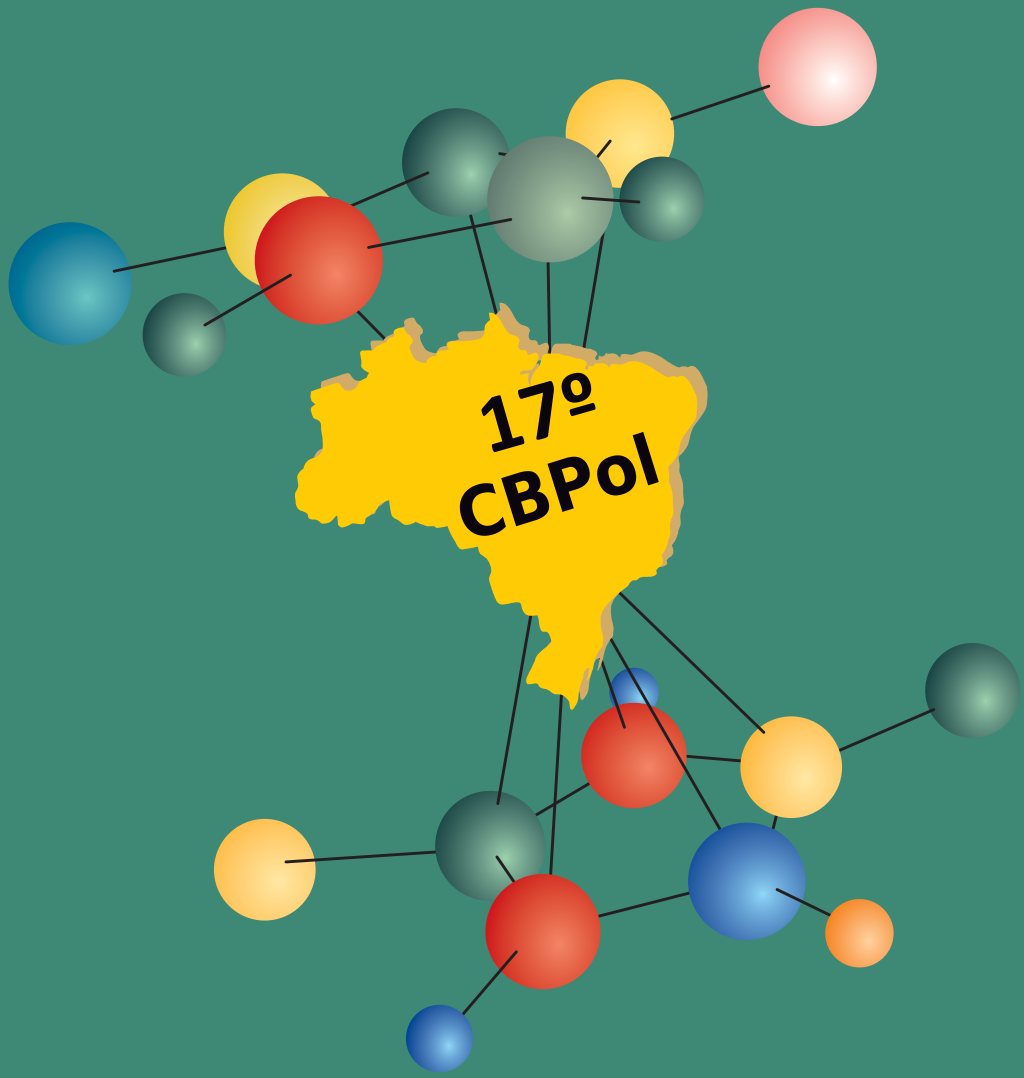

17º Congresso Brasileiro de Polímeros

29 de Outubro a 02 de Novembro de 2023

Joinville - SC

ANAIS

Associação Brasileira de Polímeros

Revisão textual e gramatical: Resposanbilidade dos respectivos autores.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Congresso Brasileiro de Polímeros (17. : 29 out. - 2 nov. 2023 : Joinville, SC)
Anais do 17° Congresso Brasileiro de Polímeros [livro eletrônico] / organização Associação Brasileira de Polímeros. -- Joinville, SC : Aptor Software, 2023.

