Towards monotonous functions approximation from few data with Gradual Generalized Modus Ponens: application to materials science

Hiba Hajri, Jean-Philippe Poli, Laurence Boudet

Towards Monotonous Functions Approximation from Few Data With Gradual Generalized Modus Ponens: Application to Materials Science

Hiba Hajri
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France
hiba.hajri@cea.fr

Jean-Philippe Poli
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France
jean-philippe.poli@cea.fr

Laurence Boudet
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France
laurence.boudet@cea.fr

***Abstract*—In this paper, we present a new approach to predict monotonous functions based on approximate reasoning and in particular on the Gradual Generalized Modus Ponens (GGMP) in fuzzy logic. We propose to optimise the parameters of such fuzzy rules with a genetic algorithm considering few experimental data. We use our approach to predict some properties of materials from their manufacturing process parameters. We automatically extract causality, seek for graduality and then set up the GGMP. We tested on both toy and real world datasets. We also discuss the importance of gradual knowledge in materials science.**

***Index Terms*—Fuzzy logic, Gradual Generalized Modus Ponens, Knowledge Extraction, Materials Science, Monotonous Function Prediction, Model Interpretability.**

I. Introduction

Discovering new materials is one of the main objectives of experts in the materials science field. It requires different experimental tasks. To guide experiments, researchers are interested in characterising relationships between the material properties (e.g. mechanical, physical) and its manufacturing parameters. Artificial Intelligence has been widely used to predict these properties from the process parameters, mostly applying machine learning and black box models, statistics, and sometimes expert systems [1].

We interest in extracting relevant knowledge from experimental data as experts would do. To this end, we propose a fuzzy logic approach to extract automatically insights from the experimental data, in the form of rules, that can be then used for properties prediction. Fuzzy logic is used here for its closeness to natural language and its ability to deal with vague knowledge and imprecise data.

One of the most important information materials researchers seek is gradual relations, for instance like “the higher the temperature, the clearer the material”. In this paper, we focus on extracting this kind of relations from an experimental dataset and formalizing them with fuzzy gradual rules. However, experimental data are often costly to produce: not only because of the raw materials cost, but also because the product characterisation is mainly performed by experts and is time consuming. We thus propose an approach that deal with few data.

The paper is structured as follows. The next section presents an overview of our approach. Section III gives the background of this work about causality, graduality and Gradual Generalized Modus Ponens (GGMP). We then present the extraction of gradual rules from data (section IV) to represent specific relations between parameters and properties, and how we use GGMP to approximate their values (sectionV). We validate our approach with synthetic data and show results in section VI. Section VII presents the results on a real world dataset and in section VIII, we compare them to the performances of other regression methods. After discussing (section IX), we finally draw some conclusions and perspectives.

This work is funded by the CEA Cross-Cutting Program on Materials and Processes Skills.

II. Approach overview

In this section, we describe our approach to extract and use relevant gradual knowledge from data to predict the materials properties from process parameters (Fig. 1).

The first part of our approach is the characterisation of links between manufacturing process parameters and materials properties: (1) causality links and (2) gradual links through gradual itemsets discovery with the GRAANK algorithm introduced in [2] and described in section III-B.

Then we combine results obtained from causality links detection and gradual itemsets detection to generate final gradual links. Each gradual link binds a set of manufacturing parameters and the property influenced by these parameters.

Each gradual link is then represented with a gradual rule, based on the GGMP (section III-C). The gradual rules and their parameters have to be set up regarding the learning dataset (section IV).

Finally, the end user presents the process parameters to the system and gets back the materials properties that should be reached. The inference is performed, obviously, by using the GGMP (section V) which is able to perform the prediction from both singleton and imprecise values.

III. Background

In this section, we give the necessary background to understand this approach.

