

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

YURI LAIO TEIXEIRA VERAS SILVA

CLUSTERIZAÇÃO AUTOMÁTICA EM SELEÇÃO DE
MATERIAIS E PROCESSOS DE FABRICAÇÃO
UTILIZANDO PSO-PARTICLE SWARM OPTIMIZATION

JOÃO PESSOA
2020

YURI LAIO TEIXEIRA VERAS SILVA

**CLUSTERIZAÇÃO AUTOMÁTICA EM SELEÇÃO DE MATERIAIS E
PROCESSOS DE FABRICAÇÃO UTILIZANDO PSO-PARTICLE
SWARM OPTIMIZATION**

Tese apresentada e aprovada, aos vinte e sete dias do mês de maio do ano de dois mil e vinte, como requisito à obtenção do grau de Doutor em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal da Paraíba

Orientador: Prof. Dr. Sandro Marden Torres

**JOÃO PESSOA
2020**

Catálogo na publicação
Seção de Catalogação e Classificação

S586c Silva, Yuri Laio Teixeira Veras.

CLUSTERIZAÇÃO AUTOMÁTICA EM SELEÇÃO DE MATERIAIS E
PROCESSOS DE FABRICAÇÃO UTILIZANDO PSO-PARTICLE SWARM
OPTIMIZATION / Yuri Laio Teixeira Veras Silva. - João
Pessoa, 2020.

101 f. : il.

Orientação: Sandro Torres, Lucidio Cabral.
Tese (Doutorado) - UFPB/TECNOLOGIA.

1. Metaheurística. 2. Agregado Graúdo. 3. Particle
Swarm Optimization. I. Torres, Sandro. II. Cabral,
Lucidio. III. Título.

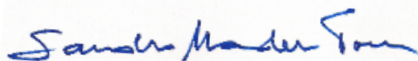
UFPB/BC

CLUSTERIZAÇÃO AUTOMÁTICA EM SELEÇÃO DE MATERIAIS E PROCESSOS DE FABRICAÇÃO UTILIZANDO PSO-ARTICLE SWARM OPTIMIZATION

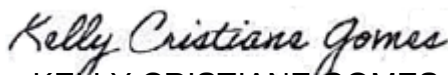
por

YURI LAIO TEIXEIRA VERAS SILVA

Tese aprovada em 27 de maio de 2020



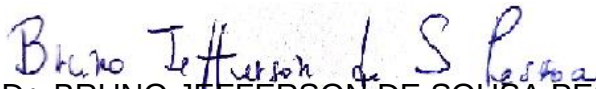
Prof. Dr. SANDRO MARDEN TORRES
Orientador - UFPB



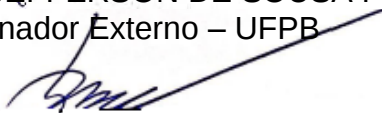
Profa. Dra. KELLY CRISTIANE GOMES DA SILVA
Examinadora Interna - UFPB



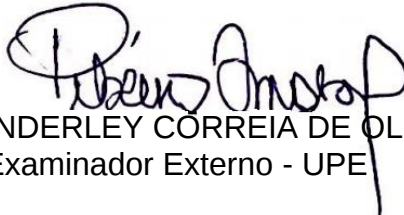
Prof. Dr. RODINEI MEDEIROS GOMES
Examinador Interno - UFPB



Prof. Dr. BRUNO JEFFERSON DE SOUSA PESSOA
Examinador Externo – UFPB



Prof. Dr. LUCIDIO DOS ANJOS FORMIGA CABRAL
Examinador Externo – UFPB



Prof. Dr. TIBÉRIO WANDERLEY CORREIA DE OLIVEIRA ANDRADE
Examinador Externo - UPE

Resumo

O volume de informações a serem coletadas e tratadas nos dias atuais é maior que a capacidade humana. A quantidade enorme de dados é uma realidade nas diversas áreas das ciências, engenharias e setores industriais. A dinâmica de aumento exponencial na quantidade de informações necessitando serem tratadas motivou a implementação de abordagens e ferramentas computacionais ao longo das últimas décadas, visando principalmente desenvolver a capacidade de tratar dados de forma automática e eficiente nas organizações. Neste sentido, o campo da ciência de mineração de dados aborda métodos e algoritmos de diferentes áreas das ciências, como pesquisa operacional, metaheurísticas e inteligência artificial. O campo das ciências e engenharia dos materiais, os materiais compostos, tais como os concretos, possuem diversas características que podem ser levadas em consideração em um método de clusterização, de modo a tornar possível a obtenção dos melhores agrupamentos possíveis por conjunto de características, conforme as métricas de desempenho estipuladas. Tais resultados podem trazer inúmeros benefícios e aplicações em decisões industriais, principalmente relacionadas aos processos produtivos e problemas de seleção de matérias-primas, podendo melhorar significativamente o desempenho operacional, reduzir custos de produção ou até mesmo impactos ambientais causados. Neste sentido, o objetivo geral do trabalho consiste no desenvolvimento de um método de clusterização para agregados de concreto da região Nordeste do Brasil, com base em suas características químicas e mineralógicas. Para tanto, desenvolveu-se uma metaheurística híbrida evolucionária visando realizar um agrupamento ótimo de dados relacionados às composições químicas e mineralógicas de tais materiais, objetivando otimizar os agrupamentos conforme as propriedades e características escolhidas, considerando ainda as restrições e métricas de desempenho associadas ao problema. Em posse dos resultados alcançados com o método, foi possível avaliar as configurações de atributos que possibilitaram uma clusterização mais eficiente em sua melhor solução, determinando os óxidos e razões mais promissores e impactantes. Além disso, o trabalho objetivou também a realização de análises de correlação entre os clusters obtidos com suas composições mineralógicas e suas densidades de probabilidades de expansão, possibilitando o estabelecimento de relações significativas entre os clusters formados e os critérios analisados. Os resultados alcançados com o método e os estudos realizados permitiram a obtenção de clusterizações de alta qualidade e eficiência (e a identificação de suas correlações com múltiplos critérios), bem como possibilitaram análises e insights inovadores acerca dos materiais e dos clusters formados, podendo assim servir de ferramenta para a tomada de decisão industrial.

Palavras-chave: Otimização por Enxame de Partículas. Algoritmo Evolucionário. Metaheurística. Clusterização. Concreto.

Abstract

The volume of information to be collected and processed today is greater than human capacity. The amount of data is a reality in the various areas of science, as engineering, industrial sectors, both in manufacturing and in the provision of services. The dynamics of exponential increase in the amount of information needing to be treated has motivated the implementation of computational approaches and tools over the last decades, aiming to develop the ability to handle data automatically and efficiently in organizations. In this sense, the field of data mining science addresses methods and algorithms from different areas of science, such as operational research, metaheuristics and artificial intelligence. The field of materials sciences and engineering, composite materials, such as concrete, have several characteristics that can be taken in a clustering method, in order to make it possible to obtain the best possible groupings per set of characteristics, as the performance metrics stipulated. Such results can bring countless benefits and applications in industrial decisions, mainly related to the production processes and problems of selection of raw materials, being able to significantly improve the operational performance, reduce production costs or even caused environmental impacts, according to the organization's objectives. In this sense, this study aims to develop an automatic clustering optimization approach that allows the determination of optimal mixtures of components and their respective attributes, so that they meet multiple optimization criteria. In this work, the criteria studied are the chemical and mineralogical compositions of rocks used as aggregates for concrete in the Northeast region of Brazil. For this, an evolutionary hybrid metaheuristic was developed in order to achieve an optimal clustering of data related to the chemical and mineralogical compositions of such materials, aiming to optimize the clustering according to the properties and characteristics chosen, considering also constraints and objectives involved in the problem. In view of the results achieved with the method, it was possible to evaluate the attribute configurations that enabled a more efficient clustering in its best solution, determining the most promising and impacting oxides and reasons. In addition, the work also aimed to carry out correlation analyzes between the clusters obtained with their mineralogical compositions and their densities of probabilities of expansion, enabling the establishment of significant relationships between the clusters formed and the analyzed criteria. The results achieved with the method and the studies carried out made it possible to obtain high quality and efficient clusters (as well as the identification of their correlations with multiple criteria), as well as enabling innovative analyzes and insights about the materials and clusters formed, thus being able to serve as a tool for industrial decision-making.

Keywords: Particle Swarm Optimization. Evolutionary Algorithm. Metaheuristic. Clustering. Concrete.

Sumário

Glossário	vii
Lista de Figuras	viii
Lista de Tabelas	x
1 Introdução	12
1.1 Definição do Tema e Justificativa	12
1.2 Objetivos	16
2 Revisão da Literatura	18
2.1 Otimização em Ciências dos Materiais	18
2.1.1 Cimento Portland	19
2.2 Clusterização de Dados	21
2.3 Medidas de Similaridade	22
2.4 Métricas de Avaliação e Validação	24
2.4.1 Métrica CS	25
2.4.2 Índice CH	25
2.4.3 Outras Métricas de Desempenho	26
2.5 Algoritmos de Clusterização	27
2.5.1 Clusterização por Particionamentos	27
2.5.2 Clusterização Hierárquica	28
2.5.3 Clusterização Baseada em Densidade	28
2.5.3.1 Clusterização Baseada em Grades	29

2.5.4	Clusterização Espectral	29
2.5.5	Clusterização Gravitacional	30
2.5.6	Clusterização por Abordagens Evolucionárias	30
3	Procedimentos Metodológicos	32
3.1	Metodologia da Pesquisa	32
3.2	Descrição Formal da Problemática de Clusterização de Materiais	34
4	Resultados e Discussões - Parte I: Formulação e Desenvolvimento da Abordagem de Otimização Multicritério	42
4.1	<i>Particle Swarm Optimization</i>	43
4.2	Representação da Solução	45
4.3	Normalização dos Dados	46
4.4	Inicialização	46
4.5	Algoritmo K-means	47
4.6	Avaliação das Soluções	47
4.7	Atualização das Partículas	49
4.8	Mecanismos de Mutação	49
4.9	Visão Geral AAKPSO	50
5	Resultados e Discussões - Parte II: Resultados Computacionais com a Aplicação da Metaheurística de Otimização Multicritério - Clusterização de Rochas com Critérios Químicos e Cristalográficos	51
5.1	Mapeamento e Estruturação das Instâncias do Problema	51
5.2	Resultados Computacionais - Instâncias Químicas	56
5.3	Resultados Computacionais - Instâncias Cristalográficas	64
6	Resultados e Discussões - Parte III: Análise e Comparação dos Resultados Alcançados por Conjunto de Instâncias	69
6.1	Análise Sistemática dos Resultados	69

6.2	Análise e Avaliação Mineralógica	73
6.3	Análise de Expansão nas Clusterizações	84
7	Considerações Finais e Trabalhos Futuros	89
	Referências	92
	Apêndice A - Banco de Dados das Britas - Análise Química - Florescência de Raios	
	X	97
	Apêndice B - Composição Mineralógica das Britas	100

Glossário

AB	:	Albita
ACO	:	<i>Ant Colony Optimization</i>
B	:	Biotita
CA	:	Calcita
CM	:	Calcita Magnésiana
DE	:	<i>Differential Evolution</i>
ED	:	Envoltória de Dados
F	:	Feldspato
GA	:	<i>Genetic Algorithm</i>
H	:	Hornblenda
LA	:	Labradorita
M	:	Mica
MI	:	Microclina
PSO	:	<i>Particle Swarm Optimization</i>
Q	:	Quartzo

Lista de Figuras

1.1	Clusterização de dados arbitrários (A) em três (B) e quatro (C) clusters . .	13
2.1	Intervalo de Temperatura para os Polimorfos do C3S	20
2.2	Clusters de densidades (a) e formas (b) distintas	22
2.3	Etapas do processo de clusterização	23
3.1	Etapas do Processo de Otimização em Clusterização no Problema Abordado	33
3.2	Ilustração do Espaço de Representação Tridimensional de Clusterização . .	35
3.3	Clusterização Arbitrária - Instância 1 - Atributos: CaO-K ₂ O-MgO	37
3.4	Clusterização Arbitrária - Instância 2 - Atributos: SiO ₂ -Fe ₂ O ₃ -Na ₂ O	38
3.5	Clusterização Instância 1 - Visão de Ternários - Atributos: CaO-K ₂ O-MgO	40
3.6	Clusterização Instância 2 - Visão de Ternários - Atributos: SiO ₂ -Fe ₂ O ₃ -Na ₂ O	40
5.1	Fluxograma de Criação das Instâncias do Problema	52
5.2	Conjunto de Instâncias Geradas a Partir dos Bancos de Dados Coletados para o Problema	56
5.3	Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Óxidos	57
5.4	Piores Soluções do Conjunto de Instâncias Químicas - Subconjunto Óxidos	58
5.5	Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões	59
5.6	Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões	60
5.7	Piores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões	61
5.8	Frequências dos Óxidos nas 20% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Óxidos	62
5.9	Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Razões	62

5.10	Menores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Razões	63
5.11	Melhores Soluções do Conjunto de Instâncias Cristalográficas - Óxidos . . .	64
5.12	Melhores Soluções do Conjunto de Instâncias Cristalográficas - Razões . . .	65
5.13	Frequências dos Óxidos nas 20% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Óxidos	66
5.14	Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Razões	67
5.15	Menores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Razões	67
5.16	Overview Melhores Soluções por Banco de Dados e Conjunto de Instâncias	68
6.1	Comparação das Maiores Frequências dos Óxidos nas 20% Melhores Solu- ções Alcançadas - Instâncias Químicas e Cristalográficas	70
6.2	Análise de Estatística de Correlação - Maiores Frequências dos Óxidos nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas .	71
6.3	Comparação das Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas	72
6.4	Análise de Estatística de Correlação - Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas .	72
6.5	Comparação das Menores Frequências das Razões nas 5% Melhores Solu- ções Alcançadas - Instâncias Químicas e Cristalográficas	73
6.6	Análise Mineralógica das Clusterizações - Melhores Soluções - Óxidos . . .	83
6.7	Análise Mineralógica das Clusterizações - Melhores Soluções - Razões . . .	84
6.8	Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 1 .	85
6.9	Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 2 .	85
6.10	Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 3 .	86
6.11	Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 1 .	86
6.12	Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 2 .	87
6.13	Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 3 .	87

Lista de Tabelas

3.1	Composição Mineralógica das Britas	34
3.2	Composição Química das Britas	34
3.3	Composição Química das Britas - Exemplo de Banco de Dados Fictício . .	36
3.4	Exemplificação de uma Instância ([CaO, MgO, Na ₂ O]) para o Problema . .	36
3.5	Instâncias Fictícias Arbitrárias para Exemplificação do Problema	37
3.6	Comparação de Clusterizações Arbitrárias da Instância Fictícia	39
4.1	Exemplo de Solução para o Problema	46
5.1	Composição Química das Britas - Método de Florescência de Raios X . . .	53
5.2	Composição Mineralógica de Extrato do Banco de Dados Coletado	54
5.3	Composição Química dos Minerais	54
5.4	Composição Química das Britas - Método Rietveld	55
6.1	Exemplo de Análise de Correlação entre a Solução de Clusterização Otimizada e suas Composições Mineralógicas - Solução da Instância [Al ₂ O ₃ , CaO e SiO ₂]	74
6.2	Instâncias de Exemplificação do Estudo de Análise Mineralógica das Clusterizações Obtidas em Suas Soluções	75
6.3	Análise Mineralógica da Clusterização da Instância [Al ₂ O ₃ , CaO e SiO ₂] .	76
6.4	Análise Mineralógica da Clusterização da Instância [CaO, SiO ₂ , e Na ₂ O] .	77
6.5	Análise Mineralógica da Clusterização da Instância [Al ₂ O ₃ , CaO e Fe ₂ O ₃] .	78
6.6	Análise Mineralógica da Clusterização da Instância [Na ₂ O, Al ₂ O ₃ e CaO] .	79
6.7	Análise Mineralógica da Clusterização da Instância $\left[\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}, \frac{\text{SiO}_2}{\text{CaO}}, \frac{\text{K}_2\text{O}}{\text{SiO}_2}\right]$. . .	80
A.1	Composição Química das Britas - Florescência de Raios X	97
A.2	Composição Química das Britas - Método de Florescência de Raios X . . .	98

A.3	Composição Química das Britas - Método de Florescência de Raios X . . .	99
B.1	Composição Minerológica de Extrato do Banco de Dados	101

Capítulo 1

Introdução

1.1 Definição do Tema e Justificativa

O volume de informações a serem coletadas e tratadas nos dias atuais é maior que a capacidade humana. A quantidade enorme de dados é uma realidade nas diversas áreas das ciências, nas engenharias, setores industriais, tanto de manufatura, como de prestação de serviços. Os acervos disponíveis em ambientes eletrônicos possuem proporções que demandam recursos computacionais cada vez maiores para viabilizar os processos decisórios. Neste sentido, as demandas da sociedade para soluções em mineração e otimização de dados surgem com uma frequência cada vez maior.

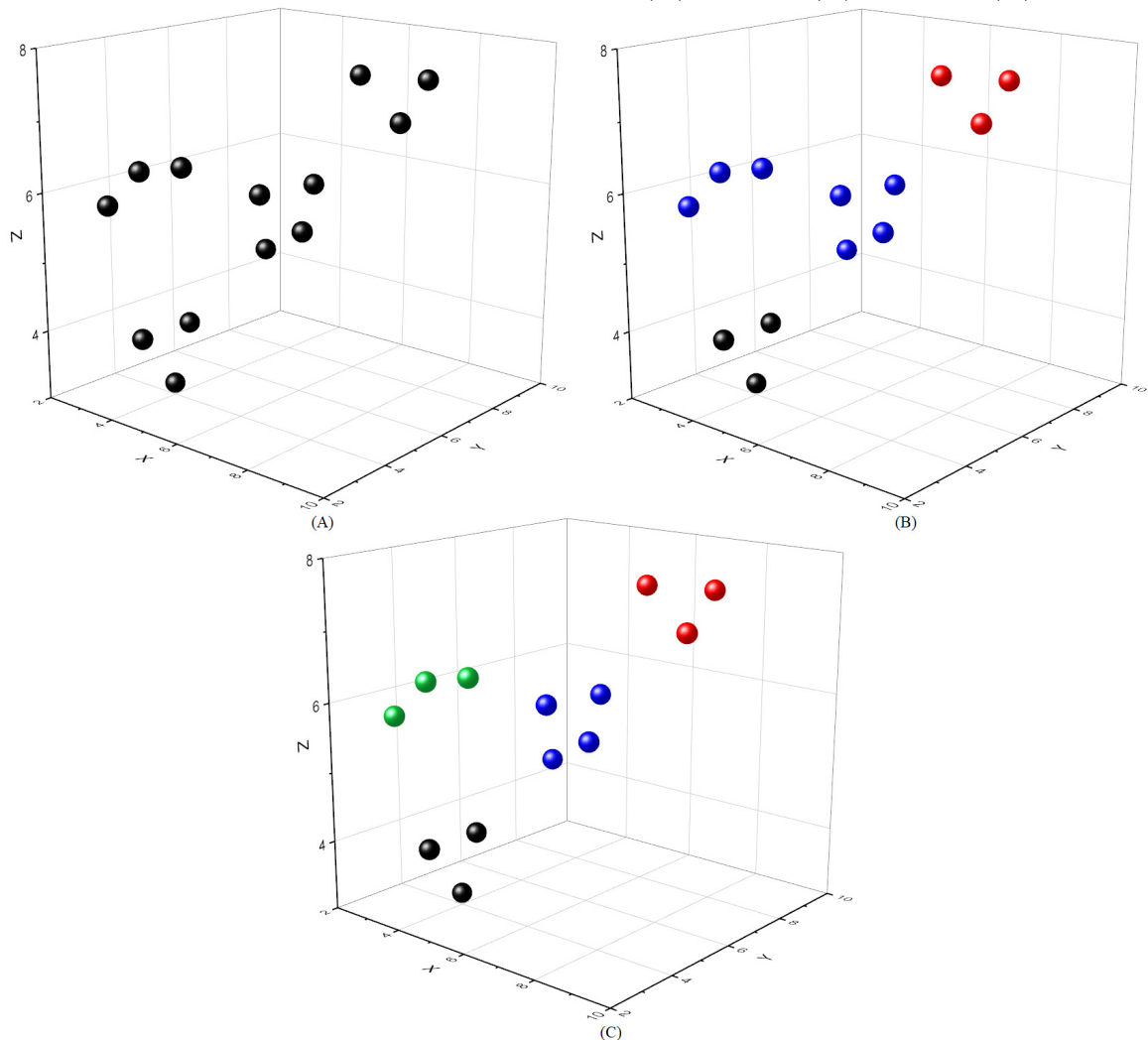
A dinâmica de aumento exponencial na quantidade de informações necessitando serem tratadas motivou a implementação de abordagens e ferramentas computacionais ao longo das últimas décadas, objetivando desenvolver a capacidade de tratar dados de forma automática e eficiente nas organizações. Muito frequentemente, cientistas e engenheiros se deparam com cenários onde existem diversas variáveis importantes que precisam ser avaliadas e agrupadas de forma a atenderem múltiplos critérios de forma otimizada. A exemplo, pode-se destacar variáveis que envolvem os impactos ambientais dos sistemas, ciclos de vida, saúde econômica das organizações, custos operacionais e processos produtivos.

