DISCOVERING FUNCTIONAL FORMULAS THROUGH CHANGING REPRESENTATION BASE

Mieczyslaw M. Kokar
Department of Industrial Engineering
and Information Systems
Northeastern University
360 Huntington Avenue, Boston, MA 02115

ABSTRACT

This paper deals with computer generation of numerical functional formulas describing results of scientific experiments (measurements). It describes the methodology for generating functional physical laws called COPER (Kokar 1985a). This method generates only so called "meaningful functions", i.e., such that fulfill some syntactic conditions. In the case of physical laws these conditions are described in the theory of dimensional analysis, which provides rules for grouping arguments of a function into a (smaller) number of dimensionless monomials. These monomials constitute new arguments for which a functional formula is generated. COPER takes advantage of the fact that the grouping is not unique since it depends on which of the initial arguments are chosen as so called "dimensional base" (representation base). For a given functional formula the final result depends on the base. In its search for a functional formula COPER first performs a search through different representation bases for a fixed form of the function before going into more complex functional formulas. It appears that for most of the physical laws only two classes of functional formulas - linear functions and second degree polynomials - need to be considered to generate a formula exactly matching the law under consideration.

1. Introduction

Learning a functional formula from observation is composed of three major steps:

- deciding which features (arguments) to choose as relevant,
- selecting a functional formula generalizing the observational data,
- performing a function fit (calculating coefficients of the formula).

Even though the problem has a very long history in mathematics only the answer to the third step (given by mathematics) can be perceived as satisfactory. As to the first step, selection of relevant features, factor analysis can be used, but only if the set from which the features are to be selected is known. In the case of selection of functional formulas, mathematics offers the Weierstrass approximation theorem (c.f., (Johnson & Reise 1982)), which says that any function (fulfilling some additional restrictions) can be approximated, with any degree of accuracy, by a polynomial (possibly of a very high degree). Application of this theorem to describing observational data with functional formulas (e.g., in generalizing results of scientific experiments) leads to two kinds of problems. First, the formulas generated in this way do not fulfill some conditions of syntactic consistency required from formulas describing results of physical measurements. For instance, a formula generated according to this theorem might result in addition of meters to seconds, kilograms to square meters, etc., which are dimensionally inconsistent. The second problem is that it usually leads to quite complex formulas which are unacceptable to humans. Humans expect formulas similar to the ones representing physical laws - simple and consistent. Because of this the Weierstrass theorem alone is not acceptable as a tool for generating functional formulas describing observational data. Nevertheless, it should not be ignored totally. It provides a very important feature - convergence of an algorithm for generating functional descriptions.

The simplicity of physical laws contrasted with the complexity of functional formulas generated using standard approximation methods suggests that there should be a more "intelligent" method of describing results of scientific experiments by functional formulas. Attempts have been undertaken to use heuristics for discovery of functional formulas. The BACON system (Langley et al. 1983) used heuristics to generate both the features (arguments of a function) and the form of the function. The drawback of this system is that it can generate a formula which is dimensionally inconsistent. ABACUS (Falkenheiner 1985) uses dimensionality principles to constrain its search space. Dimensionality is used to eliminate those formulas which are dimensionally inconsistent. The search space and the space of dimensionally consistent functional formulas partially overlap, i.e., some of the formulas belong to the search space but they do not belong to the dimensionally consistent formulas (so they are tested and then rejected by ABACUS), and some dimensionally consistent formulas cannot be generated (they are not covered by the search space).

Obviously, it is impossible to generate the space of all functional formulas, but at least the whole search space could be a subset of all dimensionally consistent formulas. This paper describes the method of generating functional formulas used by the system called COPER, which fulfills this requirement. The methodology employed in COPER is based on three main principles: meaningful functions, Weierstrass theorem and change of representation base.

In short, the principle of meaningfulness of functions says that the functional formulas must fulfill some syntactic conditions to be interpretable in the relevant domain theory. For instance, in physics, a formula resulting in addition of meters to seconds is unacceptable. The most general restrictions imposed by the domain knowledge of physics are contained in the theory of dimensional analysis. COPER utilizes it in such a way that all the formulas it generates fulfill the formal condition of meaningfulness.