PDF

Vários colaboradores.
ISBN 978-85-63273-55-0

1. Polímeros 2. Polímeros e polimerização 3. Química - Congressos I. Título.

24-188263 CDD-540

Índices para catálogo sistemático:

1. Química : Congressos 540

Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

USE OF ASSOCIATION RULES FOR EVALUATION OF STRESS CRACKING IN POLYMERS

Lucas M. Hallmann[1,2*], Edson L. Francisquetti[2] and Ruth M. C. Santana[1]

1 - Departamento de Materiais, Laboratório de Materiais Poliméricos (LAPOL), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil
lucas.hallmann@farroupilha.ifrs.edu.br
2 – Departamento de Materiais, Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS), Farroupilha, RS, Brasil

Abstract: This article discusses the application of data science techniques to the study of Environment Stress Cracking (ESC) in materials science, specifically in high-density polyethylene. ESC is a phenomenon where a slow and cumulative transfer of energy from a wetting agent or fluid applied to a surface causes a fracture due to the inability of the structure to absorb the energy fully. This article argues that data science techniques, particularly rule-based machine learning, can provide valuable insights into the factors and patterns of interaction involved in ESC. The study used data collected from bodies of PEAD subjected to the ASTM D1693 method, with fluids including Igepal CO-630, Monoethylene Glycol, neutral commercial detergent, Sodium Hypochlorite, and Hydrogen Peroxide. The variables observed in the experiment were fluid, pH, melting point, sample description, sample type, thickness, weight, absorption, bottle area, start and end time, total time, and fracture. After data collection, the authors used Python programming language and libraries such as Pandas, Numpy, Scikit-learn, MLxtend, Matplotlib, Plotly, and Seaborn to preprocess and analyze the data. The rule-based machine learning technique was applied to identify regular occurrences within the dataset, and the correlation between the variables was analyzed to understand the influence of factors on ESC. The authors conclude that the use of data science techniques provides a novel and replicable approach to analyzing ESC in polymers, which can help researchers gain a deeper understanding of this phenomenon without spending excessive time on manual analysis.

Keywords: High-Density Polyethylene, Rule-Based Machine Learning, Environment Stress Cracking

Introduction

The scientific investigation follows a systematic and methodical process for observing and understanding a phenomenon, with the use of tools to assist in the evaluation of the process and work with the data collected in the experimental phase. Data science is essential for exploratory data analysis and consolidation, but its application is still strongly linked to the fields of computing and statistics. It is important to use digital technologies for the collection, production, and sharing of information, as the demand to develop deep analytical skills can subtract time from research [1,2].

So, analyzing Environment Stress Cracking in polymers such as high-density polyethylene through the lens of data science becomes an interesting alternative. This approach, besides differing from conventional statistical analysis, can be replicated and shared without requiring the researcher to rework their analysis. Considering that a phenomenon is the result of factors acting under the same circumstance, producing the same result, if we analogously consider these factors as variables and their way of acting as a pattern of interaction, this situation is no different from an abstraction of mining association rules, included in frequent pattern mining, an area widely explored in Knowledge Discovery in Databases (KDD) [1,3].

In materials science, Environment Stress Cracking (ESC) is a phenomenon where a stress applied to a surface in contact with a wetting agent or fluid transfers energy to an area of that surface, slowly and cumulatively, to the point that the structure cannot absorb the energy fully, causing a rupture. This phenomenon occurs through repetitive mechanical stress, not chemical attack. Brittle fracture is characteristic of this rupture condition [4–6].

Considering this scenario, where research is immersed in data, we seek to demonstrate an alternative use of some data science techniques in the field of materials science, specifically to the study of Environment Stress Cracking that occurs in polymers such as the target of the study, high density polyethylene.

Experimental

Initially we started by obtaining the data used in the study. The way to perform the ESC test was based on the ASTM D1693 standard, using the folded strip method. In this test, specimens with dimensions recommended by the standard are taken from the flask. These are bent and placed on a U-shaped profile support. These specimens were taken from HDPE blown bottles. As for the fluids, Igepal CO-630, Monoethylene Glycol were used as standardized representatives, and commercial neutral detergent, Sodium Hypochlorite and Hydrogen Peroxide were used as representatives of domestic fluids. The samples were immersed in the fluids in baths at 50 ºC, for the time necessary for the samples to show damage [7,8].

The parameters observed in the experiment were: fluid, pH, melting point, sample description, sample type, thickness (a, b, c), perceived thickness, weight (initial, final and variation), absorption (in case of occurrence), flask area, date and time (initial and final), total time, fracture (direction, time and intensity). With the exception of the total time variable, all variables are categorical (nominal) [1].