A medida que a capacidade de coletar e armazenar dados das mais variadas fontes segue evoluindo nas diversas áreas do conhecimento, também aumenta a necessidade de desenvolver métodos mais eficientes para analisar e extrair informações úteis deles (SAXENA et al., 2017). O campo da ciência de dados aborda métodos e algoritmos de diferentes áreas das ciências, como pesquisa operacional, metaheurísticas e inteligência artificial, que visam descobrir informações valiosas que possam ser extraídas através dos

dados coletados.

Neste contexto, a clusterização de dados é uma das tarefas fundamentais na mineração de dados. As abordagens dessa natureza visam organizar instâncias de dados em vários conjuntos, onde cada conjunto é composto por elementos similares, enquanto diferentes conjuntos possuem elementos com características diferentes (SAJANA; RANI; NARAYANA, 2016). A Figura 1.1 ilustra o processo de clusterização de um conjunto de dados arbitrários, com três características/critérios próprios, em três (B) e quatro (C) clusters.

Figura 1.1: Clusterização de dados arbitrários (A) em três (B) e quatro (C) clusters



Fonte: Elaborado pelo autor (2020)

No exemplo apresentado, é possível perceber que a principal métrica utilizada foi simplesmente a distância Euclidiana entre os dados. Em problemáticas de clusterização de dados, métricas que envolvem o distanciamento entre os elementos são consideradas as mais comuns na literatura, especialmente devido às inúmeras possibilidades de retratar situações reais. Entretanto, há uma série de variações envolvendo outras medidas de

similaridade que visam melhor refletir realidades específicas observadas nas organizações.

Devido à natureza não supervisionada presente nas abordagens de clusterização, as ações que envolvem os processos decisórios são conhecidas por serem uma das tarefas de maior complexidade matemática e computacional na literatura, tendo sido provadas como problemas NP-Hard por Gonzalez (1982). Desse modo, é de fundamental importância que diversos fatores sejam levados em consideração na tomada de decisão sobre qual abordagem/algoritmo de clusterização a ser desenvolvido, tais como: a estruturação dos dados e das instâncias, a representação das variáveis e restrições do problema, as métricas de similaridade, as abordagens de avaliação da qualidade das soluções, entre outras questões. A forma como cada um de tais fatores serão considerados devem ser bastante estudadas e planejadas, uma vez que impactam diretamente sobre a decisão da natureza da abordagem de resolução a ser implementada, e, por consequência, na qualidade das melhores soluções a serem obtidas para o problema (GARZA-FABRE; HANDL; KNOWLES, 2017).

As aplicações de otimização em clusterização são extensas nas diversas áreas industriais, especialmente em problemas de Planejamento da Produção, Reconhecimento de Padrões e Sinais, Pesquisa de Mercado, Biotecnologia, Mineração de Dados, Agrupamento de Materiais, entre outras. Fahad et al. (2014) destacam que as aplicações envolvendo clusterização aumentaram consideravelmente nos últimos anos, especialmente devido ao crescente volume de dados produzidos pelas organizações, bem como em função de novas problemáticas que vem sendo desvendadas e estudadas nas indústrias.

O campo das ciências e engenharia dos materiais, por exemplo, é uma das áreas onde muitas aplicações de otimização foram propostas no últimos anos (STOYAN; CHUGAY, 2009). Os materiais compósitos, tais como os concretos, que consiste de uma mistura materiais cimentícios e seus agregados, em especial as britas, possuem diversas composições químicas (características) que podem ser levadas em consideração em um método de clusterização. Isto é, os valores assumidos em tais características pelos materiais são fundamentais no processo decisório de otimização sobre quais são os agrupamentos ótimos. Dado que uma abordagem de clusterização pode considerar diversas configurações distintas envolvendo as escolhas de tais características, há uma série de configurações possíveis que podem ser tratadas no processo de otimização, resultando em agrupamentos ótimos possivelmente distintos para cada conjunto de características consideradas.

Considerado o material mais utilizado no mundo, o cimento é considerado um material altamente versátil, notável principalmente por suas propriedades mecânicas e químicas. Por sua vez, seu processo de produção requer um custo ambiental extremamente elevado

em condições normais, um dos motivos que tem levado diversos pesquisadores a buscarem alternativas mais sustentáveis, não apenas do ponto de vista ambiental, como também econômico (MANZANO et al., 2011). Deste modo, estudos que envolvem os diversos tipos de agregados que são levados em consideração em seu processo de produção possuem uma relevância significativa nos diversos setores industriais. Para além disso, abordagens de otimização podem proporcionar soluções que correlacionem as características de materiais com as probabilidades de expansão em reações álcali-agregado, possibilitando abordagens preditivas na tomada de decisão industrial.

O benefício da busca pela otimização de cada possível configuração é que, após a obtenção de todas as soluções ótimas (para cada configuração existente), é possível encontrar aquelas que apresentam uma melhor qualidade de agrupamento, conforme a métrica de qualidade e avaliação adotada. Tais resultados podem trazer inúmeros benefícios e aplicações em decisões industriais, principalmente relacionadas aos processos produtivos e problemas de seleção de matérias-primas, podendo melhorar significativamente o desempenho operacional, reduzir custos de produção ou até mesmo impactos ambientais causados, conforme os objetivos da organização.

Devido a complexidade matemática e computacional associadas à problemas dessa natureza, abordagens heurísticas e metaheurísticas têm recebido uma grande atenção da literatura nas últimas décadas. Métodos de otimização dessa natureza incluem mecanismos e procedimentos que objetivam uma busca iterativa e inteligente no espaço de busca do problema, possibilitando alcançar soluções de alta qualidade e, por vezes, a própria solução ótima, a depender a dimensão e complexidade do problema (e instância) a ser tratado. Uma série metaheurísticas de naturezas e tipologias distintas têm sido desenvolvidas e propostas na literatura objetivando identificar conjuntos e padrões em problemáticas complexas.

Neste escopo, uma das principais áreas de pesquisa refere-se as abordagens de otimização evolucionárias. Em geral, são métodos que buscam representar comportamentos, movimentos e fenômenos observados na natureza. Abordagens dessa classe aplicadas em problemas de clusterização são tratadas como métodos de duas fases distintas e bem definidas. A primeira fase é representada por um modelo que visa orientar a exploração do espaço de soluções do problema ao longo das iterações, fundamentada no comportamento da natureza inspirado no método, enquanto uma segunda etapa refere-se a abordagem responsável pelas regras, objetivos e restrições de clusterização associadas à problemática a ser tratada.

Diversos algoritmos evolucionários foram propostos para outras aplicações na literatura, como Algoritmos Genéticos, *Particle Swarm Optimization*, *Ant Colony Optimization*, entre outros. Entretanto, há uma quantidade ainda limitada de trabalhos na literatura que abordaram algoritmos dessa categoria para tratar problemas de clusterização, especialmente de abordagens híbridas, caracterizadas por unirem mecanismos e estruturas de mais de um método em uma mesma abordagem. Abordagens híbridas envolvendo algoritmos evolucionários e outras metodologias de otimização foram implementadas na literatura para problemáticas em outras áreas das ciências (DENG; LIN; LO, 2012; KUO; ZULVIA, 2013; MISHRA; PANDA; MAJHI, 2014), demonstrando resultados promissores e eficientes.

1.2 Objetivos

Neste sentido, o presente projeto de tese apresenta como objetivo geral desenvolver uma abordagem de clusterização automática que permita a determinação de misturas ótimas de componentes e seus respectivos atributos, de modo que atenderem a múltiplos critérios de otimização. Nesse trabalho, os critérios estudados são as composições químicas e mineralógicas de rochas utilizadas como agregados para concreto na região Nordeste do Brasil. Para tanto, desenvolveu-se uma metaheurística híbrida evolucionária visando realizar um agrupamento ótimo de dados relacionados as composições químicas e mineralógicas de tais materiais, visando otimizar o agrupamento conforme as propriedades escolhidas, considerando ainda as restrições e os objetivos envolvidos no problema, bem como sua correlação de forma preditiva as probabilidades de expansão dos materiais em reações álcali-agregado.

De modo a viabilizar o atingimento do objeto geral em questão, foram propostos objetivos específicos para o estudo:

- Avaliar a aplicabilidade e eficiência de métodos de otimização aplicáveis ao problema de clusterização em materiais de múltiplos parâmetros;
- Formular os critérios matemáticos, computacionais e de engenharia que representem o problema tratado e suas peculiaridades, de modo a fornecer os subsídios necessários para a determinação das soluções ótimas de clusterização, incluindo os seus ranqueamentos, fundamentados em suas composições químicas, mineralógicas e de suas razões;

- Desenvolver uma metaheurística híbrida de otimização, fundamentada em algoritmo evolucionário, que permita otimizar e compreender a influência de composições químicas, mineralógicas e suas razões no processo de clusterização dos agregados;
- Aplicar o método desenvolvido e avaliar os resultados quanto à sua eficiência e desempenho matemático e computacional nas clusterizações alcançadas, bem como se há indicação da prevalência de algum elemento químico (ou conjunto de elementos) que determinem a formação das melhores clusterizações no ranqueamento.
- Analisar se há correlação, e de que forma ocorre, entre os agrupamentos obtidos para critérios químicos e cristalográficos com as composições mineralógicas, bem como com a probabilidade de expansão causada pela reação álcali-agregados.

Capítulo 2

Revisão da Literatura

Avaliar a aplicabilidade e eficiência de métodos de otimização aplicáveis à problemática de clusterização em materiais de múltiplos parâmetros.

2.1 Otimização em Ciências dos Materiais

Na era digital atual, com a crescente evolução e desenvolvimento das novas tecnologias nos mais diversos setores industriais e de serviços, dados gerados por máquinas e dispositivos têm sido cada vez mais utilizados nas decisões estratégicas e operacionais, tais como: soluções de gerenciamento do ciclo de vida, otimização da cadeia de suprimentos, desenvolvimento de novos produtos inovadores, sistemas de planejamento de produção e scheduling, bem como diversas outras aplicações. Os sistemas atingiram um volume enorme dados e a tendência é de um crescimento ainda mais acentuada nos próximos anos (BENABDELLAH; BENGHABRIT; BOUHADDOU, 2019).

Em busca de melhores desempenhos organizacionais e vantagens competitivas sustentáveis, as organizações estão buscando gerenciar o conhecimento, através do uso de informações corretas no momento certo e no formato certo. Pode-se dizer que o gerenciamento do conhecimento é uma questão crucial não só para o crescimento das organizações atuais, mas para sua sobrevivência nos mercados (DAVENPORT; PRUSAK, 2000). Neste sentido, as abordagens fundamentadas em mineração de dados estão sendo cada vez mais demandadas pelas empresas, quaisquer que sejam as áreas de suas atuações.

Ferramentas de otimização surgem também em diversas aplicações distintas dentro das ciências dos materiais, especialmente quando se trata de problemas de clusterização e empacotamento. Abordagens dessa natureza foram implementadas para representar a estrutura de líquidos e materiais vítreos (STOYAN; CHUGAY, 2009), estudar fenômenos como condutividade elétrica, fluxo de fluidos, distribuição de tensões e outras propriedades mecânicas de materiais granulares, assim como para investigar processos como sedimentação, compactação e sinterização (WILLIAMS; PHILIPSE, 2003).

Além disso, há diversas aplicações de abordagens similares envolvendo engenharia de materiais e medicina concomitantemente, como por exemplo, para planejamento de tratamento radio-cirúrgico (WANG, 1999). Em metalurgia do pó, outra área bastante estudada nas ciências dos materiais, podem ser vistas diversas aplicações para o problema de corte a laser tridimensional (GAN et al., 2004). O mesmo pode ser dito para organização e compactação de ambientes (BIRGIN; MARTINEZ; RONCONI, 2005), corte de diferentes cristais naturais, no layout estrutural de estruturas, bem como para diversas outras aplicações, conforme relatado por Hifi e Yousef (2018). Jaeger, Nagel e Behringer (1996), por exemplo, utilizaram um modelo numérico para explicar a importância do empacotamento no fluxo e compressão de materiais granulares, isto é, de que modo isso permite obter determinadas propriedades e características em tais materiais.

Recentemente, o problema da melhor proporcionalidade e clusterização possível de agregados e sua contribuição para a otimização do mix de concreto tem sido objeto de bastante discussão e realização de estudos na literatura (SOBOLEV, 2004; SOBOLEV; AMIR-JANOV, 2010). Convencionalmente, os agregados são selecionados e agrupados seguindo regras de definição estabelecidas para o mix. A seleção e a combinação dos agregados apresentam uma enorme influência na qualidade e preço dos concretos obtidos, o que os tornam alguns dos principais motivadores da atenção que tal problemática tem recebido por pesquisadores e organizações.

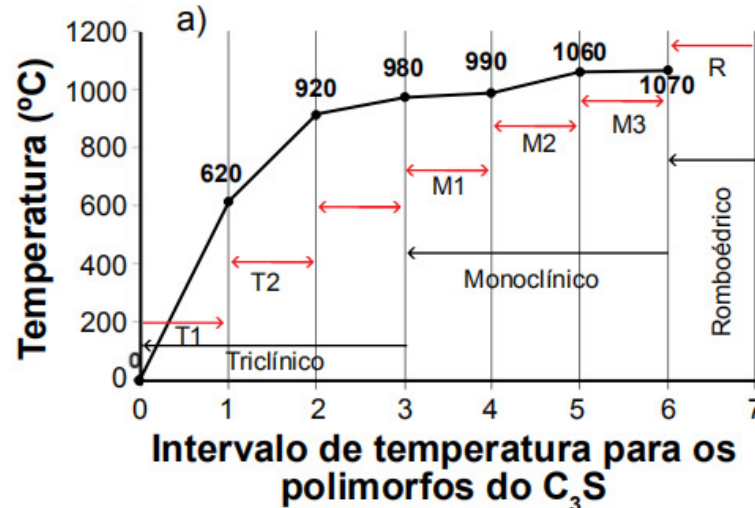
2.1.1 Cimento Portland

Considerado o material mais utilizado no mundo, o cimento é considerado um material altamente versátil, notável principalmente por suas propriedades mecânicas e químicas. Por sua vez, seu processo de produção requer um custo ambiental extremamente elevado em condições normais, um dos motivos que tem levado diversos pesquisadores a buscarem alternativas mais sustentáveis, não apenas do ponto de vista ambiental, como também econômico (MANZANO et al., 2011).

Neste contexto, a alita é o principal componente do cimento Portland, tendo uma representatividade de cerca de 70% do material. Principal constituinte presente no clínquer, a síntese do C3S dificilmente gera uma fase com apenas um polimorfo, sendo sete os polimorfos possíveis para os três sistemas cristalinos: triclinico, monoclinico e romboédrico (REN; ZHANG; YE, 2017). Na Figura (2.1) a seguir, são apresentadas as estruturas obtidas de acordo com as faixas de temperatura, onde a estrutura triclinica é obtida com temperaturas até 980 °C, a monoclinica pode ser obtida para temperaturas intermediárias, até

1070 °C e, por fim, a estrutura romboédrica é alcançada por processos com temperaturas que atinjam o patamar acima de 1070 °C.

Figura 2.1: Intervalo de Temperatura para os Polimorfos do C₃S



Fonte: (REN; ZHANG; YE, 2017)

Diversos estudos foram realizados com o intuito de identificar possíveis correlações entre várias propriedades associadas aos materiais, especialmente seus raios atômicos, eletronegatividade e estado de valência. Todavia, conforme exposto por Fernandes (2019), ainda há um enorme gap na literatura com relação a quais os principais efeitos da presença/inclusão de determinados íons em uma estrutura cristalina objeto de estudo. Apesar disso, sabe-se que o nível de incorporação de tais elementos impactam diretamente na estabilização de tais polimorfos (ÁNGELES et al., 2008). Fernandes (2019) discute ainda sobre as consequências que podem trazer a utilização de determinados combustíveis no processo de produção, enfatizando que a depender do combustível, há uma série de elementos que podem associar-se a estrutura cristalina, vindo a contribuir para que sejam obtidos resultados não esperados na obtenção do clínquer.

Manzano et al. (2011) destacam ainda que, embora que com estudos extremamente limitados neste campo de aplicação, as técnicas de simulação podem ser ferramentas críticas para superar dificuldades como as mencionadas, uma vez que permitem determinar de forma eficiente as relações envolvendo impurezas químicas, bem como examinar as propriedades resultantes no material objeto de estudo. Desse modo, abordagens de otimização bem definidas podem contribuir de uma forma extremamente relevante para a resolução de problemas dessa natureza, possibilitando o preenchimento de um gap de fundamental importância no estudo das ciências dos materiais, em especial, na otimização dos processos clusterização e produção de materiais.

2.2 Clusterização de Dados

Por definição, clusterização se trata de uma tarefa não supervisionada de mineração de dados, isto é, um processo mineração onde não é considerado conhecimentos prévios acerca dos dados que serão tratados. Abordagens dessa classe encontram um volume enorme de aplicações na análise exploratória de dados. Via de regra, objetiva-se organizar dados em um conjunto finito de categorias (pré-definidas ou não), fundamentados nas estruturas e características inerentes aos dados, na esperança de que relacionamentos significativos e previamente desconhecidos entre os eles surjam como resultado do processo.

A falta de uma definição única e adotada em todo o globo para o termo cluster na literatura impulsionou o desenvolvimento de vários paradigmas de clusterização, assim como um grande volume algoritmos referente as naturezas das problemáticas (SILVA; HRUSCHKA; GAMA, 2017; JASKOWIAK et al., 2016). Tais abordagens vem sendo aplicadas em diversas áreas das ciências, conforme destacado na Seção 2.1, como engenharias, medicina, física, computação, economia, dentre outros (ÁNGELES et al., 2008; AIDONG, 2006; DAS; ABRAHAM; KONAR, 2008).

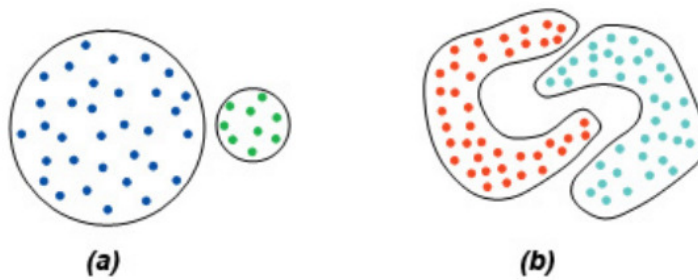
Desde que diversos problemas de clusterização de dados foi provado ser *NP – Hard* (GONZALEZ, 1982), tais paradigmas e subproblemas foram sendo propostos com cada vez mais intensidade, como, por exemplo: *Partitional Clustering*, *Hierarchical Clustering*, *Density-based Clustering*, *Grid-based Clustering*, *Correlation Clustering*, *Spectral Clustering*, *Gravitational Clustering*, dentre outros (WONG, 2016), os quais são melhor discutidos na Seção 2.5.

Em problemas que envolvem a necessidade de clusterização de dados, a primeira etapa é escolher a natureza do algoritmo de clusterização adequado para sua resolução, assim como definir seus parâmetros de maneira eficiente. O processo decisório de escolha de um algoritmo específico e sua respectiva parametrização correspondente não é uma tarefa trivial em um ambiente não supervisionado de mineração de dados. Na última década diversos fatores matemáticos foram elaborados visando orientar o processo de tais escolhas de maneira quantitativa e objetiva (NGUYEN; WOON; NG, 2015). Na literatura em questão, esses índices são geralmente referidos como critérios de validade de agrupamentos, pois também podem ser usados para avalia a qualidade dos resultados do agrupamento sob uma perspectiva estatística do problema.