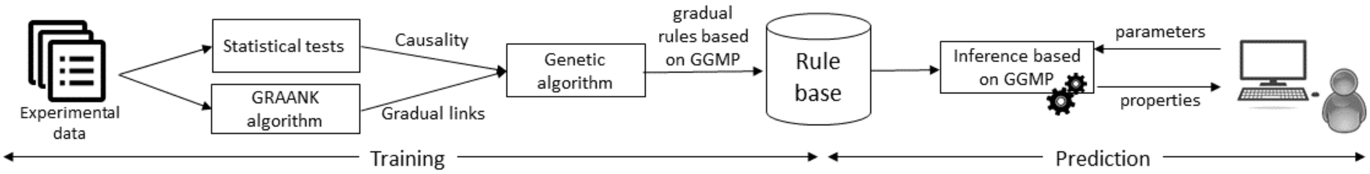

Fig. 1: Approach overview

Our work is based on fuzzy logic, which is a multivalued logic based on the fuzzy set theory. A fuzzy set is defined by a domain and a membership function that maps each element of the domain to a value between 0 and 1, expressing the membership of this element to the set.

A. Causality

Several approaches, introduced in the literature, use statistical analysis for defining the significance of the manufacturing processing parameters. The Analysis of Variance (ANOVA) and the statistical regression analysis are among the most exploited methods in materials science. For instance, in [3], ANOVA was performed to determine the level of statistical significance of the parameters. Then, a regression model based on a multi variable polynomial regression was developed.

Based on our literature study, without loss of generality, we have chosen to use ANOVA for categorical parameter, when its assumptions of normality, homogeneity of variance and independence of observations are met. A non-parametric statistical test, such as the Kruskal-Wallis test, is used otherwise. For continuous parameters, we use the statistical regression analysis to evaluate their influence on each property.

B. Graduality

The gradual dependencies were introduced to consider the common variations among attribute values. Many methods were introduced in the literature for mining gradual dependencies such as [4] where contingency diagrams are used to model fuzzy sets and linear regression analysis is applied to validate relationships between attributes. In [5], the precedence graph is used to represent the data and to detect the gradual dependencies between attributes. In [2], the GRAdual rANKing (GRAANK) is proposed by combining different existing approaches to benefit both from semantic quality and computational efficiency. Based on Kendall's tau ranking correlation coefficient, the authors evaluate candidate gradual itemsets using their concordance matrix.

For its various advantages, we use GRAANK in our work. We now present some basic notations and definitions as defined in [2].

Definition 1 (Gradual item): A gradual item A^* is defined as a pair of an attribute A associated to a variation $* \in \{\geq, \leq\}$. $A^{\geq}$ expresses an increase in A values "the higher A". $A^{\leq}$ expresses a decrease in A values "the lower A".

For instance, the gradual item $Price^{\geq}$ is interpreted as "the higher the price".

Definition 2 (Gradual itemset): A gradual itemset $GM = A_1^* A_2^* \dots A_n^*$ is a combination of n gradual items. It implies a simultaneous change between n attributes.

For example, the gradual itemset "the higher the quality of the product and the higher its price" can be formalised by the two items: $Quality^{\geq}$ $Price^{\geq}$.

Definition 3 (Gradual rule): A gradual rule GR noted as: $GM_1 \rightarrow GM_2$, is defined as a pair of gradual itemsets GM_1 and GM_2 that have to be related by a causality link. GM_1 is the antecedent of the rule GR and GM_2 represents its consequent. This causality constraint makes the difference between a gradual itemset and a gradual rule.

For example, the gradual rule, "the faster the car, then the greater the fuel consumption" can be expressed as $Speed^{\geq} \rightarrow Consumption^{\geq}$.

C. Gradual Generalized Modus Ponens

In this work, we study specifically the GGMP [6] as a way to approximate monotonous functions, and thus gradual relations between parameters and properties.

In classical logic, the Modus Ponens is used to perform inference. Its fuzzy counterpart is the Generalized Modus Ponens (GMP) that allows approximate reasoning: inference can be performed on either fuzzy sets or singletons. The special feature of the GGMP is the integration of the graduality in the inference mechanism. It is an extended version of the GMP that allows integrating the gradual hypothesis when it exists, i.e, when a monotonic relationship exists between the input and the output.

For example, for a rule: "if the price is expensive then the quality of the product is high", a gradual hypothesis can be defined by "the more the price then the more the quality of the product".