The second principle, Weierstrass theorem, justifies the use of the space of polynomials in the search for functional formulas.

The main strength of COPER, however, lies in the third principle, change of representation base. Perhaps it is worthwile to point here to an expressive example in (Charniak & McDermott 1985). In this example the numbers: 7, 60, 546, 627 are "magic." In base 10 it is hard to see what they have in common. In the representation base 9 they are: 7, 66, 77, 666, 766 and the common feature becomes much more apparent - they are composed only of two different digits, 6 and 7.

In the case of functional formulas the principle of change of representation base is related to dimensional analysis, mentioned above in reference to the principle of meaningfulness. Dimensional analysis provides rules for combining arguments of a function into a (smaller) number of dimensionless monomials. These monomials constitute new transformed arguments. The problem of finding a functional formula is reduced to finding a formula for the new transformed arguments. The interesting point is that the grouping of the arguments into monomials is not unique; usually there are a number of possible combinations. How the monomials are combined depends on which arguments have been selected as so called "dimensional base." Applying the same functional formula to the different groups of monomials (generated out of the same arguments) leads to different results. For a fixed functional formula we can either obtain a very simple resulting formula or a very complex one. In other words, the functional formula one derives depends on the selection of the dimensional base. This is analogical to the above mentioned selection of representation base.

Sections 2, 3 and 4 provide more details on the three principles employed in COPER.

2. Meaningful functions.

The notion of meaningfulness is known in the philosophy of science (measurement theory). It was investigated by many researchers, cf., (Adams et al. 1965), (Luce 1978), (Narens 1981), (Roberts 1984). In this approach functions are relations among elements called quantities. The structure of quantities is defined by some operations on them. For instance, in case of physical quantities the operations are multiplication and raising to a power. In addition to this, because numbers are also part of the quantity structure, other operations (logarithm, addition, etc.) are allowed for numbers. Meaningful functions are those that are expressed solely in terms of the operations defining the quantity structure. It is possible to show ((Luce 1978), (Kokar 1985)) that meaningful functions are invariant with respect to transformations of units in terms of which the quantities are expressed. The property of invariance is very important because in the theory of dimensional analysis (e.g., (Drobot 1953), (Whitney 1968)) there exists a theorem, called Pi-theorem, which relates the form of a function with the property of invariance.

To explain what the Pi-theorem is let us introduce some notation. Assume that a function is to relate the value of a dependent argument Z with the values of the independent arguments X1, ..., Xn, which is usually represented as:

$$Z = F(X1, \ldots, Xn).$$

The arguments X1, ..., Xn can be divided into two groups: a dimensionally independent set of arguments (they are usually called "dimensional base") and the remaining arguments being dimensionally dependent on the previous ones. Dimensional analysis provides algorithms for the division. Roughly speaking a dimensional base is such a maximal subset that the elements of it cannot be combined into a dimensionless monomial (using multiplication and exponentiation). For instance, velocity v[m] and acceleration a[m/s^2] can be part of a dimensional base because a monomial of the form:

$$v^{b1}.a^{b2}$$

can be dimensionless only for the values of b1, b2 = 0.

In this paper we are not going to discuss dimensional analysis; an interested reader is referred to the cited literature ((Drobot 1953), (Whitney 1968)). Assume that some of the arguments have been selected as the base arguments; let us denote them A1, A2, ..., Am. The rest of them will be denoted as B1, B2, ..., Br (where m+r=n). Our function takes the form

$$Z = F(A1, \ldots, Am, B1, \ldots, Br).$$

The Pi-theorem says that such a function can be represented in the form of

$$Z = f(Q1, \ldots, Qr)\cdot A1^{a1}\cdot\ldots\cdot Am^{am},$$

where each of the Q's is a dimensionless monomial of the form:

$$Qj = Bj/(A1^{aj1}\cdot\ldots\cdot Am^{ajm})$$

and dimensional analysis provides algorithms for calculating all the exponents ai, aji, (i=1,...,m; j=1,...,r).