Os conjuntos de dados a serem tratados em problemas de clusterização podem ser apresentados de diferentes formas e densidades. Desse modo falhas na compreensão de

alguns desses atributos pode tornar possível a obtenção de resultados inválidos. Determinadas abordagens de clusterização podem não ser capazes de detectar clusters como os apresentados na Figura 2.2, de densidade e/ou volume (a), e de formas côncavas (b). Assim, alguns algoritmos podem ser eficientes na obtenção de clusters com determinadas características pré-definidas (conforme sua aplicação e seus atributos), ao mesmo tempo em que podem ser ineficientes para outros problemas com características e/ou aplicações distintas.

Figura 2.2: Clusters de densidades (a) e formas (b) distintas



Fonte: Karypis, Han e Kumar (1999), Oliveira (2008).

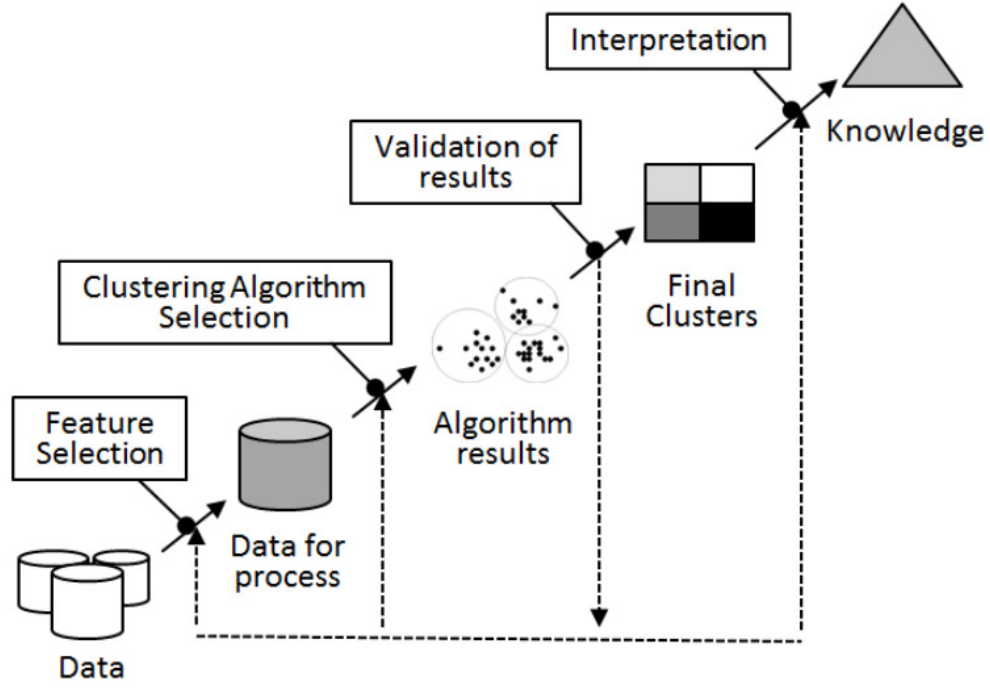
Durante o processo de clusterização, há quatro principais etapas: Coleta de dados e seleção das variáveis do problema, medidas de similaridade, metodologia de clusterização e uma etapa final, referente a avaliação e validação dos resultados alcançados. A Figura 2.3 a seguir ilustra de maneira mais clara uma visão geral do processo.

Conforme destacado anteriormente na Figura 2.3, há quatro principais etapas no processo de clusterização de dados. As seções a seguir descrevem detalhadamente as características das etapas, assim como os principais trabalhos relacionados, respectivamente.

2.3 Medidas de Similaridade

As medidas de similaridade e avaliação dos resultados alcançados são primordiais para a efetividade da abordagem de clusterização na resolução do problema. Clusters são formados a partir de elementos que possuem maior similaridade entre si, do que com elementos de outros clusters. Para que a definição dos clusters seja um processo bem sucedido, é necessário primeiramente definir as métricas que serão utilizadas para investigar a similaridade entre dois elementos do banco de dados a ser tratado. De maneira geral, leves distanciamentos entre os elementos pode significar uma grande similaridade, portanto, a medida de distância pode ser usada para quantificar a dissimilaridade, por exemplo, inclusive podendo ser considerada de diversas formas distintas. Conforme rela-

Figura 2.3: Etapas do processo de clusterização



Fonte: Halkidi, Batistakis e Vazirgiannis (2011).

tado por Nguyen, Woon e Ng (2015), Sousa (2010), além da distância e suas variantes, outros fatores também podem ser levados em consideração.

Neste contexto, a maneira mais popular na literatura de avaliar a similaridade entre elementos de um banco de dados utiliza de uma medida de distância, sendo a distância Euclidiana, um caso especial da métrica Minowsky (JAIN; MURTY; FLYNN, 1999), a mais comumente encontrada. Seguindo a métrica, dois elementos x_i e x_j tem sua distância $D_{i,j}$ estabelecida na Equação (2.1).

$$D_{i,j}^\lambda = \left(\sum_{d=1}^{d_{max}} (x_{i,d} - x_{j,d})^\lambda \right)^{\frac{1}{\lambda}} \quad (2.1)$$

A distância Euclidiana representa um caso especial, quando $\lambda = 2$. Deste modo, a distância Euclidiana é dada por:

$$D_{i,j} = \sqrt{\sum_{d=1}^{d_{max}} (x_{i,d} - x_{j,d})^2} \quad (2.2)$$

Tais métricas podem, a depender da aplicação a ser considerada, não ser eficientes para agrupar dados de grandes dimensões, uma vez que a distância entre os padrões aumentam significativamente com o crescimento da dimensionalidade. Com isso, os conceitos de próximo ou distante se tornam mais sensíveis e propensos a equívocos interpretativos, conforme relata Das, Abraham e Konar (2008). Além disso, os recursos em larga escala tendem a dominar sobre os outros recursos, podendo gerar resultados inviáveis ou ineficientes. Uma estratégia de tratamento para evitar tais problemáticas pode ser aplicada utilizando mecanismos de normalização dos dados em um intervalo comum e conhecido.

Através de um processo de normalização as distâncias obtidas que serão utilizadas posteriormente nas métricas estão sujeitas a menores falhas no processo decisório do método, especialmente quando a aplicação do problema tratado possui uma complexidade significativa e/ou quando sua dimensionalidade é consideravelmente elevada, seja pelo volume de elementos ou pela quantidade de atributos considerados no problema.

2.4 Métricas de Avaliação e Validação

Diversos autores na literatura sugeriram índices para facilitar a avaliação da similaridade nos clusters, e consecutivamente a qualidade das soluções obtidas para a tomada de decisão (HALKIDI; VAZIRGIANNIS; BATISTAKIS, 2000). A análise do distanciamento que envolvem os clusters externos, bem como entre seus elementos internos, são os dois principais fatores que envolvem a avaliação e a validação da qualidade dos clusters alcançados. Apesar de não haver um grande número de fatores, há uma enorme variação de possibilidades matemáticas e estatísticas em suas considerações para o objetivo dos problemas.

Do ponto de vista da análise externa aos clusters, elas visam comparar o resultado do agrupamento com um resultado de referência que é considerado uma premissa básica inerente à aplicação do problema. As análises que consideram o segundo fator, buscam avaliar o agrupamento apenas com o resultado em si, ou seja, a estrutura dos clusters encontrados e suas relações. Desse modo, segundo Benabdellah, Benghabrit e Bouhadou (2019), na situação em que não há um volume significativo de informações externas disponíveis, tais métricas internas são as únicas opções para a obtenção e validação dos clusters no processo.

2.4.1 Métrica CS

Uma das mais recentes e eficientes métricas reportadas na literatura, o índice CS proposto por Chou, Su e Lai (2004) possui um mecanismo de mensuração que avalia de forma inteligente o agrupamento internos aos clusters, bem como o espaçamento entre os clusters formados, possibilitando ponderações de acordo com a aplicação. O mecanismo é fundamentado da seguinte forma: Seja $D(x_i, x_j)$ o distanciamento entre os elementos x_i e x_j , conforme a Equação (2.2). O numerador da equação visa mensurar o agrupamento intracuster, isto é, o quão convergentes os pontos de um mesmo cluster estão. Para isso, é calculada a média entre as distâncias máximas calculadas para cada um dos cluster. Por sua vez, no denominador é mensurado o distanciamento entre os clusters. A Equação (2.3) a seguir ilustra as relações descritas, onde N_i indica a quantidade de pontos alocados no cluster i , enquanto K indica o máximo de clusters presentes na solução.

$$CS(K) = \frac{\frac{1}{K} \sum_{i=1}^K \left(\frac{1}{N_i} \sum_{x_i \in C_i} \max_{X_q \in C_i} [D(X_i, X_q)] \right)}{\frac{1}{K} \sum_{i=1}^K \left(\min_{j \in K, j \neq i} [D(m_i, m_j)] \right)} \quad (2.3)$$

2.4.2 Índice CH

O índice CH representa uma das primeiras métricas presentes na literatura para avaliar os resultados dos procedimentos de clusterização. Proposta inicialmente por Caliński e Harabasz (1974), a medida pode ser representada em função da razão entre a soma da dispersão dentro dos clusters e uma medida que representa a separação entre eles, utilizando os próprios clusters e as médias dos distanciamentos entre seus elementos. Das, Abraham e Konar (2008) definem a dispersão dentro do i -ésimo cluster, bem como a distância entre o i -ésimo e o j -ésimo, como as equações a seguir. $d_{i,j,t}$ calcula a distância entre dois pontos i e j quaisquer, dada uma dimensão t . Por sua vez, $S_{i,j,q}$ mensura o distanciamento médio intracuster para todos os clusters formados, onde N_i indica a quantidade de dados alocados no cluster i e $m_{i,d}$ representa o centróide do cluster i .

$$S_{i,j,q} = \left(\sum_{x \in C_i} \frac{1}{N_i} \sqrt[q]{\sum_{d=1}^{d_{max}} (x_{j,d} - m_{i,d})^2} \right)^{\frac{1}{q}} \quad (2.4)$$

$$d_{i,j,t} = \left(\sqrt[t]{\sum_{p=1}^{d_{max}} (m_{i,p} - m_{j,p})^t} \right)^{\frac{1}{t}} \quad (2.5)$$

Onde m_i representa o i -ésimo centróide, e q e t parâmetros do método escolhidos conforme a parametrização adequada a aplicação que será realizada. N_i representa a quantidade de elementos no cluster C_i . As relações entre $S_{i,j,q}$ e $d_{i,j,t}$ permitem avaliar, respectivamente, fatores intrínsecos aos clusters, e externos a eles, através do distanciamento entre o conjunto de clusters.

2.4.3 Outras Métricas de Desempenho

Diversas outras métricas foram propostas na literatura. Uma delas, bastante utilizada, é denominada Índice C. Dimitriadou, Dolničar e Weingessel (2002) destacam que a métrica busca obter o número ótimo de clusters através de uma equação que fundamentada na somas das menores distâncias intraclusters e das maiores distâncias considerado todo o conjunto de dados.

O Índice de Conectividade, por sua vez, trata-se de uma métrica proposta por Xing et al. (2005), e objetiva mensurar a distância entre cada elemento de um mesmo cluster e seu vizinho mais próximo. A métrica visa mensurar a qualidade dos agrupamentos através de pontos de observação alocados no espaço busca dos problemas, de modo que possam avaliar o agrupamento dos clusters através das menores distâncias mensuradas.

Uma das métricas mais antigas constatadas na literatura é denominado de Índice Gama. Proposta por Baker e Hubert (1975), a métrica visa realizar um procedimento de comparação de similaridade interna entre todos os clusters existente, bem como a dissimilaridade entre eles. Quanto maior o valor obtido pelo índice, melhor a qualidade da solução encontrada. Por sua vez, o Índice Dunn representa uma métrica similar ao Índice CS, dado que busca auferir, em uma única equação, a similaridade intraclusters, enquanto avalia o afastamento dos clusters no espaço.

A métrica de avaliação Tau é um aperfeiçoamento do Índice Gama reportado anteriormente. A métrica leva em consideração duas matrizes, onde a primeira contempla as distâncias euclidianas entre todos os elementos presentes no banco de dados do problema,

enquanto a segunda refere-se a uma matriz binária que indica um, caso dois elementos estejam em um mesmo cluster, e zero, caso contrário (WEBBER; MOFFAT; ZOBEL, 2010). Assim como no Índice Gama, quanto maior o valor obtido pela solução, melhor será a qualidade da solução obtida em termos de agrupamento.

2.5 Algoritmos de Clusterização

2.5.1 Clusterização por Particionamentos

Em abordagens fundamentadas em clusterização por particionamentos os dados são divididos em subconjuntos sem que haja sobreposições entre eles, de modo que cada instância de dados seja atribuída a exatamente um único subconjunto. Isto é, dada uma base de dados com N elementos, as abordagens de particionamento constroem K subconjuntos de dados, denominados clusters, de modo que $K \leq N$. A determinação de K pode ser fixa ou não, a depender da abordagem algorítmica utilizada na resolução do problema.

Via de regra, a lógica iterativa das abordagens de particionamento visa criar um particionamento inicial, geralmente através de heurísticas gulosas, e ir realocando, de maneira iterativa, os elementos em novos conjuntos, conforme atualizações realizadas no espaço busca, até que atinja um parâmetro de encerramento satisfatório. Em geral, a solução ideal é atingida quando os conjuntos alcançados com o particionamento estão próximo, quando analisados os clusters internamente, e distantes, quando analisado o distanciamento entre clusters distintos.

O k-means é um dos métodos baseados em particionamento mais clássicos, o qual aplica uma abordagem de refinamento iterativo com duas etapas principais. A primeira etapa é escolher os centróides dos clusters, enquanto a última fase é atribuir quais elementos (dados) do banco irão pertencer a cada centróide (baseado na distância Euclidiana). O centróide que estiver mais próximo de cada elemento, terá ele em seu cluster. A agilidade de sua convergência e relativa simplicidade de implementação computacional são fatores que levaram diversos estudos na área a adotarem o k-means como um dos métodos mais utilizados (XIONG; WU; CHEN, 2008).

Sua principal desvantagem se diz respeito a vulnerabilidade à sua técnica de propagação aleatória. Após definidas as posições iniciais dos centróides (geralmente de maneira gulosa), há grandes chances de que atinjam ótimos locais rapidamente e não consigam realizar buscas em um espaço considerável do espaço busca do problema, permitindo que,

por vezes, não seja possível alcançar resultados com a qualidade de clusterização esperada. Em função disso, é comum que métodos de particionamento tenham uma certa dificuldade em obter soluções de alta qualidade quando o banco de dados possui formas e tamanhos de alta complexidade matemática e/ou computacional. Uma variante bastante conhecida do k-means, é denominada k-medoids. A diferença é, basicamente, que cada cluster deverá ser representado pelo elemento mais próximo do que seria o centróide no algoritmo k-means. (SOUSA, 2010; WONG, 2016). Outras duas abordagens eficientes presentes na literatura são a k-modes (PARK; JUN, 2009) e PAM (KAUFMAN; ROUSSEEUW, 2009).

2.5.2 Clusterização Hierárquica

Nos algoritmos baseados em clusterização hierárquica, os clusters são formados seguindo uma abordagem de baixo para cima (bottom-up) ou de cima para baixo (top-down). No primeiro caso, inicialmente cada elemento do banco de dados é considerado como um cluster único. Posteriormente, o algoritmo vai, iterativamente, mesclando os clusters que apresentam maiores índices de similaridade, até que todos os clusters se unam e formem um cluster final, o maior nível hierárquico. Com isso, uma árvore hierárquica é construída para conectar todos os pontos de dados no final (GONZALEZ, 1982). Um determinado nível de profundidade da árvore pode ser escolhida para dividi-la em clusters.

De forma análoga, as abordagens top-down iniciam do nível mais alto possível, ou seja, com um único cluster que abrange todos os elementos do banco de dados original, e se divide em clusters menores até que cada cluster tenha um único elemento (ou que algum dos critérios mencionados anteriormente seja atingido). Há diversas variantes na literatura, que possibilitam que o algoritmo possa alcançar o final de sua execução quando uma quantidade específica de clusters é atingida (caso seja estipulada desde o início), ou até mesmo fundamentada nas métricas de desempenho consideradas na aplicação.

Os principais métodos que são fundamentados em clusterização hierárquica são: Abordagem Ward (JOHNSON, 1967), BIRCH (ZHANG; RAMAKRISHNAN; LIVNY, 1996), CURE (GUHA; RASTOGI; SHIM, 1998) e Chameleon (KARYPIS; HAN; KUMAR, 1999).

2.5.3 Clusterização Baseada em Densidade

Os elementos do banco de dados são categorizados em pontos principais, pontos de fronteira e pontos de ruído, de modo que todos os pontos principais são conectados com

base em suas respectivas densidades para formar um cluster. Com isso, clusters de formas arbitrárias são formados por vários algoritmos de cluster, como, por exemplo, DBSCAN (ESTER et al., 1996) e OPTICS (ANKERST et al., 1999). Os métodos baseados em densidade surgiram para buscar soluções de clusterização que identificassem tamanhos e formas consideravelmente irregulares, que é um problema considerável das outras abordagens que priorizam a estabilidade dos padrões em seus modelos. Há uma série de aplicações nas ciências nas quais uma abordagem com essas características alcançaria resultados mais eficientes e condizentes com a realidade.

2.5.3.1 Clusterização Baseada em Grades

Na clusterização baseada em grades, o espaço de busca do problema é dividido em várias porções (grades), em diferentes níveis de granularidade, de modo que possam ser clusterizados individualmente. De modo simplista, pode-se afirmar que são abordagens que incentivam uma discretização do espaço busca da problemática. A principal vantagem de abordagens dessa tipologia é que o custo computacional é relativamente constante, isto é, independente da quantidade de elementos do banco de dados, dependendo quase que exclusivamente do número de porções/grades definidas no processo de discretização do espaço. STING (WANG et al., 1997), CLIQUE (AGRAWAL et al., 1998), BANG (SCHIKUTA; ERHART, 1997), MAFIA (ADINETZ et al., 2013) e ORCLUS (AGGARWAL; YU, 2002) são alguns exemplos de algoritmos dessa natureza.

O método CLIQUE (AGRAWAL et al., 1998), por exemplo, pode encontrar automaticamente subespaços com clusters de alta densidade, sem que seja preciso qualquer suposição de distribuição de dados. Os resultados reportados na literatura indicam que o método possui uma eficiência alta na clusterização de dados de alta dimensão, cada vez mais comum nas indústrias.

2.5.4 Clusterização Espectral

Conforme definido em Wong (2016), abordagens baseadas em clusterização espectral visam encontrar mínimos locais, fazendo uso de um algoritmo iterativo, para encontrar clusters eficientes, usando para isso diferentes pontos iniciais de cluster inicial. A motivação principal dos métodos dessa natureza é fazer uso do espectro da matriz de similaridade dos dados para realizar a redução da dimensionalidade, ou seja, uma forma de realizar cortes no espaço busca, de modo a restringir o ambiente de procura do problema, conforme é melhor detalhado por Ng, Jordan e Weiss (2002). Diversas estratégias de cortes

e aplicações são apresentadas por Shi e Malik (2000), demonstrando variantes que podem ser utilizadas em métodos dessa natureza.

2.5.5 Clusterização Gravitacional

A clusterização gravitacional foi proposta inicialmente por Wright (1977). No método original, cada instância de dados é considerada como uma partícula no espaço de busca, na qual é movimentada, iterativamente, através de leis físicas que são aplicadas para estimular a movimentação e, por consequência, aumentar o espaço de busca por melhores soluções. As movimentações contam com parâmetros de aleatoriedade que estimulam com que as buscas possam, por vezes, sair de ótimos locais e visitar espaços ainda não investigados. Ao longo dos anos vários incrementos no método foram propostas pela literatura.

Gomez, Dasgupta e Nasraoui (2003) propuseram uma reformulação para o método original utilizando leis físicas diferentes, fundamentadas na mecânica clássica. Os resultados alcançados pelo método demonstraram melhores de eficiência significativas na maioria dos testes. Blekas e Lagaris (2007) e Junlin e Hongguang (2011) herdaram estruturas das abordagens anteriores, mas propuseram uma nova abordagem de busca local que trabalha em conjunto com um modelo Gaussiano. Por sua vez, Sanchez et al. (2014) adotaram uma lógica de representação de solução e de busca local distinta dos demais autores, e propuseram um método baseado em lógica fuzzy e análise multivariada no processo de tomada de decisão.