The GGMP has the same property as the GMP: if the input is a fuzzy set, the result will be a fuzzy set, and if the input is a singleton, the result will be a singleton. This will allow predicting the properties of the materials even when the parameters are set approximately.

We now describe how the GGMP works. It is not strictly necessary to understand the remainder of the article.

In order to consider the graduality, the GGMP approach is based on decomposing the fuzzy sets of the premise and the conclusion of each fuzzy rule into three parts: Smaller, Greater and Indistinguishable.

For example, for an increasing GGMP, let X and Y be two linguistic variables defined on the universes of discourse U and V respectively. A is a fuzzy set of X and B a fuzzy set of Y. We define the fuzzy rule R: "if X is A then Y is B". The universe of the premise (resp. of the consequence) is decomposed into three fuzzy sets regarding the fuzzy set A (resp. the fuzzy set B): *SmallerA*, *GreaterA* and *IndistinguishableA* (resp. *SmallerB*, *GreaterB* and *IndistinguishableB*). Using this decomposition, the inference is focused on the corresponding parts of A and B. Thus, to infer the *SmallerB* (respectively *GreaterB* and *IndistinguishableB*) part of B, the *SmallerA* (respectively *GreaterA* and *IndistinguishableA*) part of A is only used [6].

In GGMP, A and B have to be convex, normalized and continuous, and their supports have to be bounded. The definition of the membership functions of the three parts, Smaller, Greater and Indistinguishable of both the premise and the consequence of R, is based on the core and the complement of A and B [6]. Let the interval $[A_L, A_R]$ be the core of the fuzzy set A where $\forall x \in [A_L, A_R]$, $\mu_A(x) = 1$, the membership functions for the Smaller, Greater and Indistinguishable parts of A are defined as follows:

$$\phi Smaller_A(x) = \begin{cases} \overline{\mu_A(x)} = 1 - \mu_A(x) & if\ x < A_L \\ 0 & otherwise \end{cases} \quad (1)$$

$$\phi Greater_A(x) = \begin{cases} \overline{\mu_A(x)} = 1 - \mu_A(x) & if\ x > A_R \\ 0 & otherwise \end{cases} \quad (2)$$

$$\phi Indistinguishable_A(x) = \begin{cases} 1 & if\ x \in [A_L, A_R] \\ 0 & otherwise \end{cases} \quad (3)$$

GGMP has been defined for rules with multiple inputs and for monotonic decreasing relations. For more details about the GGMP, we invite the reader to refer to [6], [7].

IV. Gradual Rules Extraction from Data

As in many types of fuzzy rules, the universe of discourse has to be partitioned. Thus, to set a gradual rule based on GGMP, the following parameters have to be set.

First, we have to consider the level of granularity of each linguistic variable included in the rule. Specifically, the number of fuzzy sets partitioning the premises and the consequence universes of discourse are chosen. The number of fuzzy sets has to be defined either according to the number of data, or to the complexity of the shape of the function to approximate.

For example, let us consider the gradual rule RG: "The higher the pressure then the thicker the material". The complexity of the curve representing the material thickness values regarding the pressure is used to define the number of fuzzy sets of the two linguistic variables.

We assume that at least one triangular fuzzy set is required for a monotonic linear curve. Furthermore, in order to preserve the membership functions requirements defined in the GGMP mechanism, two half-triangles have to be added in the final partitioning, at both ends of the universe of discourse. The more complex the shape of the curve, the higher the number of fuzzy sets needed to represent the corresponding data and its graduality aspect as defined in the rule.

Once the partitioning is set, we can consider a first set of parameters: the location of the critical points of the triangular membership functions of the rule's premise and consequence fuzzy sets. More precisely, we are interested in determining the positions of the vertex of each membership function. We assume that the half-triangular membership functions have each one fixed vertex position corresponding to one extrema of the universe of discourse.

Then, we introduce a second set of parameters that measures the contribution weight of each variable in the premise of the rule in the inference of its the consequence. This is particularly useful to consider an asymmetric contribution of the different input variables. Considering these weights allows detecting the optimised contribution of each manufacturing parameter that yields to the best property values prediction.