After completing the observation stage of the experiment and obtaining the data, we started the data treatment stage. For use in the analysis, after some data transformations, the following variables were used: fluid, pH, melting point, sample type, thickness perception, absorption, total time, direction, fracture type and intensity (Fig. 1). The Python programming language was used together with Jupyter Notebook and the libraries Pandas, Numpy, Scikit-learn, MLxtend, Matplotlib, Plotly, Seaborn, all in their stable versions.

	nome	ph	ponto_fusao	tipo_amostra	espessura_percepcao	peso_resultado	sentido_fratura	tipo_fratura	intensidade_fratura
count	90	90	90	90	90	90	90	90	90
unique	5	3	2	6	2	3	4	4	5
top	Igepal	neutro	positivo	curva superior	grossa	absorveu	Transversal	Não ocorreu	zero
freq	18	54	54	15	47	69	37	35	35

Figure 1 - Overview of the base used in the analysis, in Dataframe format

Two techniques were used to analyze this data, one aiming to search for patterns to understand the fracture behavior and the other to understand the influence of factors (variables) on each other, i.e., the search for a correlation between them. The first of the techniques used is the use of association rules (rule-based machine learning), also called market basket analysis, which aims to locate regular occurrences within a dataset, in rules of the type if A (antecedent) then B (consequent). As a recurrent problem of applying this approach, we have the creation of several rules, and to separate the most significant ones, we count on statistical indexes of minimum support, minimum confidence and lift [9].

In a simplified way, we can understand support as the percentage of times that A and B occur in the base (Eq. 1). Confidence, on the other hand, represents the percentage of occurrence of B in all those of A (Eq. 2). The lift, on the other hand, calculates the chance of B occurring, if A occurs, considering the popularity of B (Eq. 3). When the lift is greater than 1, it reinforces the occurrence of the pattern. Thus, relating these indices to the Dataset itemsets, we can obtain rules (or behaviors) of great importance (relevance) [10,11].

$$supp(A) = (\#A \text{ in } B) / (\#B) \quad \textbf{(1)}$$

$$conf(A=>B) = supp(A \text{ union } B) / supp(B) \quad \textbf{(2)}$$

$$lift(A=>B)=supp(A \text{ union } B) / (supp(A)*supp(B) \quad \textbf{(3)}$$

So, we started working with the result dataset using the Apriori algorithm from the MLxtend library, aiming to search for regular itemsets. Some of the results obtained can be seen in Fig. 2. These results were obtained using a support of 0.3 (30%).

```
   itemset support >= 0.3
0.  (absorveu, positivo)
1.  (grossa, absorveu)
2.  (neutro, positivo)
3.  (absorveu, zero, 11, integro, nao ocorreu)
4.  (nao ocorreu, 11)
5.  (grossa, transversal)
6.  (microfissura, baixa)
```

Figure 2 - Itemset found with 0.3 support

Following the analysis, once we know the Itemset with the most interesting results, we implement the "*association_rules*" of the "*mlxtend.frequent_rules*" package, passing the Itemset as a parameter along with the Dataset, so we can look for frequent patterns and consequently understand the behavior presented during the experiment. In Fig. 3 you can examine some results obtained with this action.

```
   antecedents|consequents|                          support|confidence|lift
0  (microfissura)=>(baixa)                            0.31    0.87       2.71
1  (integro)=>(zero, absorveu, 11, nao ocorreu)       0.31    0.80       2.71
2  (grossa, positivo)=>(absorveu)                     0.42    0.80       2.57
3  (absorveu, 11, zero)=>(nao ocorreu)                0.31    1.00       2.57
4  (transversal)=>(grossa)                            0.32    0.78       1.50
5  (neutro)=>(positivo)                               0.44    0.66       1.11
```

Figure 3 - Some results obtained in the search for association rules

It is noticeable that some of the results of Itemset are somehow reflected in the search for rules, and after this action, we must consider whether the rules found are important to the context of the experiment. The interpretation of the obtained results should not be taken literally, but interpreted by the researcher within a specific context [11].

Results and Discussion

A possible reading of the result of rule 0 (Fig. 3), would be: "there is each micro-crack type fracture, it will be classified as a low intensity fracture with a minimum possibility of occurrence of 87%" or a less obvious rule, as the number 2: "in samples of thick thickness, exposed to fluids with positive melting point fluid absorption occurs 80% of the time". These are some of the countless possible interpretations of the many different rules generated by the use of the association rules technique.

It is also possible to concatenate other techniques and refine the understanding, such as using a correlation matrix displayed in a heatmap (Fig. 4):

