

Knowledge-based systems for formulation optimization

William H. VerDuin

ABSTRACT: *Product formulation can be a challenging task due to the number of ingredients involved and the way they interact chemically and physically. New knowledge-based artificial intelligence software designed specifically for product formulation provides an alternative to traditional statistical analysis of experimental data.*

KEYWORDS: *Adhesives, artificial intelligence, formulations, models, neural networks, optimization, problem solving, statistical methods.*

Hot-melt adhesives are called upon to meet a variety of often conflicting requirements (1). They must adhere to a variety of surfaces and must meet bond strength and durability requirements. They must meet cost and perhaps regulatory and processing constraints.

In these respects, hot-melt adhesives are similar to many other formulated products. Products such as plastics, rubber, glass, metals, coatings, petrochemicals, and pharmaceuticals are also called upon to meet performance, processing, and cost objectives and constraints (2,3). Design, that is, formulation of these products, is a critical task because it directly determines the degree to which these goals are met.

Producers of formulated products face the competitive challenges of a

worldwide economy. Buyers expect products with better performance, easier processability, and perhaps lower costs. Sellers succeed, or not, on the basis of rapid response to changing requirements. These factors have focused attention on methods of designing formulated products. This article addresses that issue of product formulation methods. Capabilities of traditional methods are compared to a new alternative based on knowledge-based computer technologies. Applications are reviewed to compare and contrast the methods.

The need: a formulation model

The properties of formulated products are determined by the ratios in which ingredients are combined and

the processing that is done to these ingredients.

Formulation can be a challenging task, because the ingredients may interact both chemically and physically. These interactions may be known in an anecdotal sense, but rarely can they be completely described in a theoretical or empirical sense. The number of ingredients and the complexity of the interactions can present a challenging design problem.

This often means that in the absence of sophisticated analytical tools, the best that can be expected is that a chemist who is expert in product formulation can identify the general direction in which a recipe should be changed to meet new requirements, but the detailed reformulation can only be established by experimentation. Product properties may also be affected by processing methods. Again, the interactions are rarely fully understood.

What is needed is a way to explicitly relate formulations to properties so that designers can find the changes in formulations to provide needed changes in properties. Such a model would minimize experimentation because the user could predict the properties that would result from a hypothetical formulation. The computer would perform "virtual experimentation."

The need to reduce experimentation, and thus the need for a formulation model, is well known. The question is one of approach. Tradi-

VerDuin is general manager with AI WARE, Inc., 11000 Cedar Ave., Cleveland, OH 44106.

tional approaches include models based on first principles and statistics. An advanced computer technology called neural networks now provides a new modeling approach. Each technique has strengths and limitations.

Modeling approaches

First principles

In high school physics, one learns equations relating force, mass, and acceleration. Chemistry rules and equations describe the ways that chemicals combine and react. These so-called first principles are fundamental rules of behavior. A model based on first principles inspires confidence because it is based on these universal truths.

This approach is limited by the complexity of many real-world applications. They are simply too complex to model accurately and completely with first principles. The dimensionality of the problem may be high. The exact nature of the interactions between variables may be unknown. In many chemical formulation applications, for example, ingredients interact both chemically and physically. The experts know that these interactions occur, but in real-sized problems they often cannot disentangle the many interactions sufficiently to model them explicitly.

Statistics

These problems have provided the rationale for the most common modeling alternative, statistics (4). The goal of statistics (and of neural nets) is to model relationships too complex or opaque for first principles. The approach is to find observable and repeatable relationships between parameters assumed to represent "cause" and those presumed to represent "effect." There is no basis on first principles, only on observed relationships in data.

Statistical models may be based on known interactions. For example, an understanding of relevant first

principles may help select the form of the statistics equation. But this level of fundamental understanding may not be available, and so the statistical model will instead be based solely on relationships observed in data.

Development of a statistical model may be viewed as "curve fitting." An equation is sought that will generate a line passing acceptably close to known data points. This process involves selection of the correct form of equation, followed by identification of the correct values of the constants in the equation. For example, a simple linear relationship may be expressed as $y = Ax + B$. Or a second-order relationship may be suspected, and so an equation form such as $y = Ax^2 + Bx + C$ might be used. Values of A, B, and C would be found to generate a line that best fits known data.

At least in theory, this process can be applied to high-dimensionality problems by finding a surface that fits known points in n-dimensional space. With simple relationships and noise-free data, the equation form may be readily apparent. In more complex cases, it may not. Expertise in both statistics and the problem at hand is required to judge accurately the appropriateness of the equation form. This judgment may be based on knowledge of underlying relationships or on analysis of the observed data. If the correct form is not chosen, the resulting model will be flawed. (Similar issues arise with neural nets, although some neural net tools support or automate selection of the proper neural net architecture.)