Finally, a genetic algorithm optimises these last two sets of parameters. The size of the individuals depends on the number of the premises in the rule and the number of the fuzzy sets identified for the premises and the consequence. Moreover, we add a constraint on the genes corresponding to the vertex positions to preserve the ascending order of the fuzzy sets vertex positions of each linguistic variable. For this, we start with defining population of genes respecting this constraint.

To sort the individuals, we use as fitness function the Root Mean Square Error (RMSE) value between the predicted values and the real ones.

V. Prediction of Properties from Parameters

Let a fuzzy rule R: "if X is A then Y is B". Let A' be a new observation for X, A' being a fuzzy set. The conclusion's membership function μ'_B is defined as follows [6]:

$$\forall y \in V, \mu_{B'}(y) = \sup_{x \in \psi(y)} \top(\mu_{A'}(x), I(\mu_A(x), \mu_B(y))) \quad (4)$$

where $\top$ is a t-norm, I is a fuzzy implication operator, and

$$\psi(y) = \{x \in U | \phi P_A(x) = \phi P_B(y)\ \ and\ \ \phi P_B(y) > 0, \\ P \in \{Smaller, Greater, Indistinguishable\}\} \quad (5)$$

where ϕP are defined by equ. 1 to 3 .

In fuzzy logic, a t-norm (unabbreviated triangular norm) is a binary operation that stands for the intersection of fuzzy sets or the conjunction in logic.

In the case of multiple crisp inputs and rules, and considering the weights we added in the definition, the equation proposed in [7] to obtain the aggregated and defuzzified conclusion y_f has been changed into:

$$y_f = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} w_i^j \mu_{B_i}(y_i^j) \mu_{B_i'^j}(y_i^j) y_i^j}{\sum_{i=1}^{m} \sum_{j=1}^{n} w_i^j \mu_{B_i}(y_i^j) \mu_{B_i'^j}(y_i^j)}. \quad (6)$$

where each $w_i^j \in [0; 1]$ represents the weight of the input j in the rule i; y_i^j represents the points at which $B_i'^j$ has positive membership values.

VI. Validation

In this section, we show that a properly set GGMP can be used to approximate monotonous functions. We remind that in our application, we have few data so we are not able to assess if the relation between parameters and properties is linear, polynomial, etc.

We thus tested on some toy monotonic functions. For each case, we generated a dataset with some points regularly sampled in the universe of discourse of the input variable for the training phase. Note that the actual function used to generate the datasets is not used by our method.

First, we tested the quadratic increasing function $f(x) = 2x^2$ on $[1; 100]$. To take into account the complexity of this relation, we used 6 fuzzy sets to partition the input and the output universes; 4 triangles and 2 half-triangles were used. The Mean Absolute Percentage of Errors (MAPE) value is about 9%.

Then, we tested a sigmoid relation defined by the function $f(x) = 1/(1 + e^{-x})$ on $[-5; 5]$. Given the complexity of this relation, we have chosen to use 7 fuzzy sets to partition the universe of discourse of the input and the output, that is to say we used 5 triangles and 2 half-triangles. The MAPE value is about 6.5%.

VII. Experimental Evaluation

In this section, we evaluate our method on an experimental dataset from Physical Vapor Deposition manufacturing method. More precisely, the dataset is provided from a case study aiming at producing thin films of zinc oxide by cathodic sputtering. It contains 59 experiments obtained from four experiments designed by the Taguchi method. This method selects only few values for each possible influencing parameter. In the experiments, there are five controlled process parameters. Four of them are considered as categorical and numerical: Pressure (Pr), Partial Pressure (PP), Power (P) and Scroll Speed (SS). Indeed, on the machine that has been used, it is only possible to select some values for these parameters. The last one, the Number of Passages (NP), is quantitative. Different types of materials properties are measured: mechanical, optical and physical.

As example of properties, we have the Deposition Speed (DeS). The deposition speed values range between 2 nm/min and 45 nm/min. We started with detecting parameters influencing this property values by using the graduality detection test. After applying the Kruskal-Wallis statistical test and statistical regression analysis, we deduced that DeS is more influenced by two parameters: Pr and P. We detected also two gradual itemsets using the GRAANK algorithm. Thus, we detected the following gradual link for the DeS (noted GL_1): $P^{\geq}, Pr^{\leq} \rightarrow DeS^{\geq}$.