Note that even the straight forward application of dimensional analysis significantly reduces the search space. Instead of searching the space of functions of n arguments we need to search the space of r=n-m arguments. Another advantage is that because Q's are dimensionless (like numbers), any of the functional formulas known for numbers can be used to generate the formula f. Therefore this method allows us to generate **only** meaningful formulas and **all** formulas admissible for numbers can be used to generate the function f (this is not the case with the function F). The system does not need to generate a formula and then test its meaningfulness; the generated formula is guaranteed to be meaningful.

3. The Weierstrass theorem.

As we mentioned before the Weierstrass theorem says that a function can be approximated with any accuracy by a polynomial, possibly of a high degree. Application of this theorem to the function F(A1,...,Am,B1,...Br) would result in formulas which might not fulfill the requirement of meaningfulness. It can however be applied to the transformed function f(Q1,...,Qr), since the arguments Q1,...,Qr are dimensionless and any operation admissible for numbers can be used here without any harm to meaningfulness. Because the Pi-theorem establishes equivalence between the two functions, the Weierstrass theorem guarantees existence of a polynomial approximating a searched function. The system can begin with the polynomial of the lowest degree (a linear function) and if some threshold accuracy of approximation is not achieved then the degree of the polynomial can be incremented and the function fit performed again.

Unfortunately, this is still not satisfactory; the degree of the polynomial may become unacceptably large. In the case of physical laws it does not guarantee obtaining results which would be an exact match with the formulas representing the laws. The usual way of solving this problem is to try out some other functional formulas, not only polynomials. There is however another possibility - changing the representation base (dimensional base). This approach has been incorporated into COPER.

4. Change of representation base.

According to the rules of dimensional analysis the set of arguments of a functional formula has to be subdivided into two sets - a dimensional base and the rest. Dimensional analysis provides algorithms for testing whether a given set of arguments satisfies the condition of a dimensional base. Three possible situations can take place:

- none of the subsets fulfills the condition of a dimensional base,
- only one subset fulfills this condition,
- there is more than one possible dimensional base.

In the first case it is not possible to represent this function with the arguments provided. This circumstance indicates that there should be some arguments that are not known to the system. This feature can be utilized nicely in the discovery system. Take for instance Ohm's law

$$U = R \cdot I,$$

where U represents voltage, I - current, and R - resistance. If COPER is asked to find a functional formula

$$U = F(I)$$

it will immediately request more arguments, because it is impossible to express U solely in terms of I using the admissible operations of multiplication and raising to a power. The resulting formula would be dimensionally inconsistent.

If the number of arguments is equal to the number of elements that must be included in a dimensional base and the set of arguments satisfies the condition of dimensional base, then the form of the function is determined uniquely. To understand this recall the Pi-theorem from Section 2. If n=m, then there are no Bj's, which means that there are no Qj's, and consequently the form of the function must be

$$Z=F(A1,\ldots,Am)=C\cdot A1^{a1}\cdot\ldots\cdot Am^{am},$$

where C is a constant numerical value. The reader can easily check that many physical laws have this form. If the dimensions of the arguments A1,...,Am are known then the values of the exponents a1,...,am can be calculated using algorithms of dimensional analysis. As an example of this situation take the formula describing pendulum period. If COPER is given arguments T[s] (pendulum's period, dependent argument), L[m] (pendulum's string length) and g[m/s^2] (acceleration of gravity), it will generate the formula

$$T = C \cdot \surd(L/g)$$

immediately, and then will calculate the value of the coefficient C using the measurement results.

In the third case, if there is more than one dimensional base among the arguments $X1,\ldots,Xn$, dimensional analysis does not give us any indication as to which base to choose. However, this may be advantageous to the process of searching for the form of the function. Instead of searching through the space of functional formulas, we can search through the space of dimensional bases first. Only if this search does not lead us to a plausible solution should we continue the search of the functional formulas space. Here is how such a search can proceed.

Step 1. Choose a form of the function f (see Pi-theorem in Section 2).

Step 2. Perform a search through the possible dimensional bases. To this end the system has to select a subset of arguments, test whether it satisfies the condition of dimensional base, if it does then express the Qj's in terms of this base (i.e., calculate exponents aji), calculate exponents ai, calculate the best coefficients for the given functional formula f, and calculate the accuracy of approximation of the experimental data by the function.

