory and practice*, New York: Van Nostrand Reinhold.

Weis, G. (Ed.), 1996, *Distributed artificial intelligence meets machine learning: learning in multi-agent environments*, Berlin Heidelberg: Springer-Verlag.

Table 1: Constraints on functional relations under scale-type characteristics.

C_n No.	Scale Types: Independent variable x_j	Scale Types: Dependent (Defined) variable x_i	Constraints*	Comments*
1	ratio	ratio	$u(kx_j) = K(k)u(x_j)$	$k > 0, K(k) > 0$
2	ratio	interval	$u(kx_j) = K(k)u(x_j) + C(k)$	$k > 0, K(k) > 0$
3	interval	ratio	$u(kx_j + c) = K(k,c)u(x_j)$	$k > 0, K(k,c) > 0$
4	interval	interval	$u(kx_j + c) = K(k,c)u(x_j) + C(k,c)$	$k > 0, K(k,c) > 0$

*c and C can be any real numbers.

Table 2: The admissible relations under scale-type characteristics

Eq. No.	Scale Types: Independent variable x_j	Scale Types: Dependent (Defined) variable x_i	Possible Relations	Comments*
1	ratio	ratio	$x_i = \alpha_*\|x_j\|^\beta$	$\beta/x_j, \beta/x_i$
2.1	ratio	interval	$x_i = \alpha \log\|x_j\| + \beta_*$	α/x_j
2.2			$x_i = \alpha_*\|x_j\|^\beta + \delta$	$\beta/x_j; \beta/x_i; \delta/x_j$
3.1	interval	ratio	$x_i = \alpha_{*a}e^{\beta x_j}$	β/x_j
3.2			$x_i = \alpha_{*b}\|x_j + \delta\|^\beta$	$\beta/x_j; \beta/x_i; \delta/x_j$
4	interval	interval	$x_i = \alpha_*\|x_j\| + \beta$	β/x_j

1) The notations α_*, β_* are α_+, β_+ for $x_j \geq 0$ and α_-, β_- for $x_j < 0$, respectively.
2) The notations α_{*a} is α_+ for $x_i \geq 0$ and α_- for $x_i < 0$, respectively.
3) The notations α_{*b} is α_{++} for $x_i \geq 0, x_j - \delta \geq 0$, α_{+-} for $x_i \geq 0, x_j - \delta < 0$, α_{-+} for $x_i < 0, x_j - \delta \geq 0$, and α_{--} for $x_i < 0, x_j - \delta < 0$, respectively.
4) The notations α/x means "α is independent of the unit x".
5) The relations in 3.1 and 3.2 are not derived from their constraints, but are inverse functions of 2.1 and 2.2.

Table 3: Algorithm to check the generality.

(S1) *Given measurement environments for* Q_s*, apply a law equation discovery system to the measurements of* Q_s*. Let* $\psi_s = 0$ *be a discovered candidate law equation.*

(S2) *Given* Q_t*, and let* Δ_{st} *be a stack of the commutation operators to derive* Q_t *from* Q_s*.* $E_t = \phi$*. For a* $\psi_s = 0$*, apply the procedure* $REASONING(\psi_s = 0, \Delta_{st}, E_t)$*.*

(S3) $E_f = \phi$*. For every* $\psi_{tk} = 0 \in E_t$ $\{$
   apply the least square fitting of $\psi_{tk} = 0$ *to the measurements of* Q_t*.*
   If the goodness of the fitting is accepted by F-test, $E_f \leftarrow E_f \cup \{\psi_{tk} = 0\}$*.*$\}$
The set of pairs $E_{st} = \{(\psi_s = 0, \psi_{tk} = 0) | \psi_{tk} = 0 \in E_f\}$ *contains highly plausible law equations in terms of the generality.*

$REASONING(\psi = 0, \Delta, E_t)$ $\{$

(P1) *If* $\Delta = \phi$*, then* $E_t \leftarrow E_t \cup \{\psi = 0\}$*, and return* E_t*.*

(P2) *Pop an operator* δ_{ij} *from* Δ*. apply* δ_{ij} *to* $\psi = 0$*, and obtain the equation set* $E = \{\delta_{ij}^h \psi = 0 | h = 1 \ or \ 1,2\}$*.*

(P3) *For every equation in* E*, apply* $REASONING(\delta_{ij}^h \psi = 0, \Delta, E_t)$*.*

(P4) *Push the operator* δ_{ij} *to* Δ*, and return* Δ *and* E_t*.*$\}$

Table 4: Performance evaluation for simulation examples.