We worked also on the film's thickness (Th) property. The thickness values vary between 7 nm and 743 nm. We deduced that Th is influenced by three parameters: SS, P and NP. We detected also three gradual itemsets using the GRAANK algorithm. Thus, we can conclude this gradual link holds for the Th (noted GL_2): $P^{\geq}, NP^{\geq}, SS^{\leq} \rightarrow Th^{\geq}$.

For each property, we extract the gradual rules from the experimental data as presented in section IV with five fuzzy sets for each linguistic variable representing the antecedent and the consequent of each rule. We used 5-fold cross validation to evaluate the performance of our approach. According to the results of the 5 folds, the prediction of the property DeS is performed with an average MAPE of 9.5% (with a standard deviation of 2.23%) and an average RMSE of 1.8 (with a standard deviation of 0.31). The prediction of the property Th is obtained with an average MAPE of 57.49% (with a standard deviation of 5.77%) and an average RMSE of 59.12 (with a standard deviation of 19.51).

VIII. Comparison with other methods

In this part, we compare the predictive performance of our approach with the performance of other predictors from the literature.

We use the same experimental dataset from Physical Vapor Deposition manufacturing method. We apply the chosen predictors to infer the values of the material properties DeS and Th based on its influencing process parameters measures. First, we use the Adaptive Neuro-Fuzzy Inference System (ANFIS). For each input, we train a model by testing several gaussian membership functions ranging from 3 to 6 and respecting fuzzy strong partitioning. We also train a model using polynomial regression method by performing different tests with several polynomial degrees ranging from 2 to 7. Then, we use the support Vector Regression (SVR) method by testing a linear and a radial basis function kernel with default hyperparameters. For ANFIS, polynomial regression and SVR, we keep the configuration that optimizes the mean and standard deviation of the metrics RMSE and MAPE. Finally, we investigate two ensemble methods: the random forest and the Extreme gradient boosting (XGBoost) with monotonicity enforced.

To compare the prediction performance of our approach to other predictors, we applied a 5-fold cross-validation using the same folds to train and test the different models. We use only singleton values for inputs, expecting singleton values for outputs, since our method is the only one to deal with more than singletons. We calculated the RMSE and MAPE metrics to evaluate the trained models for each fold. Then we considered the average and the standard deviation of the RMSE and MAPE obtained for all the folds to assess the performances of a predictor to another one.

The table I presents the results of testing the models trained using the selected methods and our method to predict the DeS values. Our method (adapted GGMP) gives the best results in terms of mean and standard deviation of RMSE and MAPE metrics. The table II shows the results of assessing the different trained models performance to predict the Th values. Our method gives the best results in terms of mean and standard deviation of RMSE. The XGBOOST model has the best performance in terms of the average of MAPE.

TABLE I: Performance of the prediction of DeS values based on Pr and P measures using the different selected predictors

	RMSE	MAPE (%)
SVR	3.61 ± 1.09	18.75 ± 7.67
Polynomial Regression	1.9 ± 0.5	11.83 ± 3.02
Random Forest	2.83 ± 1.54	12.22 ± 4.15
ANFIS	4.6 ± 3.9	18.10 ± 15.19
XGBOOST	2.62 ± 1.17	14.65 ± 9.45
Adapted GGMP	$\mathbf{1.8 \pm 0.31}$	$\mathbf{9.5 \pm 2.23}$

TABLE II: Performance of the prediction of Th values based on P, NP and SS measures using the different selected predictors

	RMSE	MAPE (%)
SVR	108.83 ± 29.09	118.31 ± 43.81
Polynomial Regression	94.19 ± 61.05	43.31 ± 22.67
Random Forest	73.5 ± 20.98	80.12 ± 38.48
ANFIS	96.5 ± 25.87	75.8 ± 32.01
XGBOOST	59.33 ± 23.23	$\mathbf{39.47 \pm 9.42}$
Adapted GGMP	$\mathbf{59.12 \pm 19.51}$	57.49 ± 5.77

IX. Discussion

The interpretation of the results allows us to draw some conclusions. Indeed, the number of fuzzy sets partitioning the premises and the consequence of the rule represents a significant hyperparameter in this approach.