2.5.6 Clusterização por Abordagens Evolucionárias

As abordagens baseadas em algoritmos evolucionários (também denominadas como clusterização de manadas), buscam representar comportamentos, movimentos e fenômenos observados na natureza. Em geral, abordagens dessa classe aplicadas em problemas de clusterização são tratadas como métodos de duas fases distintas e bem definidas. A primeira fase, é representada por um modelo que visa orientar a exploração do espaço de busca ao longo das iterações. Esse modelo tem sua estruturação definida conforme a representação comportamental da natureza inspirada no método. Uma população é iniciada, de modo que cada elemento de tal população representa uma instância do problema sendo executada paralelamente a outra, com a representação da solução indicando as posições no espaço busca. Em geral, tais posições são iterativamente atualizadas após cada fluxo completado, de modo que as melhores soluções de cada membro da população, e da população como um todo, sejam utilizadas neste processo.

O método macro rege a busca da abordagem pelo espaço de busca, objetivando que, com os membros da população, possa investigar de forma eficiente o espaço de busca do problema, possibilitando encontrar soluções de extrema qualidade. Para isso, é necessário uma segunda etapa no meio do processo, que refere-se a abordagem responsável pelas regras de clusterização propriamente dita. Via de regra, são escolhidas abordagens classificadas como métodos de particionamentos ou hierárquicos para essa segunda etapa, com adaptações que possam tornar o método como um todo mais eficiente.

Observando de forma rudimentar, métodos dessa classificação podem parecer similares as abordagens baseadas em clusterização gravitacional, onde as instâncias de dados são movidas de acordo com um modelo físico. No entanto, os detalhes intrínsecos aos métodos evolucionários são bastante diferentes. Nas abordagens de clusterização gravitacional a maior parte dos parâmetros dos métodos são fixos e estipulados por regras da física clássica, enquanto que nas abordagens evolucionárias são dinâmicas, de modo que possa melhor representar os problemas a serem tratados, permitindo inclusive estruturas de mutação durante o processo iterativo. Além disso, a macroestrutura que é representada neste tipo de abordagem, via de regra, permite que uma área maior do espaço de busca seja visitada durante a execução do método.

Há uma série de algoritmos evolucionários na literatura que podem ser utilizados como uma das etapas de um método de clusterização, alguns deles são: Algoritmos Genéticos (GA), *Particle Swarm Optimization* (PSO), *Ant Colony Optimization* (ACO) e *Differential Evolution* (DE). Entretanto, há uma quantidade limitada de trabalhos na literatura que abordaram algoritmos dessa classe para tratar problemas de clusterização. Abordagens híbridas envolvendo algoritmos evolucionários, principalmente PSO, foram implementadas na literatura para vários problemas nas mais diversas áreas das ciências (DAS; ABRAHAM; KONAR, 2008; DENG; LIN; LO, 2012; KUO; ZULVIA, 2013; MISHRA; PANDA; MAJHI, 2014; SILVA; HERTHEL; SUBRAMANIAN, 2019), demonstrando resultados promissores e eficientes, nas quais geraram insights basilares e fundamentações para o desenvolvimento da abordagem híbrida proposta no presente trabalho.

Capítulo 3

Procedimentos Metodológicos

Formular os critérios matemáticos, computacionais e de engenharia que representem o problema tratado e suas peculiaridades, de modo a fornecer os subsídios necessários para a determinação das soluções ótimas de clusterização, incluindo os seus ranqueamentos, fundamentados em suas composições químicas, mineralógicas e de suas razões.

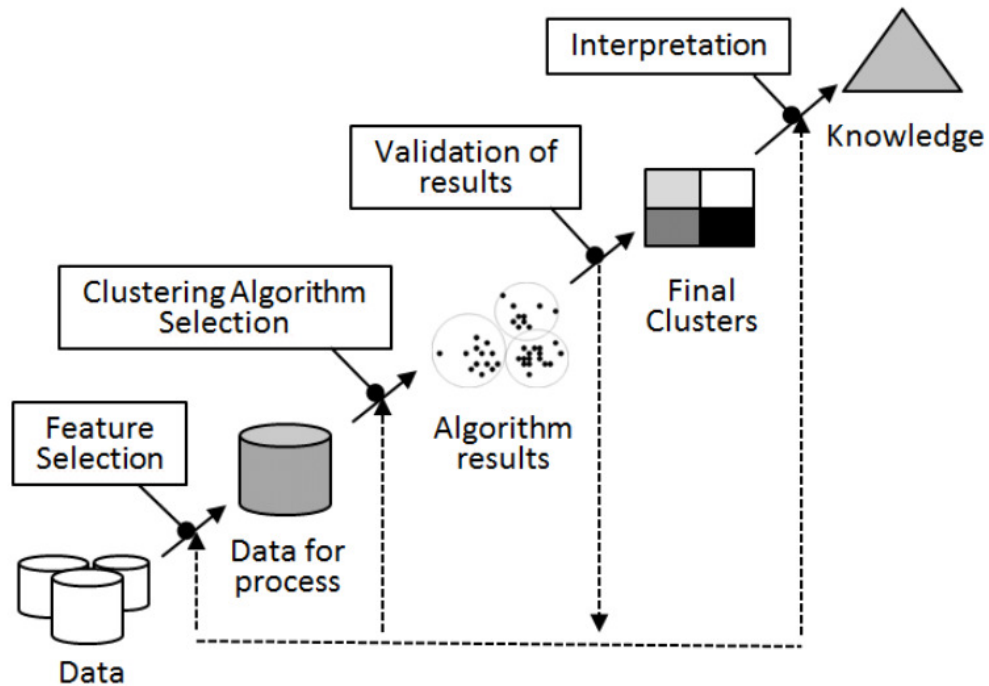
3.1 Metodologia da Pesquisa

A presente pesquisa objetivou o desenvolvimento de uma abordagem de otimização, fundamentada em uma metaheurística evolucionária híbrida, para a clusterização de dados relacionados a materiais agregados de concreto, com o intuito de encontrar as relações ótimas envolvendo os multicritérios associados aos materiais. O estudo realizado apresentou cinco fases principais, fundamentalmente importantes para os resultados finais alcançados no presente trabalho: Mapeamento e análise da literatura; Coleta de dados e processamento dos conjuntos de instâncias; Desenvolvimento da abordagem de otimização proposta; Aplicação do algoritmo de otimização nas instâncias do problema; Avaliação e análise dos resultados alcançados; Estudo de mapeamento e correlação das clusterizações com as probabilidades de expansão nas reações álcali-agregado. A Figura 3.1 descreve resumidamente as etapas adotadas para a elaboração da presente tese.

Inicialmente, fez-se necessária a realização de um mapeamento da literatura acerca de métodos de otimização em ciência e engenharia dos materiais, bem como de abordagens de clusterização em outras áreas da ciência, uma vez que não foi constatada aplicação de otimização em clusterização em engenharia de materiais na literatura. Tal estudo forneceu os insights necessários para o estudo e planejamento demandado na construção de um método de otimização robusto e eficiente, de modo a alcançar os objetivos propostos pela presente tese.

A coleta dos agregados objetos de estudo foram obtidas em diversas regiões do Nor-

Figura 3.1: Etapas do Processo de Otimização em Clusterização no Problema Abordado



Fonte: Halkidi, Batistakis e Vazirgiannis (2011).

deste do Brasil, estando sujeitas suas composições químicas e mineralógicas às suas respectivas realidades geográficas, conforme as localidades coletadas. Para que fosse possível sua utilização através do método de otimização desenvolvido, foi necessário, em primeiro lugar, proceder com análises que permitissem a obtenção de suas composições químicas. Para isso, foram realizadas duas análises: A primeira, fundamentada em fluorescência de raios X, permitiu auferir as composições químicas de todos os materiais coletados. Em seguida, foram realizadas análises Rietveld em todos os agregados, visando obter suas respectivas composições mineralógicas e, em função delas, alcançar as composições químicas. As obtenções das composições químicas através das composições mineralógicas são ilustradas e detalhadas na Seção 3.2. Desse modo, foi possível gerar os bancos de dados utilizados como objeto de estudo no problema, bem como seus respectivos conjuntos de instâncias a serem tratados. O processo é descrito detalhadamente na Seção 5.1.

Com o intuito de prover uma solução para o problema de clusterização abordado, foi elaborado e desenvolvido uma metaheurística de otimização híbrida, fundamentada em *Particle Swarm Optimization Algorithm* (PSO), que usa mecanismos propostos e adaptativos em suas estruturas, e que faz uso de um método K-means em uma das etapas de seu processamento. O método desenvolvido foi denominado *Automatic K-means Particle Swarm Optimization* (AAKPSO). O método possui mecanismos e métricas de avaliação inovadoras, as quais possibilitaram alcançar resultados extremamente sólidos e eficientes

no problema proposto. Sua implementação se deu através de linguagem de programação C++. Todas as estruturas e mecanismos desenvolvidos no método de otimização proposto, inclusive os pseudocódigos, são apresentados no Capítulo 4.

O método desenvolvido foi, em seguida, aplicado a cada uma das instâncias pertencente aos conjuntos criados a partir dos bancos de dados de agregados coletados previamente. Os resultados alcançados foram reportados de diversas formas, possibilitando uma série de análises sistêmicas e analíticas das clusterizações ótimas obtidas para todas as instâncias dos conjuntos. Do mesmo modo, foi possível realizar uma comparação entre os melhores resultados alcançados por conjuntos de instâncias e banco de dados, que permitiu a obtenção de diversos insights e a realização de estudos criteriosos, relacionando as clusterizações otimizadas com análises mineralógicas e de expansão das britas.

3.2 Descrição Formal da Problemática de Clusterização de Materiais

Conforme destacado na Seção 1.2, o objetivo geral da presente tese é propor e desenvolver uma metaheurística híbrida de otimização para problemas de clusterização aplicados às ciências dos materiais. Observando de maneira específica, o problema abordado neste trabalho se refere a otimização de clusterização de britas, sob a ótica de suas composições químicas e mineralógicas.

Em suas composições mineralógicas, as britas possuem, principalmente minerais das famílias da Sílica, Mica e Feldspato. Além disso, também possuem outros minerais, como por exemplo, a Caulinita, Calcita e Dolomita. A Tabela 3.1 descreve os principais minerais presentes nas britas. Do mesmo modo, há diversas composições químicas identificadas nas britas brasileiras, presentes nos minerais relatados. As principais composições químicas são apresentadas na Tabela 3.2.

Tabela 3.1: Composição Mineralógica das Britas

SÍLICA	MICA			FELDSPATO							OUTROS					
Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita

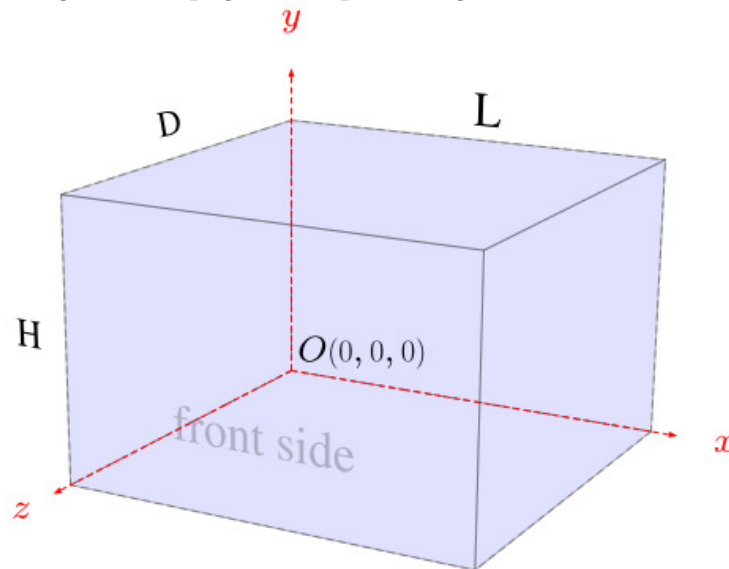
Tabela 3.2: Composição Química das Britas

Al2O3	CaO	K2O	SiO2	Fe2O3	MgO	Na2O	SO3
-------	-----	-----	------	-------	-----	------	-----

Os materiais podem ser clusterizados conforme diversas regras. No presente trabalho, as composições químicas das britas são tomadas como os atributos (características) a serem levadas em consideração no processo de clusterização. Isto é, os valores assumidos em tais características pelas britas serão os dados de entrada para o processo decisório de otimização sobre quais são os agrupamentos ótimos, indicando quais subconjuntos de britas são similares com base nas características pré-definidas. Para isso, inicialmente, foi realizado um estudo de mapeamento de instâncias a serem consideradas, com todos os detalhes reportados no Capítulo 5.1.

Desse modo, dado que o processo de clusterização em questão levará em consideração três atributos, e que há oito atributos (principais composições químicas) relatados, se fez necessário um estudo de todas as combinações possíveis entre tais características, onde cada combinação de três a três é considerada com uma configuração de instância a ser tratada pelo problema, a qual pode ser representada em um espaço tridimensional, conforme a Figura 3.2.

Figura 3.2: Ilustração do Espaço de Representação Tridimensional de Clusterização



Fonte: Elaborado pelo autor (2019)

Para ilustrar um exemplo prático do problema, uma base de dados fictícia é apresentada pela Tabela 3.3 a seguir, na qual é recortada uma instância arbitrária e apresentada, a título de demonstração do que seria uma instância para o problema, na Tabela 3.4. Em seguida, são extraídas duas instâncias (configurações) para serem tratadas (clusterizadas pelo método de otimização proposto). A primeira, utilizando os atributos CaO, K₂O e MgO, enquanto a segunda instância do problema leva em consideração os atributos SiO₂, MgO, Na₂O. Os dados referentes as duas instâncias são apresentados na Tabela 3.5. Note

que são apresentadas duas instâncias arbitrárias com o intuito de demonstrar visualmente o problema, entretanto, apesar de apenas duas instâncias serem apresentadas, há uma série enorme de quantidade de instâncias que poderiam ser mapeadas e consideradas, dado que o exemplo de banco de dados apresentado possui oito atributos.

Tabela 3.3: Composição Química das Britas - Exemplo de Banco de Dados Fictício

Brita	Composição Química							
	Al ₂ O ₃	CaO	K ₂ O	SiO ₂	Fe ₂ O ₃	MgO	Na ₂ O	SO ₃
1	9,5	2,0	3,0	7,0	1,5	0,6	3,4	5,0
2	9,0	3,5	3,0	7,0	6,0	1,5	5,5	1,0
3	11,0	3,0	5,0	8,0	5,8	5,0	5,0	5,5
4	8,0	3,3	4,5	6,5	3,0	4,5	6,5	2,0
5	10,5	2,0	5,0	7,0	2,2	4,6	3,5	2,5
6	12,0	2,4	3,5	7,2	2,0	1,0	2,5	2,0
7	12,0	3,4	6,5	4,0	2,0	2,6	6,0	1,6
8	13,5	3,0	7,5	7,5	2,6	3,2	7,0	1,0
9	11,0	3,5	3,5	7,0	1,0	2,0	3,0	7,0
10	12,5	2,8	8,0	3,6	2,0	2,3	6,5	8,5

Tabela 3.4: Exemplificação de uma Instância ([CaO, MgO, Na₂O]) para o Problema

Britas	Instância [CaO, MgO, Na ₂ O]		
	CaO	MgO	Na ₂ O
1	2,0	3,0	0,6
2	3,5	3,0	1,5
3	3,0	5,0	5,0
4	3,3	4,5	4,5
5	2,0	5,0	4,6
6	2,4	3,5	1,0
7	3,4	6,5	2,6
8	3,0	7,5	3,2
9	3,5	3,5	2,0
10	2,8	8,0	2,3

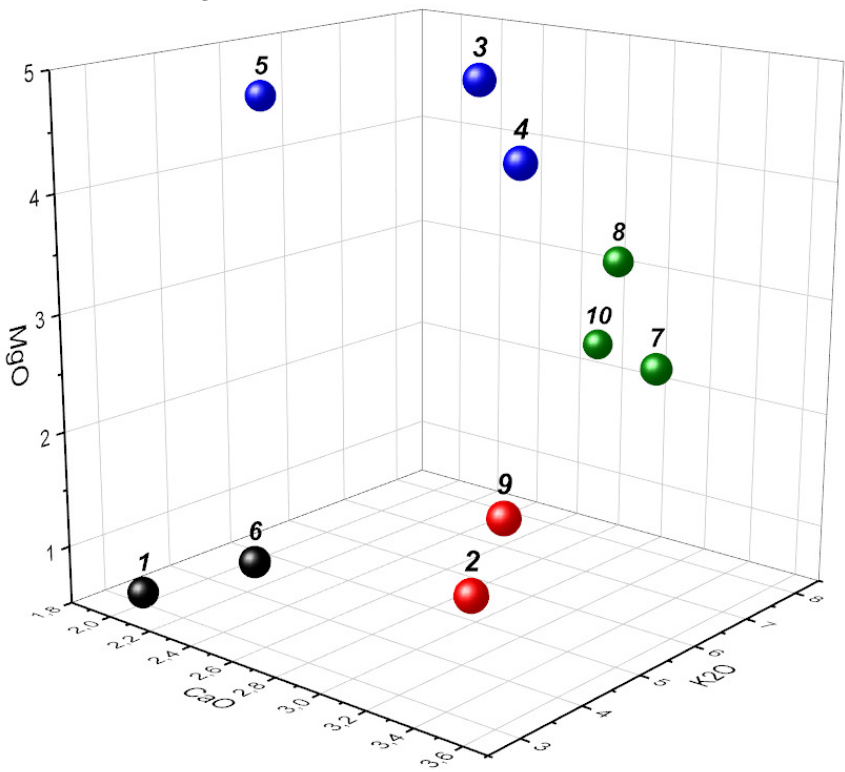
Conforme pode ser observado na Tabela 3.5, as instâncias 1 e 2 se referem as mesmas britas, enumeradas de 1 até 10, e possui todos os atributos distintos. Visando apresentar uma ilustração visual das instâncias, são apresentadas nas Figuras 3.3 e 3.4 as suas respectivas representações gráficas, assim como uma clusterização de seus dados realizada arbitrariamente, demonstrada em cores e enumeradas para uma melhor visualização.

Na clusterização realizada para a primeira instância apresentada na Figura (3.3), é possível perceber que quatro clusters foram formados, $A = \{1, 6\}$, $B = \{2, 9\}$, $C = \{3, 4, 5\}$ e $D = \{7, 8, 10\}$. A definição dos clusters se deu forma inteiramente arbitrária, levando em consideração a distância euclidiana, apenas a título de apresentação do

Tabela 3.5: Instâncias Fictícias Arbitrárias para Exemplificação do Problema

Britas	Instância 1			Instância 2		
	CaO	K ₂ O	MgO	SiO ₂	Fe ₂ O ₃	Na ₂ O
1	2,0	3,0	0,6	7,0	1,5	3,4
2	3,5	3,0	1,5	7,0	6,0	5,5
3	3,0	5,0	5,0	8,0	5,8	5,0
4	3,3	4,5	4,5	6,5	3,0	6,5
5	2,0	5,0	4,6	7,0	2,2	3,5
6	2,4	3,5	1,0	7,2	2,0	2,5
7	3,4	6,5	2,6	4,0	2,0	6,0
8	3,0	7,5	3,2	7,5	2,6	7,0
9	3,5	3,5	2,0	7,0	1,0	3,0
10	2,8	8,0	2,3	3,6	2,0	6,5

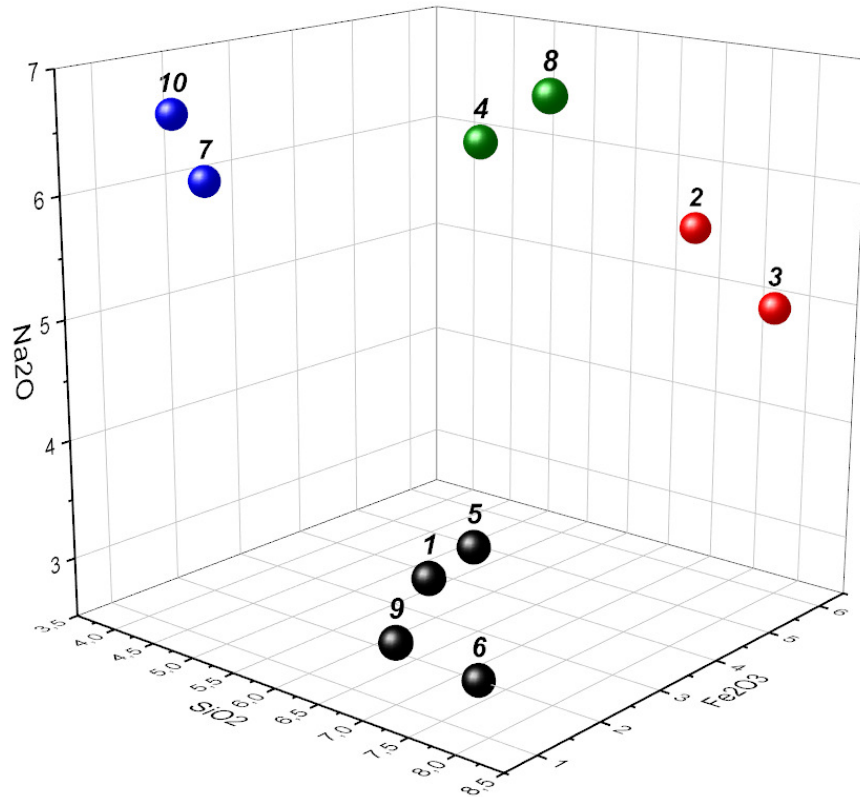
Figura 3.3: Clusterização Arbitrária - Instância 1 - Atributos: CaO-K₂O-MgO



Fonte: Elaborado pelo autor (2020)

problema.

