Tipps&Tricks with G^{raf}Compounder

No. X
Approaching Practical AI-Driven Formulation in Compound Development with Large, but complex Databases

Introduction

Formulating compound materials using AI presents unique challenges distinct from applications such as spell-checking or image analysis. Language-based models like ChatGPT rely heavily on statistical relationships between words. With very large data they generate high-accuracy results as the dataset size grows (1, 2), the large the higher the accuracy. However, in material formulation, the requirements differ: Predictions must be based on logical engineering principles rather than statistical associations alone. The scarcity of comprehensive historical data, especially of regression equations describing specific correlations between ingredients and their properties, adds complexity to this task.

Moreover, effective formulation in rubber development requires simplicity and efficiency to meet manufacturing needs, aiming to minimize the number of ingredient in a formulation to optimize cost and quality. Consequently, an AI based program must be able to predict feasible and manufacturable formulations while remaining economically viable. Here, we present a strategic approach for managing large and varied datasets for compound formulation, allowing AI to generate practical results suitable for real-world applications.

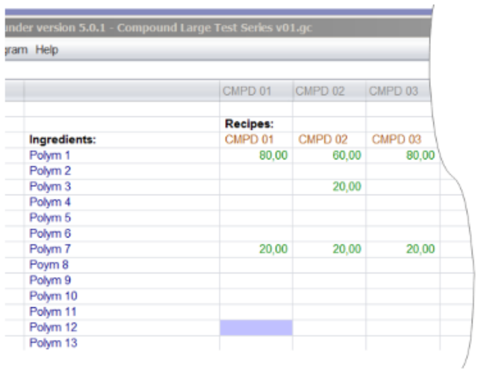

figure 1: Data set with multiple polymers

Procedure:

To illustrate the challenges of AI-driven formula development, consider a hypothetical formulation database containing a wide variety of ingredients:

1. Example Database Composition:

- **Polymers: 20 types**
- **Carbon Blacks: 20 types**
- **Whiteners: 20 types (e.g., silica, clay, calcium carbonate)**
- **Oils: 10 types (e.g., naphthenic, paraffinic)**
- **Process Aids: 20 types**
- **Anti-Aging Agents/Ozonants: 20 types**

- **Accelerators: 20 types**
- **Cross-Linkers: 5 types**

Using properties as the primary criteria for predicting a formulation, an AI model working with such large dataset may generate results with significant ingredient diversity, complicating practical use due to the number of ingredients involved. Without refining the dataset, the predicted formulations may include ingredients in minimal amounts or which are sysnonymous with others, leading to inefficient and costly manufacturing solutions. Below are three strategic methods to optimize this process for practical applicability:

2. Strategies for Practical AI-Driven Formulation:

- **<u>Database Reduction:</u> By narrowing down certain ingredient categories, we can streamline the dataset for the model. For instance, limiting the polymer selection to only two types can reduce the potential ingredient pool significantly. It can be done by setting the criteria in the "Criteria window" to zero for certain polymers. This constraint helps the AI focus on the selected essential components, reducing computational demand and enhancing the practicality of the resulting formulations.**

- **<u>Iterative Ingredient Filtering:</u> In a more labor-intensive approach, primary ingredients are prioritized, while minor ingredients are set to zero targets in the formulation. This selective filtering allows the AI to emphasize core ingredients. Although this may slightly impact the model's fitness function due to potential conflicts, the trade-off is a more manufacturable and streamlined formula. After prediction, minimal-use ingredients can be reviewed and potentially eliminated from the formulation, resulting in a simpler, more feasible rubber formulation.**

- **<u>Dataset Merging Based on Polymer Types:</u> To further refine the formulation process, separate datasets can be created for each major polymer or in combination with polymer blends. These datasets can then be merged, allowing the AI to calculate the potential impact of each polymer / -blend on the**