Neural networks

Neural networks are a completely different approach to computing. They are also known as Artificial Neural Nets (ANNs) and simply neural nets. They are lumped together with expert systems and fuzzy logic in the broader category of knowledge-based systems or artificial intelligence because they share the

common goal of enabling computers to capture and apply knowledge, not just manipulate numbers and letters.

Neural nets differ from conventional computer technologies (and expert systems, another knowledge-based technology) in that they are not procedural in nature (5,6,7). The essence of conventional programming is to specify procedures to be applied to data. The programmer explicitly defines the operations and the bounds on these operations to accomplish the desired computational goal. Computer languages enable programmers to manage the flow of data, define procedures, and manage the interaction between these procedures. Each program has a start, a logical progression, and an end.

Neural nets provide an interesting contrast. They are not programmed as such, but are taught by example. Neural nets learn to solve a problem by being shown examples of situations and associated solutions. The net learns a relationship that is valid between each of the sample situations and the associated solutions. Upon finding this relationship, the net is able to generalize to provide an appropriate solution to a new situation.

Neural nets have several interesting properties. They can improve their own performance. They can adapt. They can discover relationships in data. This means that a computer-based system can now learn by itself to design better products and perform manufacturing tasks more effectively.

The formulation of hot-melt adhesives is an excellent application. The design problem is to vary the recipe to provide new product properties. If a neural net is provided data on previous adhesive recipes, along with data on the properties of products made according to that recipe, the neural net can discover the relationship between the two. In this case, that relationship is a product formulation model that can be used to design new or better products. The user

Formulation

need not know how changes in recipe affect changes in properties. The user need only provide historical data that capture that relationship.

A comparison

Reports on the relative merits of statistics and neural nets by users of both indicate strengths in statistical methods, including widespread familiarity among users and a great variety of supporting products. Tyson Gill and Joel Shutt, reporting on their formulation applications for Glidden and Lord Corp., respectively, identified a number of benefits of neural nets over statistics (8). Statistics are subject to error in instances with sparse data, particularly when those data do not include data points representing all combinations of the extreme values of each variable. This situation is not uncommon when available resources preclude a formal Design of Experiments (DOE) program and collection of the entire data set specified by the DOE program.

Statistical methods are also subject to error in applications involving intercorrelation between parameters. Statistical methods encounter difficulty because they are not well suited to separating the individual influence of each variable. A practical, if not particularly accurate, approach in these cases is to treat the problem as a large number of independent single-variable problems rather than one multivariable problem. This simplifying assumption may lead to significant modeling errors in some cases.

Neural nets handle multiple interacting variables as such, without artificially simplifying the problem. Mixture problems, such as product formulation, are an example of this type of problem. The proportions of ingredients in a mixture are clearly independent. Yet the effect of each ingredient can depend on the level of each of the other ingredients, due to chemical and physical interactions between ingredients. Users report that because of the difficulties with

statistical methods in accommodating these interactions, neural net models are often more accurate than statistical models for mixture problems.

Gill and Shutt reported that first principles, statistics, and neural nets each have their strengths in model application. In particular, they reported that:

- Models based on first principles are preferred when feasible.
- Statistics are a more established technology than neural networks, and so may be more comfortable to many.
- Statistical methods are effective in lower-order relationships involving only independent variables.
- Neural networks can provide better results than statistics in multidimensional and time-varying situations, and when variables exhibit interdependencies.
- Neural networks are more effective than statistics with data not meeting rigorous DOE criteria. Neural nets have been shown to provide more accurate models when data are not available for all combinations of extreme values of all variables.

The goal: An optimal design

A good model provides accurate answers to the "what-if" question, the question of "What will happen to properties for a particular change in formulation?" But from the formulator's perspective, this "what if" capability may be necessary but not sufficient. To paraphrase Tyson Gill, the goal is not to find out "what if" a particular formulation is tried; instead, designers are more interested in "how to get" to new product properties.

The chemist's interest is in meeting new objectives. These objectives may include performance, price, and manufacturability, and these objectives may be conflicting. The chemist may call upon a formulation model to

do this. But since the model answers "what if," not "how to get," questions, the chemist must guide the model through a directed search of "what if" exploration. By intelligently repeating the "what if" question, the chemist will eventually learn to ask the right question, the one that yields the answer of the desired product attributes and thus answers the more difficult "how to get" question.