An insufficient number of fuzzy sets and few experimental data limit the ability to learn an accurate model. When the data are very few, we can use almost as many partitions as different values for each parameter and property. It will learn the segments between two points. When the data are more numerous, it is important to choose a partitioning that allows good performances, depending on the precision needed by the application. In addition, if the partitioning has an impact on the learning time, it does not affect the inference time (only few memberships are used to perform the GGMP).

In the experiments above, we made different tests to select the number of fuzzy sets to use for each gradual rule. It would be interesting to automate the selection of the optimal number.

We have also checked the gradual rules extracted from the data with materials scientists and they confirm both the interest of our approach and the results we obtained.

The proposed approach allows extracting useful knowledge from few data. On the one hand, other methods, such as ANFIS, require large datasets and their interpretability is not always guaranteed due to the number of the generated rules [8]. On the other hand, Sugeno approaches needs to know a priori the shape of the function to approximate.

The advantage of our method is:

- it is a good predictor for properties from process parameters;
- it is directly representable with a sentence, e.g. “the more ... the more”, which is understandable by the end user;
- it extracts gradual rules that can cohabit with other types of rules to manage other relations between parameters and properties;
- it deals with singleton values as well as fuzzy sets that can represent imprecise values or intervals.

Indirectly, the gradual rules can help during the search for optimal process parameters to get some specific properties. For instance, with a rule like “the higher the temperature, the clearer the material”, it is obvious that to obtain the clearest material, the temperature must be set to its maximal value.

X. Conclusion

In this paper, we present an original use of the GGMP to approximate monotonous functions. It does not need any assumption about the shape of the monotonous function to approximate and can deal with both a single value and fuzzy sets. In this paper, we focused on the availability of few data, but without loss of generality, our approach could also work with more data. This method has thus the advantage of extracting useful knowledge from the data, interpretable by humans, instead of only providing the results as other approaches provide.

In this work, we focus only on gradual rules. We extract them from the data by detecting causality, then graduality, and finally setting up a set of gradual rules that are based on GGMP.

We successfully applied our approach to materials science and showed we are able to predict the properties of the materials from process parameters. We also compared to other models and our approach sometimes outperforms XGBoost.

Since the approach has been approved by materials scientists and since the gradual links discovered have been verified, we will continue to improve this approach by generalising to other kinds of relations.

References

[1] J.-P. Poli, H. Hajri, and L. Boudet, *Artificial Intelligence for Materials Science and Engineering*. EDP Sciences, 2021, pp. 311–336.

[2] A. Laurent, M.-J. Lesot, and M. Rifqi, “Graank: Exploiting rank correlations for extracting gradual itemsets,” in *International Conference on Flexible Query Answering Systems*. Springer, 2009, pp. 382–393.

[3] A. K. Singh and R. S. Prakash, “Simulation based two level factorial designed experiments for optimisation of selective laser sintering process,” *International Journal of Computer Aided Engineering and Technology*, vol. 1, no. 3, pp. 322–337, 2009.

[4] E. Hüllermeier, “Association rules for expressing gradual dependencies,” in *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, 2002, pp. 200–211.

[5] L. Di-Jorio, A. Laurent, and M. Teisseire, “Mining frequent gradual itemsets from large databases,” in *International Symposium on Intelligent Data Analysis*. Springer, 2009, pp. 297–308.

[6] P.-N. Vo, M. Detyniecki, and B. Bouchon-Meunier, “Gradual generalized modus ponens,” in *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2013, pp. 1–7.

[7] P.-N. Vo and M. Detyniecki, “Towards smooth monotonicity in fuzzy inference system based on gradual generalized modus ponens,” in *8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*. Atlantis Press, 2013.

[8] M. N. M. Salleh, N. Talpur, and K. Hussain, “Adaptive neuro-fuzzy inference system: Overview, strengths, limitations, and solutions,” in *International Conference on Data Mining and Big Data*. Springer, 2017, pp. 527–535.