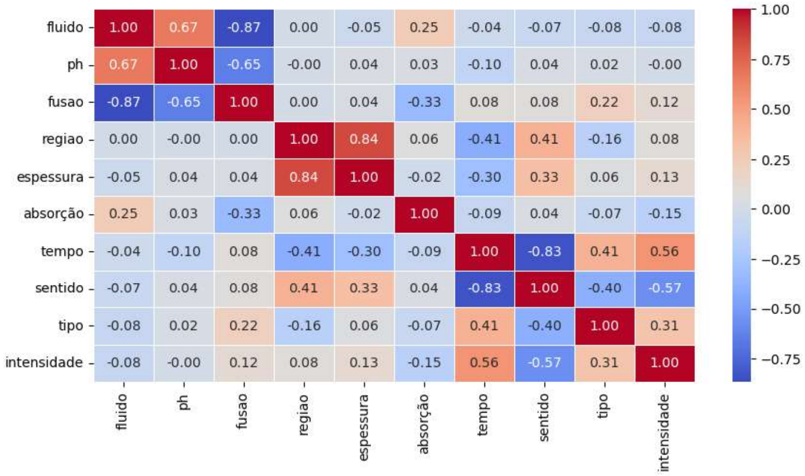

Figure 4 - Correlation matrix using Pearson's method, in the same Dataset used for the association rules.

We can see that we have a strong correlation between variables such as the thickness of the sample and the region from which it was taken, where this can imply in cross-sectional fractures in 78%, considering the inference between the patterns obtained in the association rules together with the correlation shown in the matrix.

Conclusions

Using unusual approaches to understanding a phenomenon can open new paths, especially when it comes to replicating and sharing techniques. Feeding an already created database, which can be shared among several researchers online, as well as the programming codes used for data analysis, saves time and avoids rework. This also allows for explicit assimilation of the technique and its execution, facilitating replication.

In the case of polymer behavior, it is possible to have very valuable insights by crossing the responses of the approaches and building knowledge, as for example when we realize that the region of the flask determines the direction of rupture more than the wall thickness, although transverse fractures are common in thicker samples. This kind of detail would perhaps go unnoticed if only traditional approaches to analysis were used.

Acknowledgements

The authors thank the Laboratory of Polymeric Materials at UFRGS and Materials Department of IFRS (campus Farroupilha) for supporting the research and CAPES and CNPq for the aid granted.

References

1. Rodrigues AA, Dias GA. Estudos sobre visualização de dados científicos no contexto da Data Science e do Big Data. Pesquisa Brasileira em Ciência da Informação e Biblioteconomia (PBCIB) [Internet]. 2017 Jun 30 [cited 2023 Apr 12];219–28. Available from: https://pbcib.com/index.php/pbcib/article/view/34774/

2. Brito A da C, Tenorio MB. Habilidades primárias para trabalhar com ciência de dados e big data. Revista Alomorfia [Internet]. 2019 [cited 2023 Apr 10];3:01–12.

Available from: https://www.alomorfia.com.br/index.php/alomorfia/article/view/10/6
3. Goldschmidt R, Passos E. Data Mining: um guia prático. 1st ed. Vol. 4. Rio de Janeiro: Elsevier Editora Ltda; 2005.
4. Reddy D V, Butul B. A COMPREHENSIVE LITERATURE REVIEW OF LINER FAILURES AND LONGEVITY [Internet]. 1999 [cited 2023 Apr 25]. Available from: https://archive.epa.gov/region5/waste/clintonlandfill/web/pdf/cl_044.pdf
5. Scheirs J. Compositional and failure analysis of polymers: a practical approach. John Wiley & Sons; 2000.
6. Queiroz G de C. Resistência dos polietilenos ao stress cracking ambiental (ESC - environmental stress cracking) [Internet]. 2006 Jan [cited 2023 Apr 22]. Available from: https://ital.agricultura.sp.gov.br/arquivos/cetea/informativo/v18n1/v18n1_artigo3.pdf
7. ASTM International. Standard Test Method for Environmental Stress-Cracking of Ethylene Plastics. D1693 [Internet]. 2020;15:1–11. Available from: www.astm.org,
8. Hallmann LM, Francisquetti EL, Santana RMC. Avaliação do tensofissuramento em frascos soprados de PEAD com fluídos domissanitários. In: 24 Cbecimat [Internet]. Águas de Lindóia; 2022 [cited 2023 May 8]. Available from: https://www.monferrer.com.br/METALLUM/CBECIMat2022/PROGRAMA-AF.pdf
9. Borgelt C, Kruse R. Induction of Association Rules: Apriori Implementation. In: Physica H, editor. Compstat [Internet]. Heidelberg: Physica-Verlag HD; 2002 [cited 2023 May 9]. p. 395–400. Available from: https://link.springer.com/chapter/10.1007/978-3-642-57489-4_59
10. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93 [Internet]. New York, New York, USA: ACM Press; 1993 [cited 2023 May 9]. p. 207–16. Available from: https://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf
11. Al-Maolegi M, Arkok B. An Improved Apriori Algorithm for Association Rules. International Journal on Natural Language Computing [Internet]. 2014 Feb [cited 2023 May 9];3. Available from: https://arxiv.org/abs/1403.3948