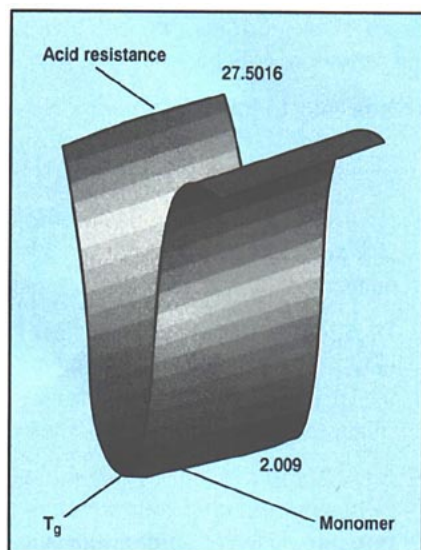
In the course of developing several neural net-based formulation support systems for applications in rubber and plastic products, AI WARE, Inc. found that quite a few expert formulators were not seeking merely the "what if" capability of a traditional model but the "how to get" capability. They wanted to save the time required to repeatedly ask "what if." More importantly, they noted that "what if" analysis did not directly address all formulation requirements.

In real-world design problems, some design goals are critical, others less so. Furthermore, some design goals may require that an exact value be achieved, but other goals may be more loosely constrained. Perhaps the value must be held only within a certain range, not to a single target value. In addition, there may or may not be preferred values within this range.

Design recommendations that do not recognize these types of constraints were seen by these formulators as not responsive to real-world needs and therefore of theoretical interest only. Solutions that did not take advantage of prioritization and allowed ranges of values were seen to be not optimal because they were artificially constrained.

In the course of developing several custom formulation support systems, the difficulty was discovered in answering the "how to get" question. The obvious solution of developing an inverse formulation model was tried. Instead of starting with formulations, and from these, predicting properties, the idea was to start with properties and then estimate formulations.

1. Paint formulation monomer response surface



Unfortunately, this approach did not work. The formulation model was neural net-based, but the limitations found were not due to neural net capabilities. The larger issue was that if a mathematical mapping (or model) was found from formulation to properties, in general a unique mapping did not exist in the reverse direction. The inverse model would estimate a formulation that, at least theoretically, would provide the given properties. But this formulation was generally not the one that would best meet the multiple objectives of cost and performance while accommodating processing and materials constraints.

Thus the real need was not a “what if” formulation model alone but also a “how to get” functionality that would directly map properties to ingredients. Since a formulation model could not be legitimately inverted, an alternative approach was required. This “how to get” functionality was provided by linking a neural net-based model with an optimizer.

The following description of the use of a commercial software product called CAD/Chem (Computer Aided Design Tool for Chemists) (9) illustrates the benefits of “how to get” functionality. It also highlights the role of advanced computer technolo-

gies in formulation support. Neural nets provide easy and robust development of formulation models. Expert systems and fuzzy logic are also employed; they provide important support to the neural net-based modeler and the integrated optimizer.

Expert systems capture human problem-solving expertise in the form of rules (10,11). Expert systems are becoming commonplace in diagnostic and decision support applications. In this case, an embedded expert system functions as a system manager, managing the interaction of the neural net and the optimizer.

Fuzzy logic enables objectives to be stated qualitatively rather than quantitatively (12). This approach is receiving increasing attention in control applications as an easier and less rigid way to specify processing goals. In this case, fuzzy logic concepts are embodied in Desirability Functions, which users select to define property preferences within a range of allowed values.

The resulting integrated system provides the following functionalities:

- The neural net modeler provides models not limited by dimensionality or degree of nonlinearity.
- The user may directly express formulation and processing goals, constraints, and preferences.
- The implementation and application details of the advanced technologies employed are hidden from the user, who may instead focus on the design task at hand.

Applications of the integrated technology formulation system cover a wide variety of industries. Two representative applications are described in the following sections.

Paint formulation

Paint must meet a variety of potentially conflicting requirements, such as coverage; resistance to wear, dirt, and sunlight; ease of application; and

cost. The design task is to identify the paint formulation, or recipe, that best meets these multiple objectives.

In this case, Glidden’s goal was to enhance acid resistance. In an acid resistance study, seven formulation variables and three performance variables were measured, with two of the performance variables being the acid resistance measured by two different methods.

Data were imported into the integrated technology formulation system from a spreadsheet print file. The neural net-based model was developed, and the optimizer was configured to maximize the value of the results of the first acid resistance test and to minimize that of the second test. An unconstrained optimization was performed to find the best formulation without imposing any constraints on the region in which the program could search for an optimum formulation.

This search found an optimum outside the typical design space. For this particular formulation, the glass transition temperature (T_g) and crosslinker levels were below the range that the formulation chemist had specified.

The chemist wondered if this unexpected result was valid, and thus further explored this design approach. A second optimization trial was performed. In this case, the optimization space was restricted to include only the anticipated best-design region. Constraints were specified to restrict T_g and crosslinker to the experimental range previously specified by the chemist, and the optimization calculations were repeated.