Figura 3.4: Clusterização Arbitrária - Instância 2 - Atributos: SiO_2 - Fe_2O_3 - Na_2O



Fonte: Elaborado pelo autor (2020)

Por sua vez, através da clusterização efetuada na segunda instância (3.4) foram obtidos os clusters $A = \{1, 5, 6, 9\}$, $B = \{2, 3\}$, $C = \{7, 10\}$ e $D = \{4, 8\}$. Ambos os agrupamentos podem ser melhor observados através da Tabela 3.6.

Note que, levando em consideração os atributos presentes na primeira instância, as britas 3, 4 e 5 estão presentes no mesmo cluster (C), demonstrando que possuem fortes similaridades levando em consideração os três atributos (CaO - K_2O - MgO) da instância em questão. Por sua vez, é possível observar que as mesmas três britas apresentam características significativamente distintas na segunda instância, quando são considerados outros três atributos (SiO_2 , Fe_2O_3 e Na_2O), uma vez que a brita 3 está presente no cluster B, a brita 4 no cluster D, enquanto a brita 5 pertence ao cluster A. Isto é, as três britas estão pertencendo a clusters distintos, enfatizando suas dissimilaridades em relação aos atributos considerados. De forma análoga, podemos verificar diversos outros insights dessa natureza observado as clusterizações das duas instâncias.

Tais observações demonstram que os materiais, em especial as britas, podem ser clusterizados de formas completamente diferentes a depender dos atributos que são levados

Tabela 3.6: Comparação de Clusterizações Arbitrárias da Instância Fictícia

Britas	Clusters (Inst. 1: CaO-K ₂ O-MgO)	Clusters (Inst. 2: SiO ₂ -Fe ₂ O ₃ -Na ₂ O)
1	A	A
2	B	B
3	C	B
4	C	D
5	C	A
6	A	A
7	D	C
8	D	D
9	B	A
10	D	C

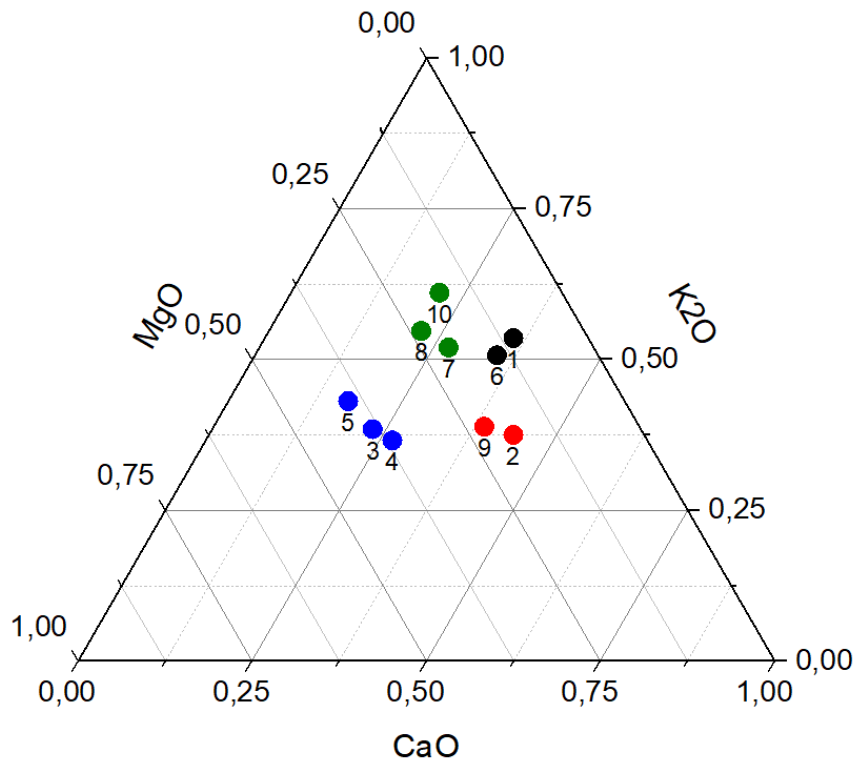
em consideração em um processo de otimização em clusterização. Em outras palavras, algumas britas podem ter fortes similaridades com outras, quando levada em consideração alguns atributos, enquanto podem ter enorme dissimilaridade com as mesmas, quando levadas em consideração características distintas.

Assim, o intuito da busca pela otimização de cada possível configuração é que, após obter todas as soluções ótimas (para todas as configurações existentes), serão obtidos os melhores agrupamentos (e insights) relacionadas à todas as possibilidades de atributos que possam ser levadas em consideração, bem como, também é possível encontrar aquelas que apresentam uma melhor qualidade de agrupamento, conforme a métrica de desempenho a ser adotada pelo método de otimização do problema. Isso pode trazer inúmeros benefícios e aplicações em decisões industriais, principalmente relacionadas aos processos produtivos e à problemáticas de seleção de matérias-primas. Tais aplicações podem ainda otimizar significativamente o desempenho operacional, promover a redução de custos de produção, bem como impactos ambientais causados, conforme as estratégias organizacionais adotadas pela empresa.

Como as instâncias reais tendem a apresentar um volume de dados significativamente maior do que o exemplo apresentado, a visualização tridimensional pode, por vezes, se tornar comprometida e não apresentar de forma intuitiva os clusters formados. Nestes casos, a visualização através de ternários pode ilustrar melhor os resultados obtidos através da abordagem de resolução do problema. As Figuras 3.5 e 3.6 ilustram as representações em ternários das soluções arbitrárias destacadas nas instâncias fictícios apresentadas na Tabela 3.5.

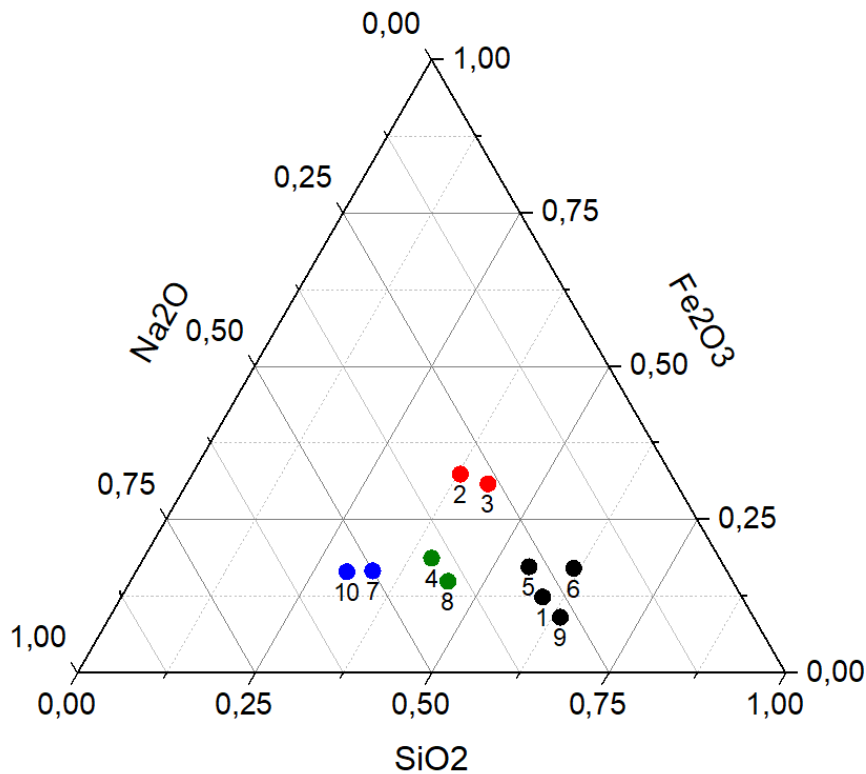
Os bancos de dados obtidos, tanto através da análise Rietveld, como por intermédio do uso de fluorescência de raios X, foram utilizados como base de entrada para a aplicação do

Figura 3.5: Clusterização Instância 1 - Visão de Ternários - Atributos: CaO-K₂O-MgO



Fonte: Elaborado pelo autor (2020)

Figura 3.6: Clusterização Instância 2 - Visão de Ternários - Atributos: SiO₂-Fe₂O₃-Na₂O



Fonte: Elaborado pelo autor (2020)

método de otimização em clusterização proposto e explanado no Capítulo 4. No Capítulo 5.1, é realizado um estudo elaborado visando o mapeamento e criação das instâncias objeto de estudo do problema. Os resultados alcançados após aplicação do método proposto, bem como os insights gerados, são reportados individualmente por banco de dados, bem como comparados de forma analítica e sistêmica em seguida, no Capítulo 6.

Capítulo 4

Resultados e Discussões - Parte I: Formulação e Desenvolvimento da Abordagem de Otimização Multicritério

Desenvolver uma metaheurística híbrida de otimização, fundamentada em algoritmo evolucionário, que permita otimizar e compreender a influência de composições químicas, mineralógicas e suas razões no processo de clusterização dos agregados.

Uma vez que problemas de clusterização podem ser classificados como *NP-Hard* (GONZALEZ, 1982), é extremamente custoso, computacionalmente, obter soluções exatas e ótimas, especialmente se considerando um conjunto com grandes quantidades de elementos a serem clusterizados. Uma dificuldade adicional diz respeito às métricas de avaliação de qualidade das soluções, as quais muitas vezes precisam ser bem mensuradas para que possam retratar, com eficiência, os objetivos buscados na aplicação. Deste modo, abordagens heurísticas, aproximativas e de simulação têm recebido uma atenção especial por parte da comunidade científica para esse tipo de problema, conforme destacado no Capítulo 2.

Neste capítulo, é apresentado detalhadamente a abordagem de resolução desenvolvida para tratar o problema de clusterização considerada neste trabalho. O método implementado trata-se de uma abordagem híbrida, fundamentado em um *Particle Swarm Optimization Algorithm* (PSO), que usa mecanismos propostos e adaptativos em suas estruturas, e que faz uso de um método K-means em uma das etapas de seu processamento. O método desenvolvido foi denominado *Automatic K-means Particle Swarm Optimization* (AAKPSO).

Com o método proposto, foi possível alcançar soluções de qualidade no problema tratado, e demonstraram um desempenho significativamente melhor em comparação com os próprios métodos desenvolvidos executados isoladamente, enfatizando a importância das estruturas e mecanismos implementados no método proposto.

4.1 Particle Swarm Optimization

Algoritmos evolucionários têm sido amplamente utilizados em diversas problemáticas na literatura de otimização, principalmente devido ao alto poder de busca aliado a um custo computacional significativamente reduzido comparado com outras abordagens metaheurísticas presentes na literatura.

Dentre os algoritmos evolucionários, destaca-se o algoritmo *Particle Swarm Optimization* (PSO), que é uma metaheurística baseada em busca populacional que simula o comportamento social de um enxame de partículas em busca por comida. Neste método, é definido inicialmente o tamanho da população (quantidade de partículas) e a posição de cada partícula no espaço representa uma solução para o problema. As partículas têm suas posições alteradas a cada iteração em função de vetores que representam a sua velocidade (incluindo sua direção) no espaço em busca dos alimentos. A atualização dos vetores de velocidade, via de regra, são representados através de funções que levam em consideração, de algum modo, as melhores posições já visitadas localmente (da própria partícula) e globalmente (da população como um todo).

Por tal motivo, é extremamente importante que as métricas de avaliação de desempenho estejam bem definidas e implementadas, pois elas serão responsáveis por guiar a movimentação iterativa das partículas em busca dos alimentos, representado pela melhor solução do problema. Desse modo, temos que a cada iteração t , de máximo de t_{max} iterações, a i -ésima partícula do enxame (população) P terá sua posição $x_{i,j}^t$ atualizada através de sua velocidade $v_{i,j}^{t+1}$ em todas as dimensões j . A Equação (4.1) a seguir destaca o processo de atualização das posições das partículas para a próxima iteração.

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \quad (4.1)$$

Onde $x_{i,j}^{t+1}$ e $v_{i,j}^{t+1}$ representam, respectivamente, a posição (solução) e a velocidade de partícula i em todas as dimensões j do espaço de busca na iteração $t + 1$. A velocidade da partícula é responsável por direcionar a mudança nas posições das partículas a cada iteração do método, e é obtida através da Equação (4.2) a seguir.

$$v_{i,j}^{t+1} = Wv_{i,j}^t + C_1r_1(p_{i,j} - x_{i,j}^t) + C_2r_2(p_{g,j} - x_{i,j}^t) \quad (4.2)$$

Na equação, $p_{i,j}$ representa a melhor posição (local) que a partícula i já esteve no espaço de busca durante toda a execução do método, isto é, a posição onde ela obteve os melhores desempenhos em termos de qualidade da solução obtida. Por outro lado, $p_{g,j}$ indica a melhor posição global, ou seja, a melhor posição já alcançada por qualquer partícula naquele instante, considerando todo o período de execução do método até então desde a primeira iteração.

Há uma série de parâmetros que são considerados em um algoritmo PSO, bem como existem várias formas de considerá-los na literatura, uma vez que o desempenho do método depende diretamente da definição de tais parâmetros. Na atualização da velocidade das partículas, há cinco parâmetros principais que guiam o processo de busca iterativa do método, W , C_1 , C_2 , r_1 , r_2 . W corresponde ao peso de inércia do método, o qual é responsável por ponderar o impacto que a velocidade atual da partícula terá na atualização da velocidade na próxima iteração do método, e, conseqüentemente, da atualização da posição da partícula. Eberhart e Shi (2000) observaram em estudos experimentais, que valores próximos de 0.90 para o peso de inércia tendem a gerar resultados mais eficientes. Após testes preliminares, foi definido, no método proposto, W variando de forma decrescente de 0.95 até 0.80, em proporções de $\frac{0.15}{t_{max}}$, a cada iteração, sendo $t_{max} = 200$. Destaca-se, ainda, que para o presente trabalho foi considerado uma população inicial de 100 partículas. Tal definição ocorreu após testes paramétricos indicarem que não havia ganho significativo em qualidade de solução na medida em que populações mais elevadas eram testadas, enquanto o aumento do custo computacional era significativo. Os testes foram realizados com populações entre 50 e 300 partículas, com intervalos de 50 partículas.

Os parâmetros r_1 e r_2 representam os fatores estocásticos do algoritmo, possibilitando que, por vezes, uma partícula possa visitar regiões do espaço busca que ainda não foram visitadas, possibilitando investigar uma área ampla. No método proposto, os parâmetros são valores entre 0 e 1, que são obtidos de forma aleatória através de uma distribuição uniforme. As formas de definição de tais parâmetros são relativamente padrões na literatura, com algumas leves propostas de alterações quando ocorrem.

Por sua vez, os parâmetros C_1 e C_2 representam os coeficientes de aceleração das partículas. Eles são usados para ponderar a contribuição cognitiva e social do comportamento das partículas, representados, respectivamente, pela melhor posição local e global. Logo, C_1 indica a confiança de uma partícula com base apenas em sua melhor posição histórica (desde a iteração inicial), enquanto C_2 representa a confiança da partícula em relação à população como um todo, através da melhor posição global. Deng, Lin e Lo (2012) destacam que tais valores podem ser utilizados e atualizados de diferentes formas,

a depender das estratégias a serem empregadas na resolução do problema.

Estudos reportados na literatura concluíram que valores entre 1,90 e 2,10 para os coeficientes de aceleração das partículas tendem a ser mais promissores na busca por soluções de melhor qualidade (GUANGYOU, 2007; SHI; EBERHART, 1998). Após a realização de testes de parametrização preliminares, os parâmetros C_1 e C_2 foram definidos como 2,10 e 1,95, respectivamente.

Nos tópicos seguintes, são apresentados, detalhadamente, cada etapa da abordagem de resolução proposta e os respectivos pseudocódigos de cada uma das etapas, bem como a apresentação do pseudocódigo geral integrando todas as estruturas relatadas.

4.2 Representação da Solução

Como destacado anteriormente, a posição de cada partícula no espaço representa uma solução para o problema a ser tratado. Neste caso, para o desenvolvimento de uma abordagem fundamentada em um algoritmo PSO, é necessário, inicialmente, avaliar a melhor forma de representação da solução da problemática a ser tratada, isto é, como serão representadas as posições das partículas de modo que retratem as soluções buscadas, bem como torne possível a avaliação de sua qualidade e a viabilidade em relação às restrições do problema.

A estrutura das partículas é definida como possuindo N variáveis, que é definida como um produto da quantidade máxima de clusters que podem ser formados (K_{max}), somados pelo produto de tal quantidade pelo número de atributos considerados A , conforme a Equação (4.3).

$$N = K_{max} + (K_{max}A) \quad (4.3)$$

Desse modo, para aplicação na solução proposta, após a realização de testes de parametrização, foi considerado como limitante máximo a quantidade 10 clusters, enquanto o número de atributos foi definido como 3, totalizando uma quantidade de 40 variáveis na representação da solução. Todas as variáveis podem assumir valores entre 0 e 1. As 10 variáveis iniciais são responsáveis por indicar se o cluster j está ativo (caso assumam valores ≥ 0.5) ou inativo (caso contrário), enquanto as demais 30 variáveis representam as posições dos centróides de cada cluster j , três a três, respectivamente, dado que para

o problema, por definição, são considerados três atributos.

A Tabela (4.1) ilustra a representação de uma solução supondo que o número máximo de clusters fossem três, para simplificar a apresentação visual.

Tabela 4.1: Exemplo de Solução para o Problema

0.585	0.268	0.921	0.252	0.196	0.855	0.823	0.902	0.254	0.541	0.387	0.773
Ativo	Inativo	Ativo	Centróide 1		Centróide 2			Centróide 3			

4.3 Normalização dos Dados

Antes do início propriamente dito do método de resolução, há uma etapa de normalização dos dados de entrada. O objetivo da normalização é de que todos os elementos objetos do processo de clusterização tenham os valores de seus atributos/características entre 0 e 1, uma estratégia eficiente em problemáticas dessa natureza para evitar falhas interpretativas e de avaliação no decorrer do método. Para que isso seja possível, todos os dados presentes na instância a ser tratada são divididos pelo maior valor de atributo presente no banco de dados. Ao final da execução do método, a melhor solução obtida é retornada em seu formato original, multiplicando pelo mesmo valor previamente utilizado no processo de normalização.

4.4 Inicialização

No ato de inicialização do método, valores iniciais aleatórios são gerados para todas as variáveis de todas as partículas, isto é, tanto para as que indicam se os clusters estão ativos ou inativos, como para as respectivas posições de seus centróides no espaço de busca. No caso dos valores relacionados as posições dos centróides, são observados, no processo de geração inicial aleatória, os limitantes de cada atributo do problema, baseado no menor e maior valor de cada atributo.