					F-test		
Cases	Source	True Target	Candidate	Identified	0%	5%	20%
Kepler	$T = 297.2\dot{\omega}_a^{-0.75}$	$T = 5.39 \times 10^{-10} a^{1.5}$	$T = 297.2\alpha_*{}^{-0.75}\lvert a\rvert^{-3/4\beta}$	$T = 5.67 \times 10^{-10} a^{1.49}$	100%	100%	100%
Heat Trans.	$\dot{H} = K(T_{c1} - T_{c2})$	$\dot{H} = K(T_{a1} - T_{a2})$	$\dot{H} = K(\alpha_1 \log\lvert T_{a1}\rvert - \alpha_2 \log\lvert T_{a2}\rvert + (\beta_{1*} - \beta_{2*}))$		0%	0%	0%
			$\dot{H} = K(\alpha_{1*}\lvert T_{a1}\rvert^{\beta_1} - \alpha_{2*}\lvert T_{a2}\rvert^{\beta_2} + (\delta_1 - \delta_2))$	$\dot{H} = K(0.993T_{a1} - 0.998T_{a2})$	100%	100%	100%
			$\dot{H} = K(\alpha_1 \log\lvert T_{a1}\rvert - \alpha_{2*}\lvert T_{a2}\rvert^{\beta_2} + (\beta_{1*} - \delta_2))$		0%	0%	0%
			$\dot{H} = K(\alpha_{1*}\lvert T_{a1}\rvert^{\beta_1} - \alpha_2 \log\lvert T_{a2}\rvert + (\delta_1 - \beta_{2*}))$		0%	12%	0%
El. Amp.	$V_o = \frac{R(1+h_fe)}{R+R_{BE}} V_i$	$A_o = A_i + 4.34 \log \frac{R(1+h_fe)}{R+R_{BE}}$	$A_o = \frac{\beta_i}{\beta_o} A_i + \frac{1}{\beta_o} \log\left(\frac{\alpha_{i*a}}{\alpha_{o*a}} \frac{R(1+h_fe)}{R+R_{BE}}\right)$	$A_o = A_i + 4.71 \log \frac{R(1+h_fe)}{R+R_{BE}}$	100%	100%	53%
			$A_o = \pm\left(\frac{\alpha_{i*b}}{\alpha_{o*b}} \frac{R(1+h_fe)}{R+R_{BE}}\right)^{1/\beta_o} \lvert A_i + \delta_i\rvert^{\frac{\beta_i}{\beta_o}} - \delta_o$		7%	0%	0%
			$A_o = \pm\left(\frac{\alpha_{i*a}}{\alpha_{o*b}} \frac{R(1+h_fe)}{R+R_{BE}}\right)^{1/\beta_o} e^{\frac{\beta_i}{\beta_o} A_i} - \delta_o$		0%	0%	0%
			$A_o = \frac{\beta_i}{\beta_o} \log\lvert A_i + \delta_i\rvert + \frac{1}{\beta_o} \log\left(\frac{\alpha_{i*b}}{\alpha_{o*a}} \frac{R(1+h_fe)}{R+R_{BE}}\right)$		0%	0%	0%
Pendulum	$\dot{x} = A\omega \cos \omega t$	$\dot{x} = A\omega \cos \arcsin(x/A)$	$\dot{x} = A\omega \cos \omega \alpha_* \lvert x\rvert^{\beta}$		0%	0%	0%

Table 5: Computation time and noise robustness.

The upper row for each example shows the results for the proposed method and the lower row the results for SDS.