A different optimum formulation was found with T_g and crosslinker levels in the middle of their specified ranges. Product properties requirements were met, but the level of a monomer had been changed from the usual 22% to 0.9% in this newly calculated optimum formulation.

This was interesting for two reasons. First, it suggested that the use of the expensive monomer could be reduced by 96%. Second, it also sug-

Formulation

gested a radically different design approach. This conclusion was initially suspect since the results were so different from the conventional wisdom.

A sensitivity analysis helped explain this result. Acid resistance according to the first test method was plotted as the vertical axis, with Tg and monomer levels the horizontal axes.

The monomer response surface, as shown in **Fig. 1**, has two peaks of nearly equal height. Constraining the ingredient levels to the anticipated best design region excluded the region of space which contained the global optimum and resulted in the use of the lower local optimum. In other words, conventional wisdom focused on one leg of the "U"-shaped response surface. However, the integrated technology formulation system discovered the cost-effective alternative represented by the other leg.

Adhesive formulation

Adhesive formulation is also a challenging task. Adhesives must meet requirements such as bond strength; resistance to extreme temperatures, salt water, and sunlight; ability to adhere to dirty or oily surfaces; ease of application; and cost.

In the application described below, Lord Corp. developed an adhesive that formed thermally stable bonds to a number of different substrates. This new product would enable commercial adhesive users to reduce the number of different adhesives to be ordered and stocked. Production control would be simplified, but not at the expense of performance.

The new formulation faced two challenges. The first was to accommodate a range of substrates. The second was to accommodate formulation limitations imposed by the manufacturing process.

A certain adhesive additive was known to promote adhesion in low concentrations, but it could be detrimental in greater concentrations. Experience also showed that different formulations of this additive were

required to form thermally stable bonds to different substrates. Thus the first design challenge was to find an additive formulation effective at low concentrations for the intended range of substrates.

In this application, manufacturing constraints were clear. The several elements in the adhesion promoter could only be combined in specific ratios. This in turn limited formulation opportunities. A new adhesive formulation was developed in a three-step process. First, the neural net-based formulation model learned the relationship between product formulation and resulting properties. A statistical approach would have been difficult, since common statistical methods assume that ingredient proportions are independent, but in this case process limitations impose dependencies in the form of specific allowable proportions between ingredients.

The formulation model was developed using data covering seven concentrations of adhesion promoter from a total of 38 different formulations.

In the second design step, the integrated optimizer found the formulation that would best adhere to the many substrates. Design objectives were defined directly, rather than relying on the common but slower "what if" approach.

As the third design step, the optimization was rerun without the formulation constraints imposed by the process. This enabled identification of a formulation that better met functional requirements and in turn enabled improvements in adhesive cost and performance to be balanced against the required process improvement costs.

The reported benefits for this application included development of a useful new product and a significant reduction in design time. The several days required to fully explore design alternatives with the integrated technology formulation system accelerated a design process that otherwise might have taken an additional year.

Conclusions

From these studies, we can draw the following conclusions:

- It is easy to miss design opportunities due to the complexity of interactions between ingredients.
- Real-world designs must accommodate constraints and preferences.
- Designers may find significant benefits in performance and cost by fully exploring design alternatives.
- New computer-based technologies have enabled new formulation approaches and opportunities, and have resulted in the development of higher performance, lower cost products. ▮

Literature cited

1. Gauthier, M.M., *Advanced Materials and Process Magazine*, p. 26 (July 1990).
2. VerDuin, W.H., in *Handbook of Expert Systems in Manufacturing* (Maus, R., and Keyes, J., Eds.), McGraw-Hill, New York, 1991, p. 515-524.
3. VerDuin, W.H., *J. of the Adhesive and Sealant Council, Inc.* 22(1):269(1992).
4. Hamburg, M., *Statistical Analysis for Decision Making*, Harcourt, Brace & World, Inc., New York, 1970.
5. Pao, Y.-H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA, 1989.
6. Caudill, M., *AI Expert Magazine*, 2(12):46(1987).
7. Caudill, M., *AI Expert Magazine*, 3(2):55(1988).
8. Gill, T., and Shutt, J., *Scientific Computing & Automation Magazine* 5(10):19(1992).
9. "CAD/Chem Custom Formulation System Technical Overview," product literature, AIWARE, Inc., Cleveland, OH, 1992.
10. Ferrada, J.J., and Holmes, J.M., *Chemical Engineering Progress Magazine*, p. 34 (April 1990).
11. Keyes, J., *AI Expert Magazine* 4(11):50(1989).
12. "Fuzzy Logic: A 21st Century Technology," product literature, Omron Electronics, Inc., 1991.

Received for review March 3, 1994.

Accepted April 12, 1994.

Presented at the 1994 TAPPI Hot Melt Symposium.