Criteria					
	From	To	To	To	To
Polym 1			100	100	100
Polym 2			0		100
Polym 3				0	100
Results					
MV(1+4)100°C		50	53,91	49,95	50,03
MH-ML	14	16	15,61	15,32	15,33
M100%-Mpa	25	30	2,501	2,500	2,495
M300%-Mpa		150	11,761	11,378	11,290
TS-Mpa		250	24,664	22,634	22,242
EB-%		550	544,79	523,09	524,25
H-°ShA	65	65	65,00	65,00	65,00
H-°ShA-Aged	65	65	70,97	71,35	71,16

figure *2: Property Criteria: Column "From" and Column "To"*
Criteria on Polymers:
Column 3: Polymer1 (=100) , but no Polymer 3 (=0),
Column 4: Polymer1 (=100) , but no Polymer 2 (=0)
Column 5: Polymer1 (=100) , Polymer 2 (=100) , Polymer 3 (=100)
Properties showing no big differences.

	gc-unconfirmed	gc-unconfirmed	gc-unconfirmed
Ingredients:	**Mixture1**	**Mixture2**	**Mixture3**
Polym 1	72,00		6,83
Polym 2		82,00	68,66
Polym 3			4,52
Polym 4	28,00		0,45
Polym 5		14,00	10,15
Polym 6			0,57
Polym 9			6,28
Cyclate 2		4,00	5,11
Filler 1		2,05	0,64
Filler 2	3,71	2,52	3,79
Ingredient 01	2,60	2,84	2,42
Ingredient 03	1,70	2,50	2,38
Ingredient 04	0,30	0,50	0,95
Ingredient 05			0,26
Ingredient 07	0,41	1,00	1,85
Ingredient 08		0,20	0,15
CB 01			1,05
CB 02			2,81
CB 04		30,00	48,93
Cb 05	45,00	22,00	1,09
Ingredient 11	0,14	4,50	1,66
Ingredient 12			1,47
Ingredient 13	10,00		0,38
Ingredient 15			0,11
Ingredient 16	2,00		0,08
Ingredient 18		0,60	1,00
Ingredient 22	0,50	1,00	0,33
Ingredient 24	1,00	0,25	0,61
Ingredient 25		0,15	0,11
Ingredient 27			0,02
Ingredient 30		2,40	4,00
Ingredient 31	11,00	5,50	4,21
Ingredient 32	1,20	2,00	1,94
Ingredient 36	0,30		0,03
Ingredient 38	0,80	0,80	0,80
Ingredient 39	0,33		
Ingredient 40			0,02
Ingredient 45	0,13	0,25	0,35

figure 3*: Calculated compound according criteria given in figure 2.*
Mixture 3 seems not practical with no limits on polymers, perhaps review requires too much effort.
Mixture 1 seems okay, but sill some potential for ingredient consolidation
Mixture 2 needs more / some ingredient review compared to Mixture 1

predicted properties. By holding the target properties constant and varying only the polymer composition, the AI can assist in identifying optimized polymer blend ratios that align with specific performance goals. However, while AI can perform calculations and predict potential outcomes based on data inputs, the results must be interpreted and evaluated by human experts. This approach provides a structured means for exploring material influences within the AI model.

3. Conclusion

These three strategies—database reduction, iterative filtering, and dataset merging based on polymer types—offer a practical approach for managing large databases while ensuring AI-driven formula predictions that are both useful and manufacturable. By applying these methods, AI can assist in achieving engineering criteria and creating economically viable formulations, making it a valuable tool in compound development across diverse manufacturing contexts. However, human judgment remains essential to assess the practicality, manufacturability, and overall effectiveness of AI-predicted formulations. While this approach provides a structured means for exploring material influences within the AI model, final decision-making must rest with human experts to ensure the results meet real-world requirements and practical applications.

Literature:

1. **Manon Bischoff, How to teach a computer talking (Wie man einem Computer das Sprechen beibringt) GE, Sepktrum 09.03.2023**
2. **Partha Pratim Ray, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, https://doi.org/10.1016/j.iotcps.2023.04.003**

H-JG_Consulting
2024-11-11
Hans-Joachim Graf