Example	Num. of quantities	Num. of data 50		500
		CPU time(sec)	Error %	Error %
Kepler	2	2.4	5.2%	2.1%
		10.3	3.4%	2.5%
Heat Trans.	4	3.6	0.5%	0.4%
		27.7	24%	3.2%
El. Amp.	5	4.9	8.5%	4.9%
		74.9	46%	3.7%

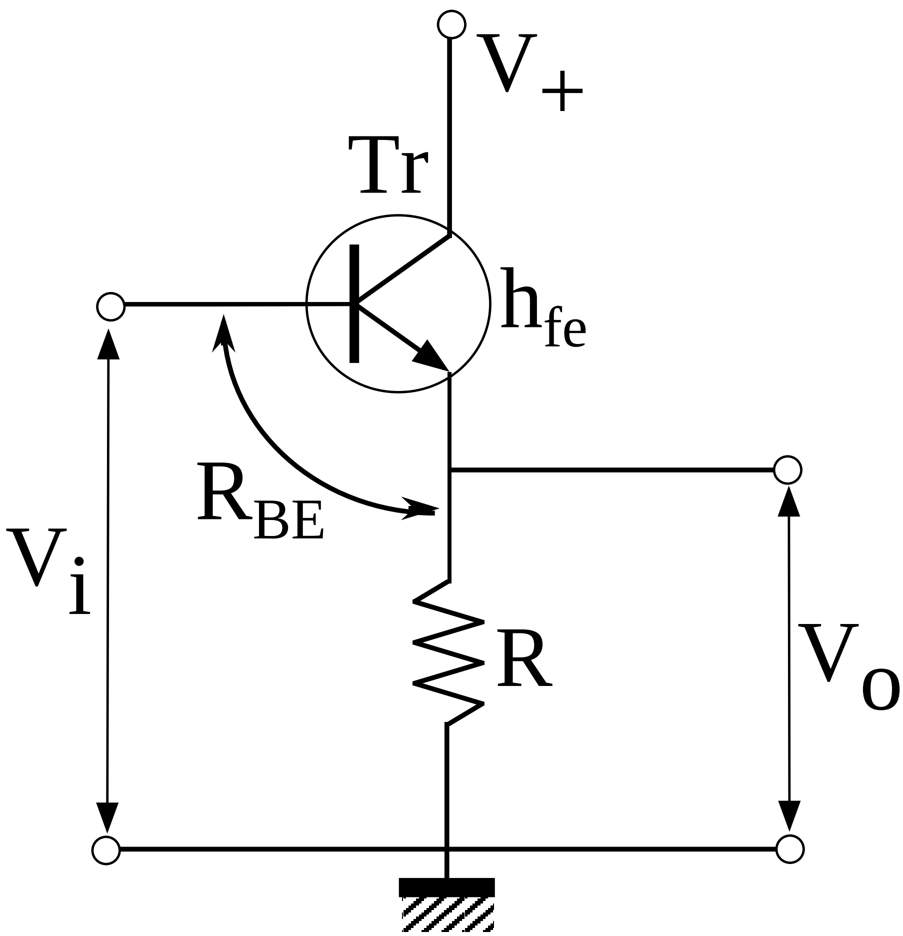

Figure 1: A circuit of amplifier.

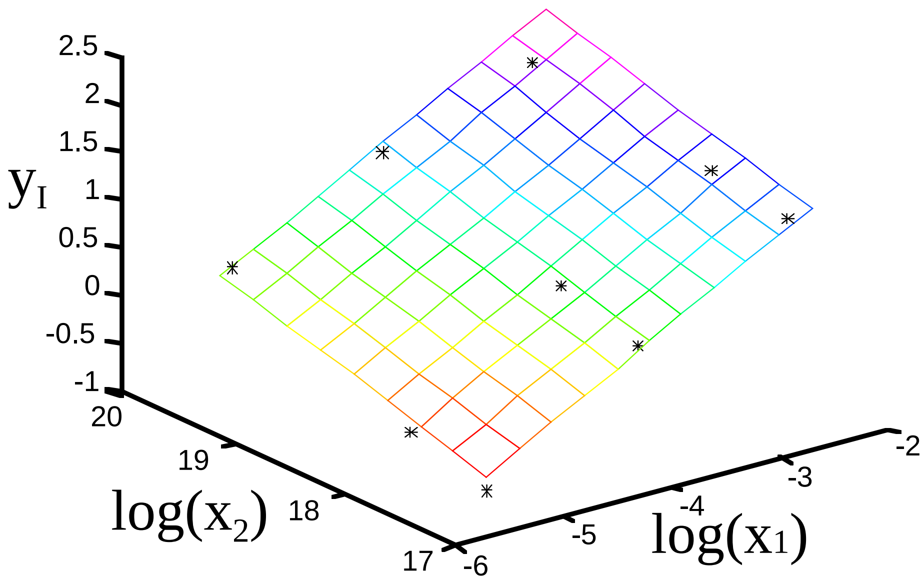

Figure 2: Plot of Eq.(6): $y_I = 0.63 \log x_1 + 0.34 \log x_2 - 2.9$.