Tal estratégia se faz importante para que não sejam geradas soluções iniciais em áreas do espaço de busca que não possuam relevância ou possibilidades de serem promissoras, dado que não há elementos próximos daquelas regiões. Em testes realizados, a estratégia adotada demonstrou uma aceleração considerável na convergência para a melhor solução encontrada quando comparado com o método estritamente guloso. Além disso, a consideração dos limites estratégia permite, por definição, que as restrições de valores que possam ser assumidas pelas variáveis estejam dentro da região de viabilidade do problema. Ao

final da etapa de inicialização, todas as partículas iniciam com velocidades $v_{i,j}^t = 0$, sendo atualizadas apenas após o término da primeira iteração.

4.5 Algoritmo K-means

Visando refinar as partículas, o método K-means é executado após cada iteração do método logo após a atualização das posições (em função das novas velocidades das partículas) e da realocação dos pontos serem realizadas pelo PSO. Inicialmente, o procedimento herda as partículas e clusters devidamente atualizados pelo PSO. Em seguida, o método então atualiza todos os centróides ativos baseados na média de todos os elementos/pontos que fazem parte de cada um dos respectivos clusters. Em seguida, após a atualização dos centróides, uma nova alocação dos elementos é realizada, desta vez, considerando as novas posições assumidas pelos centróides.

O objetivo é buscar a melhor convergência possível nas regiões de busca de cada partícula em cada uma das iterações. Para isso, o procedimento anterior é repetido, iterativamente, até que ao menos um dos dois critérios de paradas estipulados sejam atingidos. O primeiro deles, diz respeito ao número de iterações do K-means, que é limitado de 30 iterações. O segundo critério diz respeito a ausência de ajustes nas posições dos centróides (e, por consequência, dos elementos presentes nos clusters) entre duas iterações consecutivas do K-means. O Algoritmo (4.1) ilustra de forma mais detalhada o procedimento.

Algorithm 4.1 Algoritmo K-means

```

1: Input:  $P$ 
2: while  $iter \leq iter_{max}$  do
3:   for  $i \in P$  do
4:     Determine os novos centróides para cada cluster  $k$ , de cada partícula  $i$ 
5:     for  $j \in N_{dados}$  do
6:       Para cada ponto, encontre o centróide ativo mais próximo
7:       Aloque o ponto no centróide mais próximo
8:     end for
9:   end for
10: end while

```

4.6 Avaliação das Soluções

Em uma problemática de clusterização, a etapa de avaliação das soluções (partículas) de uma abordagem é um dos passos mais importantes do método, uma vez que é fun-

damental para que uma varredura eficiente da abordagem de resolução seja executada, bem como para seja possível encontrar soluções eficientes dadas às métricas de qualidade alinhadas ao problema. Para a aplicação objeto de estudo, foi definida como métrica de avaliação uma versão adaptada do índice CS, denominada neste trabalho de ACS.

O índice CS representa uma das mais recentes e eficientes métricas reportadas na literatura (CHOU; SU; LAI, 2004), e possui um mecanismo de mensuração que avalia de forma inteligente o agrupamento internos aos clusters, bem como o espaçamento entre os clusters formados. Na métrica proposta adaptada, a Equação (4.4) descreve a regra de avaliação das soluções encontradas, onde leva em consideração, no numerador, a maior distância de cada ponto para um outro ponto qualquer do mesmo cluster, ao invés das distâncias de cada ponto para seus respectivos centróides. Tal adaptação foi fundamental para o objetivo de maior compactação dos clusters, condizente com a aplicação realizada. A modificação possibilita que o agrupamento buscado seja mais padronizado, uma vez que penaliza aqueles pontos que possuem uma distância significativa de um outro pertencente ao mesmo cluster. Note que, na Equação (4.4), temos que $D(x_i, x_j)$ indica o distanciamento entre os pontos x_i e x_j . Por sua vez, K indica a quantidade máxima de clusters ativos naquela partícula, m_i e m_j representam os centróides dos clusters i e j , enquanto N_i indica a quantidade de pontos alocados no cluster i .

$$ACS(K) = \frac{\frac{1}{2K} \sum_{i=1}^K \left(\frac{1}{N_i} \sum_{x_i \in C_i} \max_{X_q \in C_i} [D(X_i, X_q)] \right)}{\frac{1}{8K} \sum_{i=1}^K \left(\min_{j \in K, j \neq i} [D(m_i, m_j)] \right)} \quad (4.4)$$

Uma outra adaptação realizada, objetivou ponderar de forma mais incisiva o fator intracuster em detrimento dos distanciamentos entre os clusters, fundamental dada a aplicação prática da abordagem. Para isso, foi definido um peso de 80% para a aproximação dos elementos internos aos clusters, e de 20% para o afastamento entre os cluters. Conforme foi discutido na Seção 2, o denominador, que penaliza a proximidade entre os centróides, é fundamental para a qualidade em termos de heterogeneidade entre os clusters formados.

Numa abordagem PSO, cada partícula da população busca pela solução ótima levando em consideração a melhor posição global corrente da população (denominada gbest), e as melhores posições pessoais já encontrada em uma certa vizinhança, para cada partícula (chamadas pbests). Por definição, a melhor posição local de cada partícula representa

a melhor posição já ocupada por cada uma delas desde o início do método. Ao longo das iterações é verificado, através do procedimento de avaliação, se houve uma melhoria na solução em função da nova posição de cada partícula, se houver ocorrido, sua melhor posição local é atualizada. Se a posição $x_{i,j}^t$ obtida por i na iteração t for melhor do que a $p_{i,j}$, $p_{i,j}$ é atualizada. Caso contrário, nenhum procedimento é realizado. Por sua vez, a melhor posição global é atualizada sempre que uma das novas posições de uma das partículas apresentar um desempenho melhor do que a melhor posição global corrente.

4.7 Atualização das Partículas

Durante a execução do método proposto é realizado um procedimento de atualização das posições das partículas baseados em suas respectivas velocidades. O processo de atualização é baseado nas Equações (4.5) e (4.6), utilizando os parâmetros discutidos e definidos no início deste Capítulo.

$$v_{i,j}^{t+1} = Wv_{i,j}^t + C_1r_1(p_{i,j} - x_{i,j}^t) + C_2r_2(p_{g,j} - x_{i,j}^t) \quad (4.5)$$

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \quad (4.6)$$

Por fim, é importante destacar que, durante o procedimento de atualização, há um mecanismo de checagem da viabilidade das novas posições obtidas. Caso alguma das soluções não seja viável, é executado um mecanismo de viabilização. Em geral, soluções inviáveis são representadas por determinadas dimensões das partículas assumirem valores menores do que zero, ou maiores do que um, fugindo assim dos limites de viabilização. Nestes caso, é aplicado um processo de viabilização fundamentado em função exponencial que retorna um valor entre 0 e 1 para qualquer valor de entrada fora dos limitantes.

4.8 Mecanismos de Mutação

A etapa de mutação é responsável por realizar a diversificação do enxame entre uma iteração e outra. Para cada partícula, há uma pequena probabilidade de que ocorra uma mutação em alguma dimensão de sua posição no espaço, ou seja, que seja alterada sua posição em relação as demais.

Tal procedimento só pode ser executado quando a iteração corrente obedecer a regra que permite as mutações serem executadas, isto é, só podem haver mutações quando

$t < t_{max} * pMut$, onde $pMut = 0.5$. O objetivo desse parâmetro na mutação é permitir que, no início, o método consiga percorrer por um espaço de busca maior, priorizando a convergência na segunda metade das iterações. O procedimento é descrito no Algoritmo (4.2). Note que, quando uma mesma partícula não alcançar melhoras em seu desempenho após 10% do número máximo de iterações ocorrerem consecutivamente, $i = 1$, ativando o mecanismo de mutação.

Algorithm 4.2 Procedimento de Mutação

```

1: Input:  $P$ 
2: for  $i \in P$  do
3:   if  $mut(i) == 1$  (mecanismo para determinar se  $i$  sofrerá mutação) then
4:     Selecione aleatoriamente uma dimensão  $j$  da posição (solução) da partícula  $i$ 
5:     Gere aleatoriamente um novo valor (viável) para a dimensão  $j$  da partícula  $i$ 
6:   end if
7: end for

```

4.9 Visão Geral AAKPSO

De modo a integrar todas as estruturas e mecanismos apresentados ao longo do presente Capítulo, o Algoritmo (4.3) a seguir descreve de forma concisa e agrupada, o passo a passo da execução da abordagem de resolução.

Algorithm 4.3 Algoritmo AAKPSO

```

1: Leia o arquivo de instância que armazena os dados de entrada
2: Normalize os dados de entrada baseado nos limitantes estipulados
3: Gere as soluções iniciais e velocidades para o enxame  $P$ 
4: Aloque todos os pontos baseado nas novas posições das partículas
5: Verifique as condições de restrições envolvendo os clusters formados
6: Execute o procedimento K-means em  $P$ 
7: Avalie as soluções alcançadas
8: Inicialize e armazene pbests  $p_{i,j}$  e gbest  $p_{g,j}$ ;
9: while  $t \leq t_{max}$  do
10:   Compute e atualize as velocidades de todas as partículas  $v_{i,j}^{t+1}$ 
11:   Atualize as posições do enxame  $x_{i,j}^{t+1}$ 
12:   Verifique as condições de restrições envolvendo as partículas
13:   Realoque todos os pontos baseado nas novas posições das partículas
14:   Verifique as condições de restrições envolvendo os clusters formados
15:   Execute o procedimento K-means em  $P$ 
16:   Avalie as soluções alcançadas
17:   Avalie e atualize os pbests  $p_{i,j}$  e gbest  $p_{g,j}$ , caso necessário;
18: end while

```

Capítulo 5

Resultados e Discussões - Parte II: Resultados Computacionais com a Aplicação da Metaheurística de Otimização Multicritério - Clusterização de Rochas com Critérios Químicos e Cristalográficos

Aplicar o método desenvolvido e avaliar os resultados quanto à sua eficiência e desempenho matemático e computacional nas clusterizações alcançadas, bem como se há indicação da prevalência de algum elemento químico (ou conjunto de elementos) que determinem a formação das melhores clusterizações no ranqueamento.

5.1 Mapeamento e Estruturação das Instâncias do Problema

Visando alcançar o objetivo proposto para o trabalho, foi necessário, inicialmente, investigar as configurações de instâncias que seriam tratadas pelo método proposto. Para isso, foram coletadas amostras de britas de várias regiões do Brasil, cada uma delas com composições mineralógicas e químicas distintas. A partir dessa coleta, foram realizadas análises com o intuito de estabelecer as composições químicas das britas, as quais foram auferidas de duas formas: Primeiramente, através de fluorescência de raios X, e em um segundo momento, por intermédio de análise Rietveld (difração de raios X). Neste último caso, a quantificação das composições químicas foram realizadas através das quantificações mineralógicas obtidas. O fluxograma apresentado na Figura 5.1 descreve melhor o processo.

Desse modo, como pode ser observado através do fluxograma, a partir da mesma amostra de britas coletada inicialmente, são gerados dois bancos de dados (um para cada

Figura 5.1: Fluxograma de Criação das Instâncias do Problema



Fonte: Elaborado pelo autor (2020)

quantificação realizada), de modo que cada um deles geram, portanto, um conjunto de instâncias para o problema, totalizando, dois grandes conjuntos de instâncias.

A primeira abordagem de investigação da composição química dos materiais foi por intermédio do método baseado em fluorescência de raios X. Após os ensaios laboratoriais serem realizados, foram obtidas as composições químicas para todas as britas. Um extrato com 20 britas do banco de dados gerado é apresentado na Tabela 5.1, enquanto que no Apêndice A são discriminadas as composições químicas para todas as britas conforme resultados obtidos através do método de fluorescência de raios X.

Através da Tabela 5.1, cada brita possui valores distintos para cada atributo, que na presente problemática são representados pelas composições químicas dos materiais. Por sua vez, a construção do segundo banco de dados foi realizada em duas etapas. Primeiramente, foram estabelecidos as composições mineralógicas para todas as britas coletadas através de uma abordagem de difração de raios X - Rietveld. Em um segundo momento, as composições mineralógicas obtidas foram tratadas matematicamente (baseado nas composições químicas de cada mineral presente nos agregados, Tabela 5.3) para obter-se as composições químicas referentes as britas. A Tabela 5.2 apresenta as composições mineralógicas obtidas para o mesmo extrato de britas considerados na Tabela 5.1, enquanto o banco de dados completo pode ser observado no Apêndice B.

Após a investigação mineralógica e o devido processo de conversão das composições, foi possível auferir as composições químicas das amostras de britas através do método de difração de raios X - Rietveld, as quais são apresentadas, no mesmo extrato considerado anteriormente, na Tabela 5.4. As composições químicas referentes ao banco de dados completo são detalhadas no Apêndice C.

Tabela 5.1: Composição Química das Britas - Método de Florescência de Raios X

Brita	Composição Química							
	Al ₂ O ₃	CaO	K ₂ O	SiO ₂	Fe ₂ O ₃	MgO	Na ₂ O	SO ₃
1	14,50	1,75	5,75	71,97	2,05	0,38	2,78	0,07
2	14,33	2,13	5,62	70,70	2,28	0,76	3,17	0,06
3	14,40	4,19	4,14	66,37	4,80	1,02	3,34	0,12
4	15,26	4,13	5,07	65,56	4,47	1,01	2,96	0,14
5	13,60	2,42	5,22	69,50	4,60	0,56	2,62	0,16
6	15,02	2,77	5,89	67,22	4,21	0,67	2,57	0,11
7	14,56	2,13	3,50	70,47	3,52	1,71	3,05	0,03
8	14,30	2,11	5,71	70,67	2,26	0,79	3,13	0,06
9	15,42	3,72	4,34	66,13	3,17	1,62	4,04	0,05
10	15,22	3,11	5,98	66,59	3,62	0,74	2,82	0,10
11	13,90	19,73	2,61	52,53	4,92	3,21	1,57	0,27
12	9,86	40,19	1,09	35,40	5,96	3,95	0,67	0,39
13	15,33	0,66	6,79	73,38	1,00	0,82	1,65	0,00
14	17,04	4,13	4,22	64,52	3,90	1,34	3,04	0,12
15	16,68	7,36	2,27	59,50	6,26	3,18	2,37	0,15
16	18,01	1,41	5,95	68,68	1,61	0,42	2,87	0,30
17	18,08	4,20	0,29	54,22	10,74	3,89	3,66	0,04
18	18,14	2,56	3,76	63,70	5,80	2,30	2,27	0,20
19	15,70	1,97	4,93	70,87	2,09	0,58	2,94	0,13
20	16,58	3,18	4,56	68,95	2,28	0,60	2,63	0,08
...	...	...	...	...	...	...	...	...

Conforme apresentado nas tabelas anteriores, há oito principais atributos no problema a ser tratado, representados pelas composições químicas Al₂O₃, CaO, K₂O, SiO₂, Fe₂O₃, MgO, Na₂O e SO₃. Tomando como base as explanações realizadas na descrição do problema, a clusterização dos materiais é tratada no presente trabalho em configurações de três atributos/características. Isto é, cada combinação de três atributos (independente da ordem), gera uma instância a ser tratado e resolvida pelo método proposto. Como há oito características, matematicamente a quantidade de instâncias a serem geradas é um processo de análise combinatória, representada por uma combinação de oito, três a três, que totalizam 56 instâncias.

Entretanto, frações entre os atributos também podem gerar novos atributos. Isso implica que, a divisão entre os valores obtidos para CaO e K₂O, por exemplo, para todas as britas, geram um novo atributo, neste caso, $\frac{CaO}{K_2O}$. O mesmo vale para todas as demais possibilidades de frações entre os atributos primários. Desse modo, foi necessário a implementação de um algoritmo iterativo que pudesse elencar todas as possibilidades de instâncias a serem mapeadas, levando em consideração a existência destes novos atributos destacados, denominados de atributos secundários para diferenciar dos iniciais.

Tabela 5.2: Composição Mineralógica de Extrato do Banco de Dados Coletado

RIETVELD - QUANTIFICAÇÃO MINERALÓGICA (%)																	
	SÍL.	MICA			FELDSPATO								OUTROS				
Amostra	Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita
1	38,4	2,7	0,0	0,0	26,3	0,0	0,0	0,0	32,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
2	18,4	4,3	0,0	0,0	30,5	0,0	0,0	0,0	43,9	0,0	0,0	0,0	2,9	0,0	0,0	0,0	0,0
3	17,3	10,9	0,0	0,0	36,1	0,0	0,0	0,0	21,3	0,0	0,0	0,0	10,2	4,2	0,0	0,0	0,0
4	1,6	19,4	0,0	0,0	61,5	0,0	0,0	0,0	0,0	17,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	34,7	15,8	0,0	0,0	42,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	7,0	0,0
6	41,5	4,1	0,0	0,0	32,5	0,0	0,0	0,0	21,6	0,0	0,0	0,0	0,4	0,0	0,0	0,0	0,0
7	37,4	0,0	12,3	0,0	41,3	0,0	0,0	0,0	4,3	0,0	0,0	0,0	0,0	4,8	0,0	0,0	0,0
8	25,3	3,3	0,0	0,0	40,6	0,0	0,0	0,0	18,5	0,0	0,0	0,0	12,4	0,0	0,0	0,0	0,0
9	11,6	0,0	10,0	0,0	41,6	0,0	0,0	0,0	23,3	0,0	0,0	0,0	13,6	0,0	0,0	0,0	0,0
10	16,7	0,0	5,6	8,8	20,6	0,0	0,0	0,0	37,6	0,0	0,0	10,7	0,0	0,0	0,0	0,0	0,0
11	28,2	0,0	14,1	0,0	16,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	10,7	16,9	0,0	14,1
12	24,9	3,8	0,0	0,0	24,4	0,0	0,0	0,0	28,3	11,7	0,0	0,0	6,9	0,0	0,0	0,0	0,0
13	38,9	2,7	8,8	0,0	4,4	0,0	0,0	0,0	21,6	7,5	6,9	9,3	0,0	0,0	0,0	0,0	0,0
14	15,1	6,6	0,0	0,0	12,0	0,0	0,0	0,0	18,8	10,9	0,0	21,9	14,7	0,0	0,0	0,0	0,0
15	8,9	8,9	5,8	0,0	0,0	0,0	39,2	0,0	12,0	0,0	9,5	0,0	15,9	0,0	0,0	0,0	0,0
16	27,0	2,3	0,0	0,0	30,5	0,0	0,0	0,0	32,6	0,0	0,0	7,6	0,0	0,0	0,0	0,0	0,0
17	17,5	0,0	0,0	0,0	67,4	0,0	0,0	0,0	0,0	0,0	0,0	12,5	0,0	2,6	0,0	0,0	0,0
18	40,0	10,1	15,8	0,0	21,6	0,0	0,0	0,0	7,2	0,0	0,0	0,0	4,7	0,6	0,0	0,0	0,0
19	29,0	5,4	5,9	0,0	32,9	0,0	0,0	0,0	21,0	0,0	0,0	0,0	5,9	0,0	0,0	0,0	0,0
20	27,4	11,0	11,6	0,0	20,5	0,0	10,6	0,0	17,7	0,0	0,0	0,0	0,0	1,2	0,0	0,0	0,0
21	33,6	1,0	4,7	0,0	25,8	0,0	0,0	0,0	34,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
22	34,8	9,8	9,2	0,0	30,3	0,0	0,0	0,0	0,0	0,0	14,6	0,0	0,0	1,2	0,0	0,0	0,0
23	33,4	3,0	16,8	0,0	38,8	0,0	0,0	0,0	7,2	0,0	0,0	0,0	0,0	0,8	0,0	0,0	0,0
24	18,6	7,4	8,5	0,0	12,4	0,0	0,0	0,0	26,3	11,7	0,0	0,0	14,6	0,5	0,0	0,0	0,0
25	22,2	7,5	0,0	0,0	28,6	0,0	0,0	0,0	29,0	0,0	0,0	0,0	11,7	1,1	0,0	0,0	0,0
26	22,8	8,9	7,7	0,0	29,8	0,0	0,0	0,0	26,4	0,0	0,0	0,0	4,5	0,0	0,0	0,0	0,0
27	28,2	2,8	4,2	0,0	31,0	0,0	0,0	0,0	33,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
28	32,3	4,7	0,0	0,0	28,7	0,0	0,0	0,0	19,0	0,0	12,8	0,0	2,6	0,0	0,0	0,0	0,0
29	29,1	0,8	2,4	0,0	27,3	0,0	0,0	0,0	40,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Tabela 5.3: Composição Química dos Minerais