Step 3. Select the best representation, i.e., one for which the accuracy of approximation (approximation error) is the best.

Step 4. If the accuracy for the selected formula exceeds some threshold value then select another form of the function f and start from Step 1, otherwise stop.

In COPER such an approach has been implemented and the results are very promising. The space of functional formulas currently includes polynomials. In future implementations the space will be extended to other functional formulas. COPER starts its search with a polynomial of the lowest degree - a linear function. It tests all the possible bases. In the case of physical laws such an exhaustive search is feasible for large amounts of experimental data. In many cases of physical laws an exact match is achieved with a first degree polynomial, i.e., there exists a dimensional base for which the function

$$Z=(Co+C1\cdot Q1+\ldots+Cr\cdot Qr)\cdot A1^{a1}\cdot\ldots\cdot Am^{am}$$

gives an exact match with the formula representing a particular physical law. The example presented in the next section is intended to show how this algorithm works.

One of the directions for the future research is to investigate the influence of noise. The antinoise protection that COPER has stems from the fact that its decisions are based on the whole measurement results available at the time.

5. Example.

We will show the results of an application of the described method to the discovery of the functional formula representing uniformly accelerated motion:

$$S=Vo\cdot t+a\cdot t^2/2.$$

In this formula S[m] stands for distance (expressed in meters), Vo[m/s] - initial velocity, a[m/s^2] - acceleration, and t[s] - time. The system knows values of S for many different values of Vo, t, a, and the units of measurement (in square brackets). In this example 1000 of such values were generated out of the above formula. In practice they could be obtained from experiments.

The goal for COPER is to discover the above formula given the experimental results and knowledge about dimensions of the arguments. Obviously, the system does not have any knowledge about the form of the function. It knows only that it is supposed to generate a formula

$$S=F(Vo,\ a,\ t)$$

describing the experimental results.

In this particular case the dimensional base must consist of two arguments (any three arguments could be combined into a dimensionless monomial). Therefore there are at most three possible choices: (Vo, t), (Vo, a) and (a, t). Any one of the pairs may be selected since all fulfill the condition of a dimensional base. Below we represent the application of the Pi-theorem to this function for all possible selections of a dimensional base (note that in each case the new argument Q1 is dimensionless).

Table 1. Application of the Pi-theorem to S=F(Vo, a, t)

Base	Resulting formula
Vo, t	$S = f(Q1)\cdot Vo^2\cdot a^{-1} = f(t/(Vo^1\cdot a^{-1}))\cdot Vo^2\cdot a^{-1}$
Vo, a	$S = f(Q1)\cdot Vo^1\cdot t^1 = f(a/(Vo^1\cdot t^{-1}))\cdot Vo^1\cdot t^1$
a, t	$S = f(Q1)\cdot a^1\cdot t^2 = f(Vo/(a^1\cdot t^1))\cdot a^1\cdot t^2$

As we can see the problem of selecting a functional formula has been reduced from a function of three arguments to a function of one argument. In the first step of the search procedure COPER assumes that the functional formula for f is linear, i.e.,

$$f(Q1)=Co+C1\cdot Q1,$$

and starts searching through different dimensional bases. It represents the problem in the forms shown in the above table and performs a function fit. It then calculates the values of the coefficients Co and C1. Both the coefficients and the degree of accuracy of the fit are different for the three bases. The coefficients of the function and the degrees of accuracy obtained are represented in Table 2.

Table 2. Coefficients and accuracies for different bases

Base	Co	C1	Accuracy
Vo, t	-186294	2135	4.78E+6
Vo, a	1	0.5	1.40E-2
a, t	0.5	1	1.84E-4

The reason for the wide variation in the degrees of accuracy (ten orders of magnitude) becomes apparent when we substitute the values of the coefficients Co, C1 into the formulas and perform some simple symbolic operations leading to the elimination of parentheses as in the following table.