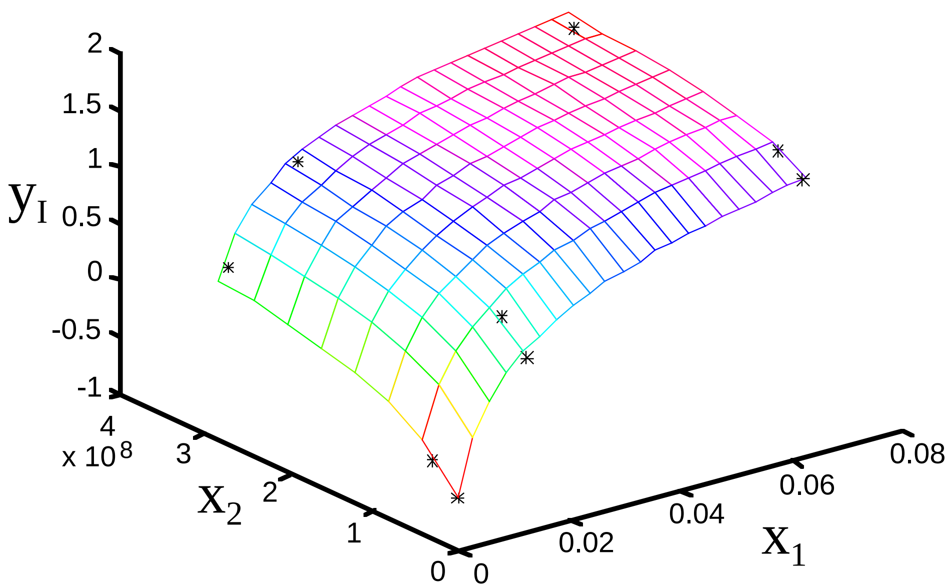

Figure 3: Plot of Eq.(7): $y_I = -7.9 x_1^{-0.23} x_2^{-0.11} + 3.5$.

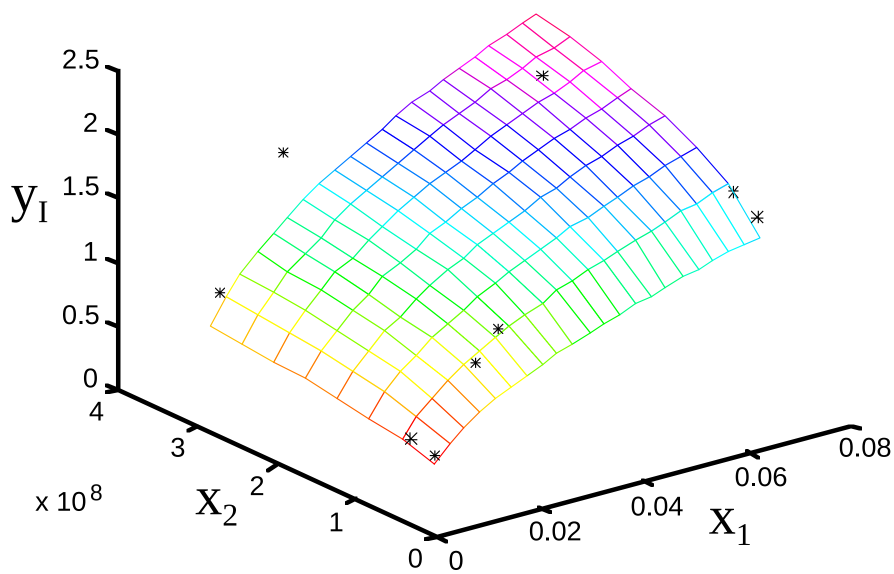

Figure 4: Plot of Eq.(14): $y_I = 0.164x_1^{0.449}x_2^{0.207}$.