QUANTIFICAÇÃO QUÍMICA DOS MINERAIS (%)																	
	SÍL.	MICA			FELDSPATO								OUTROS				
Mineral	Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita
CaO	0,0	0,0	0,0	0,0	1,1	0,0	19,2	0,0	0,0	12,4	0,0	4,2	13,7	0,0	56,0	45,1	30,4
SiO ₂	100,0	41,6	45,2	43,0	67,4	58,2	44,4	67,7	64,8	53,1	64,8	63,4	51,2	45,6	0,0	0,0	0,0
SO ₃	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
MgO	0,0	23,2	0,0	28,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	19,6	39,5	0,0	10,9	21,9
Al ₂ O ₃	0,0	11,8	38,4	12,2	20,4	26,6	35,8	19,2	18,3	30,0	18,3	23,1	10,9	0,0	0,0	0,0	0,0
Fe ₂ O ₃	0,0	8,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,4	0,0	0,0	0,0	0,0
K ₂ O	0,0	10,9	11,8	11,2	0,0	0,0	0,0	4,4	16,9	0,0	16,9	0,0	0,0	0,0	0,0	0,0	0,0
Na ₂ O	0,0	0,0	0,0	0,0	11,2	6,9	0,6	8,7	0,0	4,6	0,0	9,3	0,0	0,0	0,0	0,0	0,0

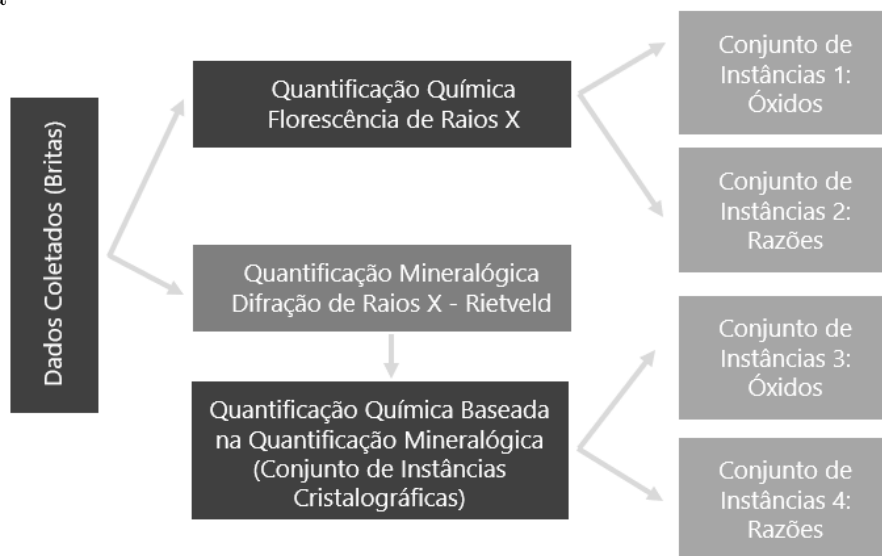
Tabela 5.4: Composição Química das Britas - Método Rietveld

Brita	Composição Química							
	Al ₂ O ₃	CaO	K ₂ O	SiO ₂	Fe ₂ O ₃	MgO	Na ₂ O	SO ₃
1	11,65	0,28	5,81	78,36	0,22	0,62	2,95	0,00
2	15,07	0,72	7,90	70,64	0,43	1,58	3,41	0,00
3	13,64	1,78	4,79	67,06	1,15	6,21	4,04	0,00
4	20,07	2,84	2,10	60,37	1,60	4,50	7,68	0,00
5	10,51	3,60	1,71	69,95	1,31	4,42	4,76	0,00
6	11,09	0,40	4,10	79,25	0,35	1,02	3,63	0,00
7	13,90	0,44	2,17	75,69	0,00	1,88	4,62	0,00
8	13,38	2,12	3,50	72,29	0,58	3,20	4,54	0,00
9	18,02	2,30	5,11	66,17	0,33	2,66	4,65	0,00
10	16,76	0,68	8,00	68,02	0,00	2,54	3,31	0,00
11	8,68	13,92	1,67	50,23	0,00	7,30	1,80	0,00
12	14,86	2,66	5,20	71,00	0,48	2,24	3,26	0,00
13	14,18	1,37	6,13	75,25	0,22	0,62	1,71	0,00
14	16,58	4,42	3,90	65,30	0,90	4,41	3,89	0,00
15	22,95	9,69	5,27	54,58	1,12	5,18	0,22	0,00
16	14,20	0,65	5,77	74,43	0,19	0,54	4,12	0,00
17	16,59	1,25	0,00	72,03	0,00	1,03	8,71	0,00
18	13,47	0,87	4,18	73,26	0,95	3,49	2,42	0,00
19	14,07	1,15	4,83	72,65	0,59	2,40	3,68	0,00
20	16,96	2,25	5,56	67,76	0,91	3,02	2,35	0,00
...	...	...	...	...	...	...	...	...

Através do algoritmo implementado foi possível mapear e criar todas as possíveis instâncias, de três a três atributos, as quais totalizaram 1428 instâncias. Devido a extensão das informações (em função da quantidade de instâncias), as configurações relacionadas a cada uma delas é apresentada em Anexo. Com isso, os dois conjuntos de instâncias obtidos inicialmente (conforme demonstra a Figura 5.1), são subdivididos em dois subconjuntos, cada uma. O primeiro, corresponde às instâncias que levam em consideração apenas os óxidos isolados apresentados, enquanto o segundo, leva em consideração os novos atributos gerados a partir das razões entre os óxidos, conforme destacado anteriormente. Assim, o fluxograma destacado na Figura 5.2 a seguir ilustra os conjuntos de instâncias finais obtidos a partir dos bancos de dados.

Após a aplicação do método de otimização desenvolvido nos conjuntos de instâncias, foi possível obter a melhor solução alcançada para cada instância do problema, geradas por intermédio das combinações de três a três atributos em cada conjunto. A seguir, nas Seções 5.2 e 5.3, os resultados obtidos são reportados e devidamente analisados. Em seguida, tais no Capítulo 6, todos os resultados são comparados entre si, com o intuito de realizar uma análise sistêmica dos resultados alcançados, assim como objetivando melhor

Figura 5.2: Conjunto de Instâncias Geradas a Partir dos Bancos de Dados Coletados para o Problema



Fonte: Elaborado pelo autor (2020)

compreender as melhores soluções e composições acerca do problema tratado.

5.2 Resultados Computacionais - Instâncias Químicas

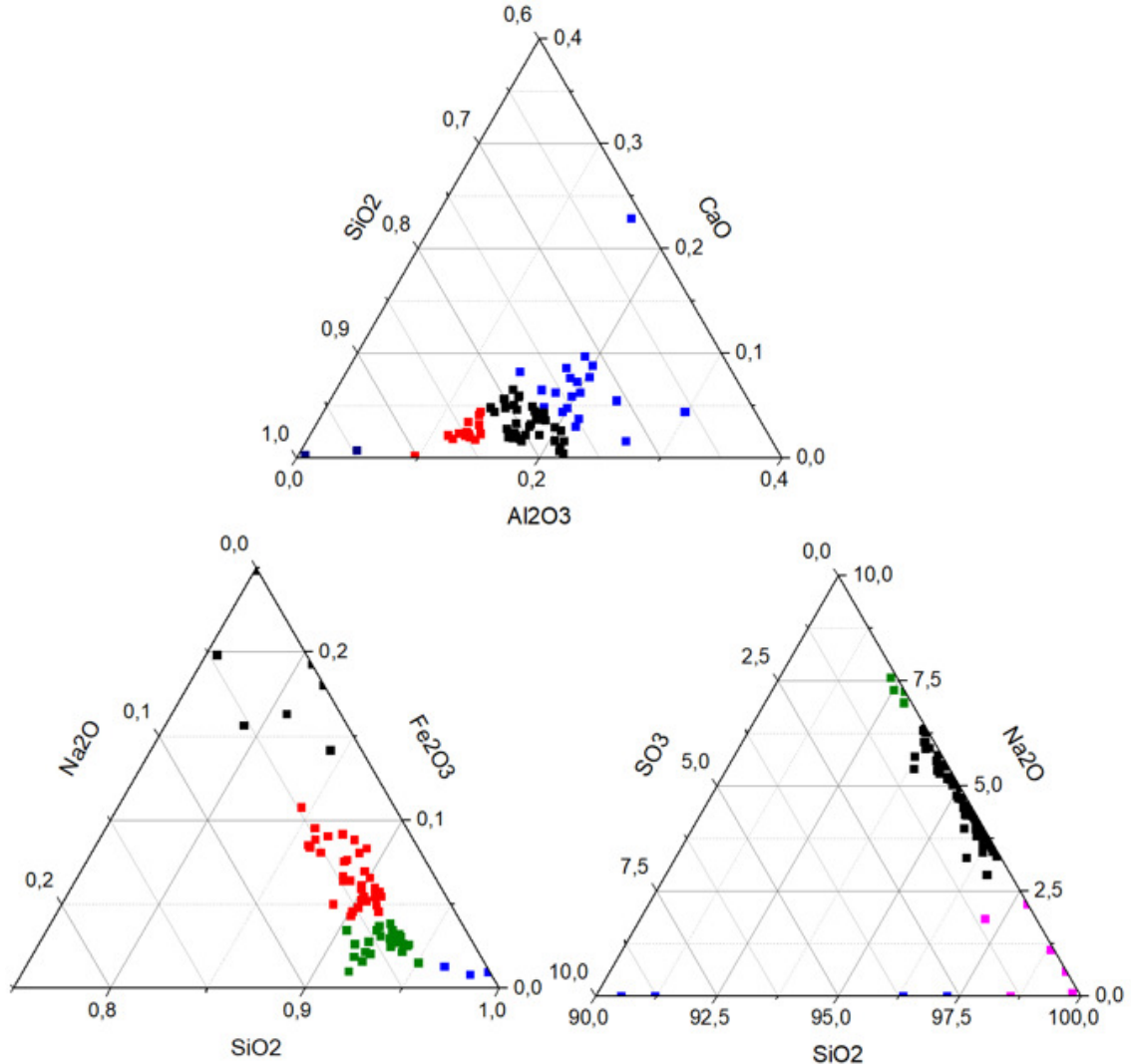
A presente seção objetiva apresentar os resultados computacionais alcançados após a execução do método de otimização desenvolvido em todas as instâncias químicas de seus subconjuntos, as quais foram obtidas através do banco de dados referentes às análises de fluorescência de raios X nos materiais coletados. Dessa forma, o método foi aplicado, inicialmente, nas 56 instâncias do subconjunto de óxidos e, posteriormente, executado nas 1428 instâncias advindas das combinações entre as razões dos óxidos considerados no problema.

Após o procedimento de otimização executado pelo método, cada melhor solução, referente a cada uma das instâncias, bem como seus respectivos resultados, foram coletados e tratados através da realização de um procedimento de ranqueamento das soluções obtidas para cada um dos dois conjuntos de instâncias (óxidos e razões), de modo que fosse possível reportar os resultados de forma mais eficiente e ilustrativa, respeitando as particularidades envolvidas nas duas classes de conjuntos.

Para o conjunto de instâncias referente aos óxidos, as três melhores soluções obtidas foram para as instâncias $[\text{Al}_2\text{O}_3, \text{CaO}, \text{SiO}_2]$, $[\text{SiO}_2, \text{Fe}_2\text{O}_3, \text{Na}_2\text{O}]$ e $[\text{SiO}_2, \text{Na}_2\text{O}, \text{SO}_3]$, respectivamente, demonstrando uma grande participação dos óxidos SiO_2 , CaO e Na_2O

na efetividade do processo de clusterização. A Figura 5.3 ilustra as soluções obtidas em ternários.

Figura 5.3: Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Óxidos

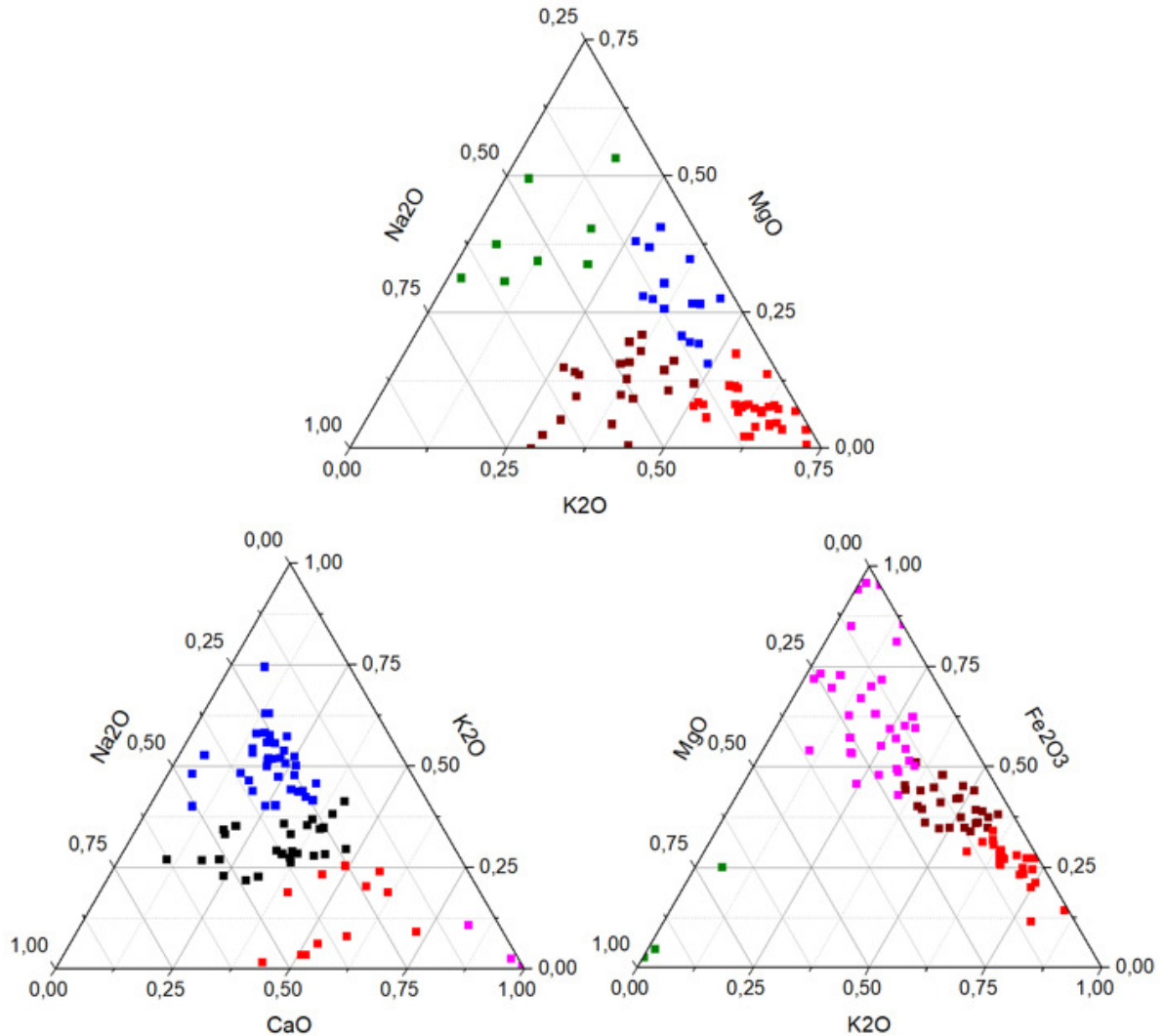


Fonte: Elaborado pelo autor (2020)

Conforme pode ser observado, as três melhores soluções tiveram os agregados clusterizados em quatro ou cinco clusters, a depender da instância, com representações bem definidas e delimitadas nos ternários representados. Com o intuito de melhor compreender as variáveis que envolvem a problemática, também foram analisadas as piores soluções dentre as melhores soluções obtidas. Considerando o conjunto de instâncias óxidos, as três piores soluções obtidas foram $[\text{K}_2\text{O}, \text{MgO}, \text{Na}_2\text{O}]$, $[\text{CaO}, \text{K}_2\text{O}, \text{Na}_2\text{O}]$ e $[\text{K}_2\text{O}, \text{Fe}_2\text{O}_3, \text{MgO}]$, respectivamente. As soluções são apresentadas na Figura 5.4.

Conforme pode ser observado, há uma forte presença dos óxidos MgO e K_2O nas soluções apresentadas. Além disso, trata-se de dois óxidos que não estiveram presentes

Figura 5.4: Piores Soluções do Conjunto de Instâncias Químicas - Subconjunto Óxidos



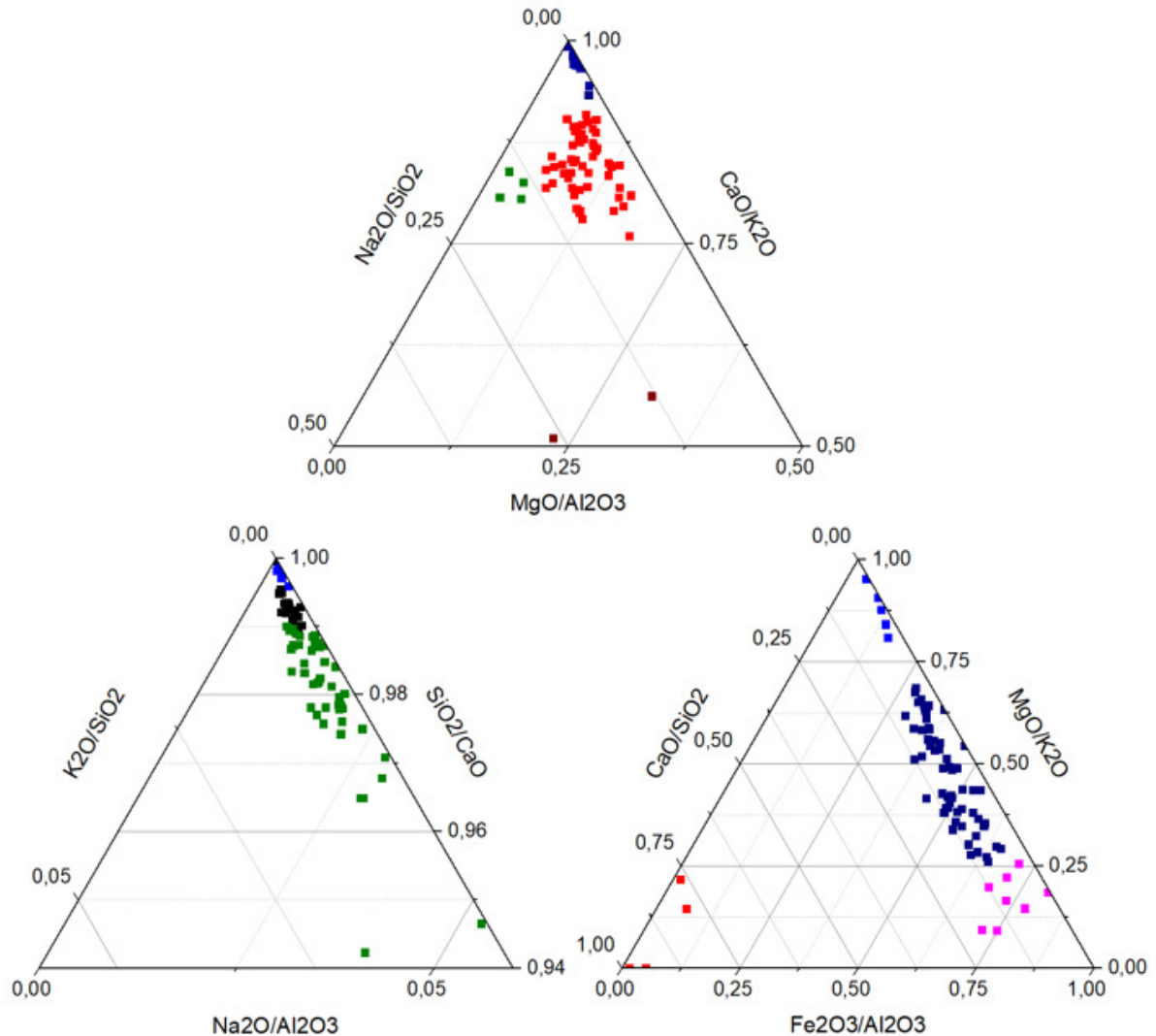
Fonte: Elaborado pelo autor (2020)

em nenhuma das melhores soluções destacadas anteriormente, reforçando uma tendência de dissimilaridade associadas aos óxidos isolados no processo de clusterização. Através da Figura 5.4 é possível também constatar a forte dissimilaridade destacada, com os dados extremamente espaçados e, por vezes sem uma definição e/ou delimitação eficiente dos clusters.