Table 3. Final functional formulas for different bases

Base	Resulting formula
Vo, t	$S = Co\cdot Vo^2/a+C1\cdot Vo\cdot t =$ $-186294\cdot Vo^2/a+2135\cdot Vo\cdot t$
Vo, a	$S = Co\cdot Vo\cdot t+C1\cdot a\cdot t^2 =$ $1\cdot Vo\cdot t+0.5\cdot a\cdot t^2$
a, t	$S = Co\cdot a\cdot t^2+C1\cdot Vo\cdot t =$ $0.5\cdot a\cdot t^2+1\cdot Vo\cdot t$

For the bases (Vo, a) and (a, t) we received an exact match with the original formula describing this law. Therefore no more searching is required; the first order polynomial satisfies the requirement of plausibility (low value of accuracy).

6. Conclusions.

The approach to searching for a functional formula describing scientific experimental results presented in this paper has been tested on many physical laws with positive results. Its strength lies in the fact that it generates very simple functional formulas exactly matching physical laws. This is achieved through changing the representation base (dimensional base) before going into more complex functional formulas. The approach also has been tested on real data, i.e., on the results of scientific experiments obtained by measuring a physical process for which the functional formula was not known. The results of these investigations have been described partially in (Kokar 1975, 1978). Here again COPER's formulas were simple while accurate. Still, there are physical laws for which COPER cannot generate an exact formula, e.g., if a logarithm is part of the formula. It can come up with a polynomial which describes the experimental data with sufficient precision (Weierstrass theorem), but the formula is too complex (too high degree of the polynomial). Research is underway to incorporate further heuristics for searching the space of functional formulas (not only polynomials). In any case, the idea of changing description base proved to be very useful in the process of discovery of functional formulas describing physical laws.

REFERENCES

[1] Adams, E., W., Fagot, R., F., and Robinson, R., E., (1965), "A theory of appropriate Statistics," Psychometrica, 30, pp. 99-127.

[2] Charniak, E., McDermot, D., (1985), Introduction to Artificial Intelligence, Addison-Wesley, pp. 616-617.

[3] Drobot, S., (1953), "On the Foundations of Dimensional Analysis," Studia Mathematica, 14, pp. 84-89.

[4] Falkenhainer, B., (1985), "Proportionality Graphs, Units Analysis, and Domain Constraints: Improving the Power and Efficiency of the Scientific Discovery Process," Proceedings of the Nineth International Joint Conference on Artificial Intelligence, August 1985, Los Angeles, California, pp. 552-554.

[5] Johnson, L., W., & Riess, R., D., (1982), Numerical Analysis, Addison-Wesley, p. 205.

[6] Kokar, M., (1975), "The Choice of the Form of the Mathematical Model Using Dimensional Analysis," (in Polish), Inzynieria Chemiczna, V, 1, pp. 103-119.

[7] Kokar, M., (1978), "A System Approach to Search of Laws of Empirical Theories," Current Topics in Cybernetics and Systems, Berlin-Heidelberg-New York.

[8] Kokar, M., M., (1985), "On Invariance in Dimensional Analysis," Technical Report, MMK-2-85, College of Engineering, Northeastern University, Boston, Massachusetts.

[9] Kokar, M., M., (1985a), "Coper: A Methodology for Learning Invariant Functional Descriptions," in R.S. Michalski, J.G.Carbonell, and T.M.Mitchell (Eds), Machine Learning: A Guide to Current Research, Kluwer Academic Publishers.

[10] Langley, P., Bradshaw, G., L., Simon, H., A., (1983), Rediscovering Chemistry with the Bacon System. In R.S.Michalski, J.G.Carbonell, and T.M.Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach, Tioga Pub., pp. 307-330.

[11] Luce, R., D., (1978), "Dimensionally Invariant Numerical Laws Correspond to Meaningful Qualitative Relations," Philosophy of Science, 45, pp. 1-16.

[12] Narens, L., (1981), "A General Theory of Ratio Scalability, with Remarks about the Measurement-Theoretic Concept of Meaningfulness," Theory and Decision, 13, pp. 1-70.

[13] Roberts, F., S., (1984), "On the theory of meaningfulness of ordinal comparisons in measurement," Measurement, 1, pp. 35-38.

[14] Whitney, H., (1968), "The Mathematics of Physical Quantities," American Mathematical Monthly, 75, part I and II, pp. 115-138, and 227-256.