Os mesmos critérios do estudo realizado para os óxidos foram levados em consideração para avaliar os resultados obtidos para o conjunto de instâncias referente as razões dos óxidos. Desse modo, as instâncias onde foram obtidas as três melhores soluções após o procedimento de otimização, foram: $[\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}, \frac{\text{SiO}_2}{\text{CaO}}, \frac{\text{K}_2\text{O}}{\text{SiO}_2}]$, $[\frac{\text{SO}_3}{\text{Al}_2\text{O}_3}, \frac{\text{MgO}}{\text{K}_2\text{O}}, \frac{\text{SO}_3}{\text{SiO}_2}]$ e $[\frac{\text{MgO}}{\text{Al}_2\text{O}_3}, \frac{\text{MgO}}{\text{K}_2\text{O}}, \frac{\text{MgO}}{\text{SiO}_2}]$. A Figura 5.5 a seguir ilustra as melhores soluções destacadas.

É perceptível, através das soluções relatadas, que há uma forte similaridade entre as

Figura 5.5: Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões

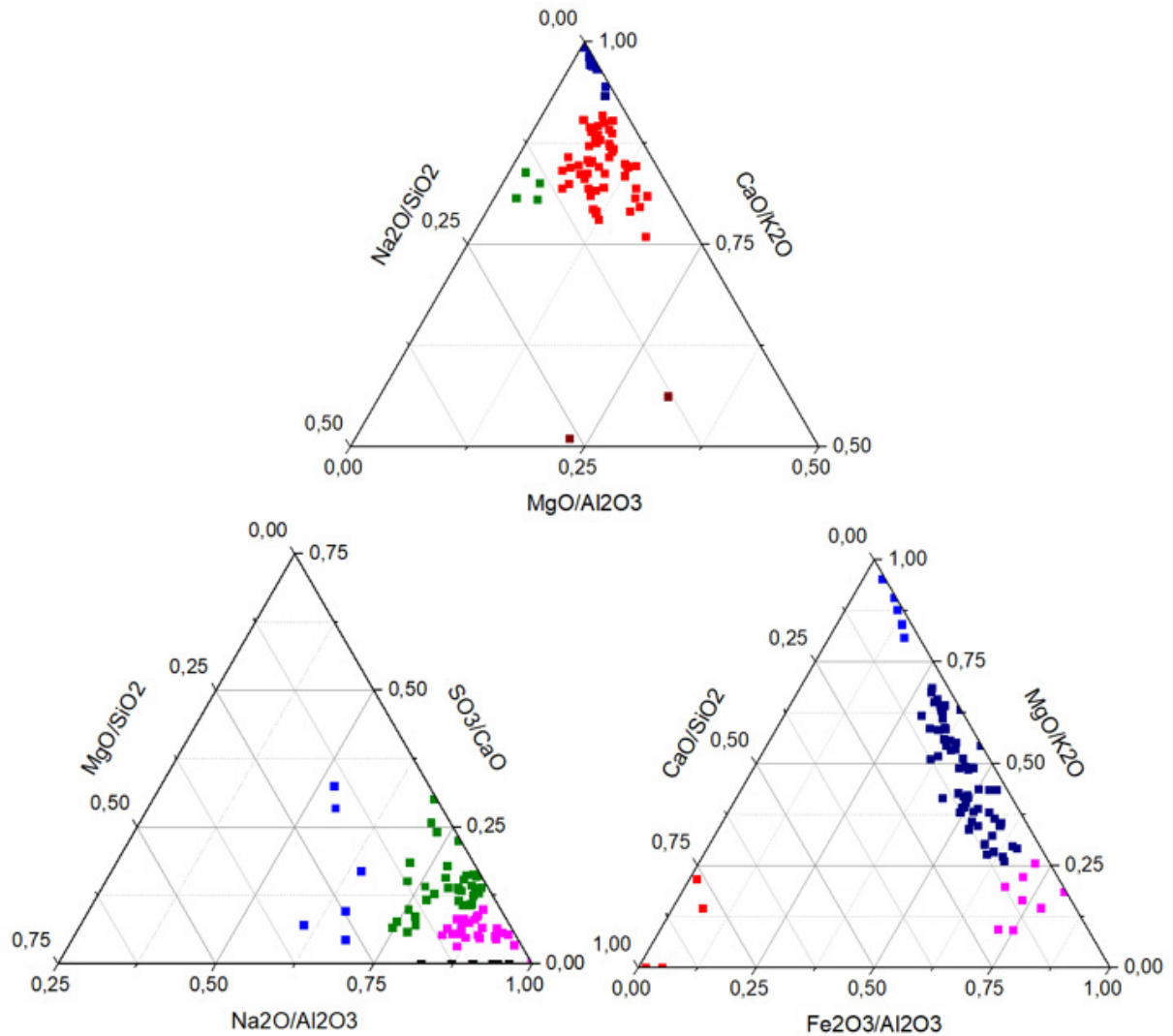


Fonte: Elaborado pelo autor (2020)

clusterizações realizadas nas soluções, abrangendo de três a quatro clusters, a depender da solução observada. Também foi realizada uma avaliação das melhores soluções consideradas puras, isto é, soluções onde as três razões consideradas nas instâncias sejam compostas por óxidos independentes. Deste modo, para que a instância seja considerada pura, os seis óxidos envolvidos nas razões da instância devem ser independentes. De modo a tornar o ranqueamento de soluções puras ainda mais independente, foi garantido que cada uma das razões existentes na melhor solução ranqueada seguinte não possua nenhuma das razões presentes na melhor solução de instância anterior. Com isso, as três melhores soluções de instâncias puras são apresentadas na Figura 5.6.

Por sua vez, a Figura 5.7 as instâncias que obtiveram o pior desempenho dentre as que compõe o conjunto de instâncias de razões dos óxidos, sendo possível observar uma

Figura 5.6: Melhores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões

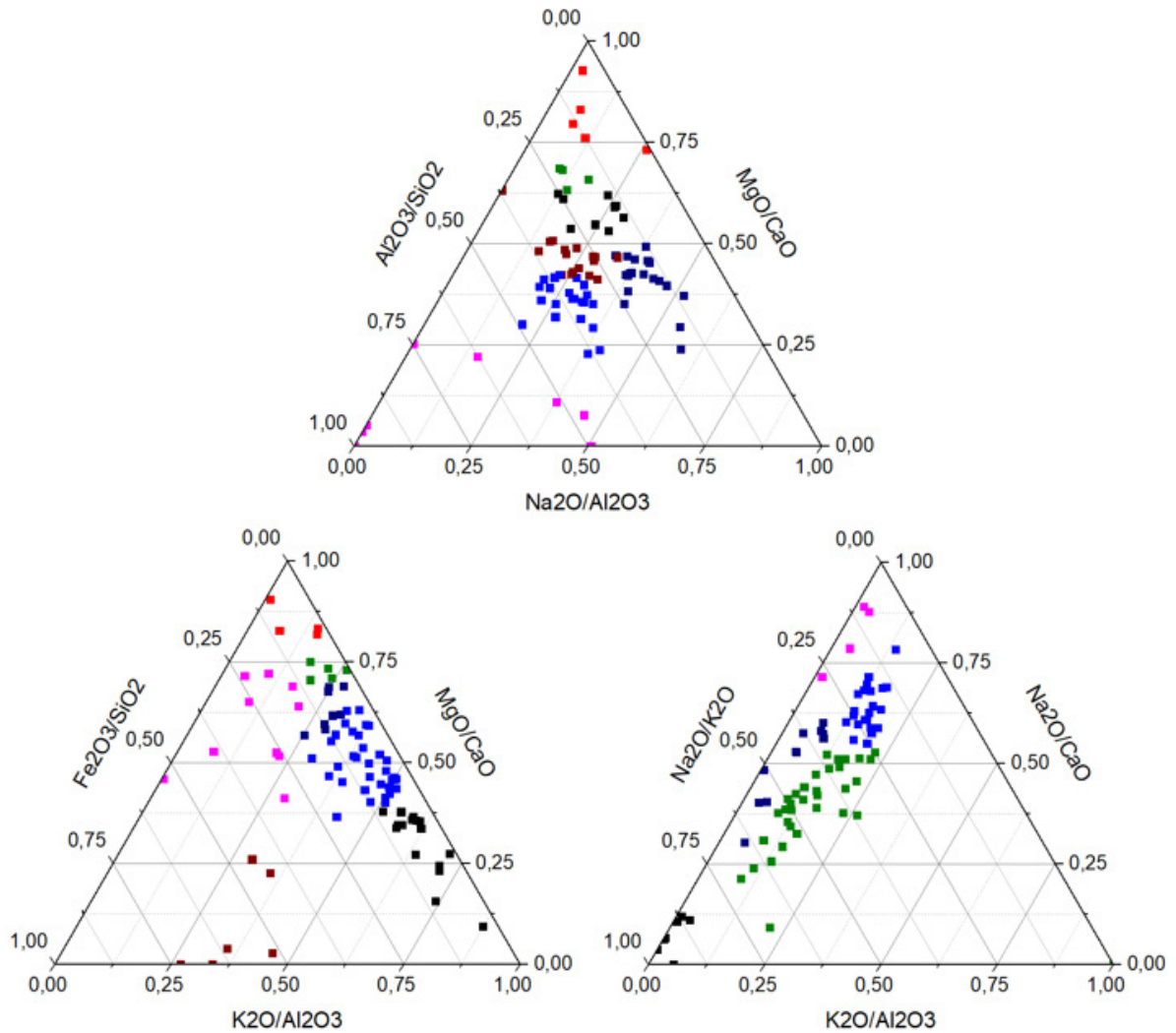


Fonte: Elaborado pelo autor (2020)

descoordenação significativa entre os clusters formados, bem como uma falta de linha de tendência no espaço amostral.

Devido ao grande volume de instâncias, especialmente considerando o conjunto de razões dos óxidos, optou-se por avaliar, também, a frequência dos óxidos e de suas razões nas 20% e 5% melhores soluções obtidas, respectivamente, para cada um de seus conjuntos de instâncias. A definição dos percentuais utilizados para tal representação levou em consideração o volume significativo de instâncias para os conjuntos. Testes preliminares foram realizados com percentuais variando entre 10% e 50%, no caso do conjunto de óxidos, e entre 2% e 15%, no caso do conjunto de razões. Os resultados dos testes de parametrização possibilitaram a visualização de que os percentuais definidos representam de forma eficiente os resultados obtidos para os dois conjuntos de instâncias.

Figura 5.7: Piores Soluções do Conjunto de Instâncias Químicas - Subconjunto Razões

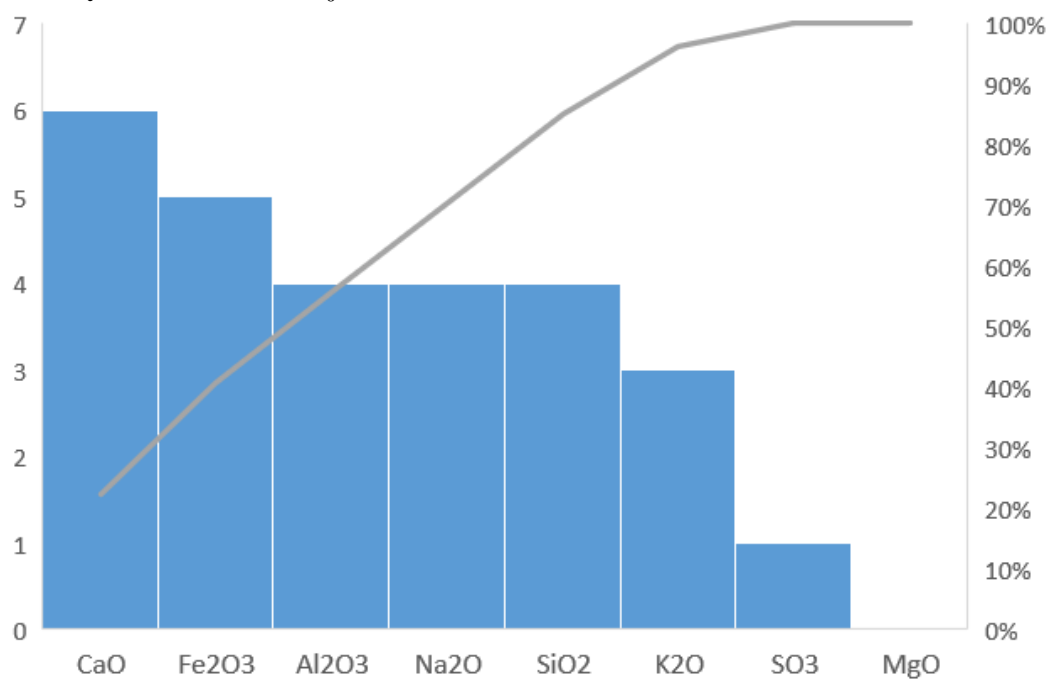


Fonte: Elaborado pelo autor (2020)

Desse modo, dado que há uma quantidade significativa de instâncias, apresentar e levar em consideração apenas as três melhores e piores instâncias levaria a conclusões limitadas (ou até mesmo precipitadas) acerca da problemática tratada no presente estudo. Assim, o objetivo principal da análise em questão é investigar quais são os óxidos e razões mais promissores na obtenção de clusterizações de maior qualidade, isto é, quais destes atributos possibilitam um melhor e mais eficiente agrupamento. Em posse disso, é possível auferir, não apenas quais são as melhores soluções (e suas instâncias) obtidas para o problema, mas definir um mapeamento de eficiência em torno de todos os atributos/componentes que estão presentes na problemática, e consequentemente, no processo decisório. As Figuras 5.8 e 5.9 ilustram as frequências alcançadas nas melhores soluções.

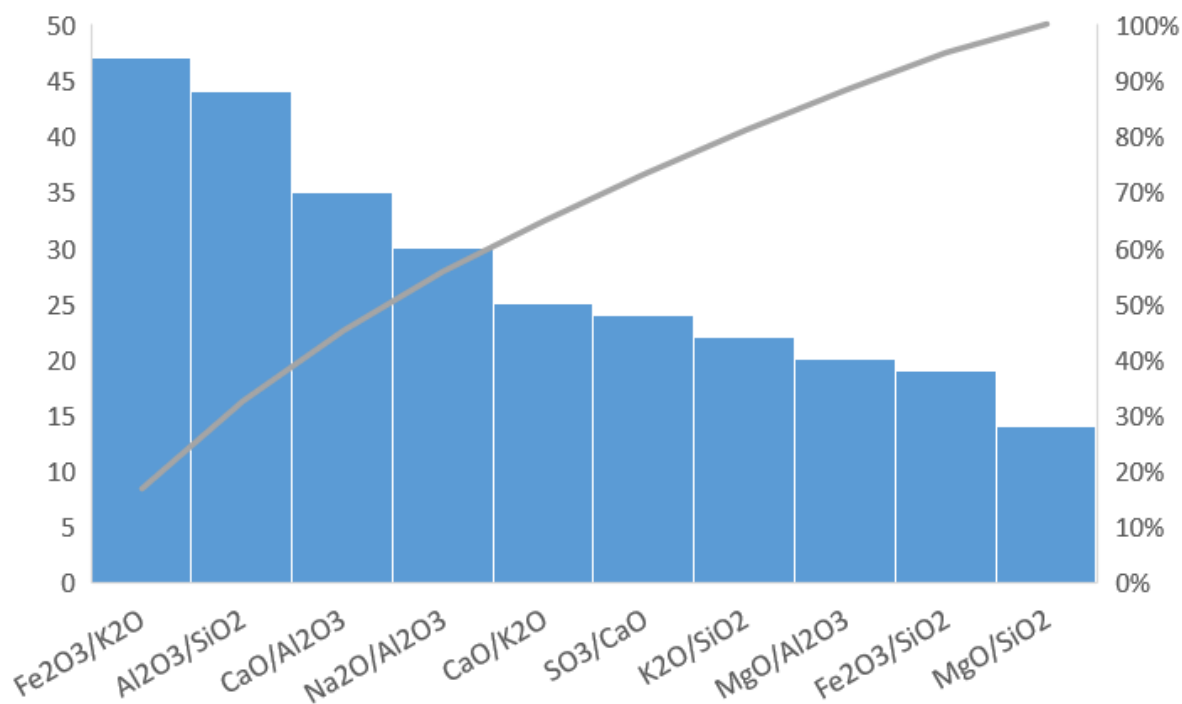
Conforme pode ser observado na Figura 5.8, há uma forte presença de CaO , Fe_2O_3 ,

Figura 5.8: Frequências dos Óxidos nas 20% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Óxidos



Fonte: Elaborado pelo autor (2020)

Figura 5.9: Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Razões

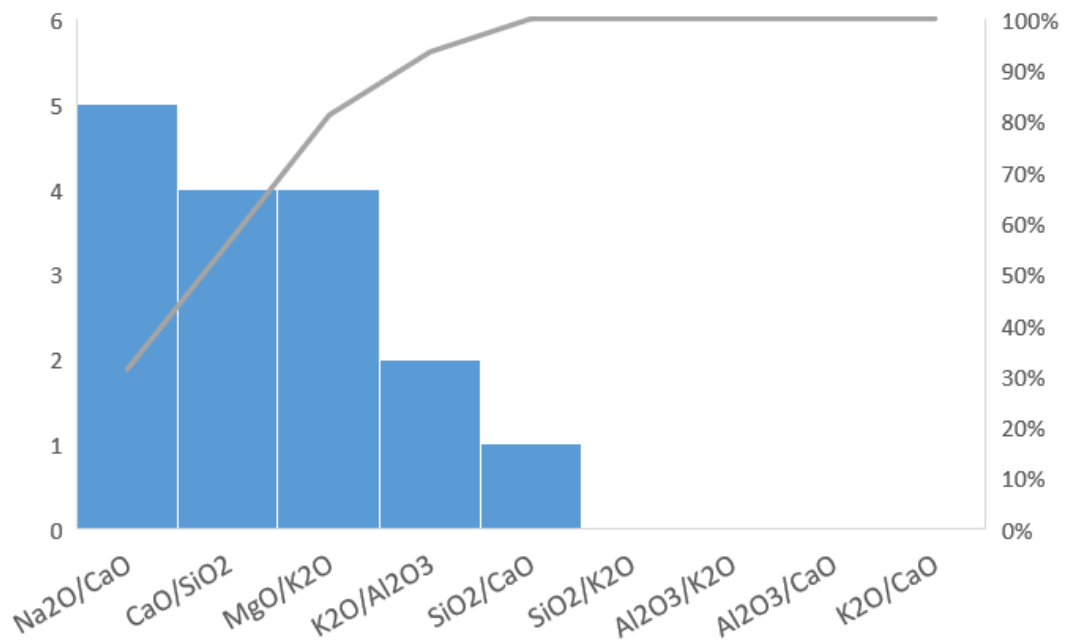


Fonte: Elaborado pelo autor (2020)

Al_2O_3 e SiO_2 nas melhores soluções encontradas, ao mesmo tempo em que K_2O , SO_3 e MgO representam os óxidos com menor frequências nas 20% melhores soluções referentes ao conjunto de instâncias de óxidos. Tais análises permitem inferir que as três melhores soluções estão dentro do esperado quando avaliado um conjunto maior das melhores soluções. Além disso, é possível perceber uma grande influência de alguns óxidos na formação de clusters de maiores níveis de qualidade.

De forma análoga, foram investigadas as frequências relativas as 5% melhores soluções obtidas para o conjunto de instâncias referentes as razões dos óxidos. As razões $\frac{\text{Fe}_2\text{O}_3}{\text{K}_2\text{O}}$, $\frac{\text{Al}_2\text{O}_3}{\text{SiO}_2}$, $\frac{\text{CaO}}{\text{Al}_2\text{O}_3}$ e $\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}$, conforme ilustra a Figura 5.9, foram as que melhor impactaram positivamente na qualidade das clusterizações realizadas, possibilitando uma agrupamento eficiente dos agregados. Visando uma melhor compreensão da influência dos atributos na qualidade das soluções obtidas, foram auferidas também as razões com menores frequências nas melhores soluções. A Figura 5.10 ilustra as piores frequências alcançadas, enfatizando as razões $\frac{\text{SiO}_2}{\text{CaO}}$, $\frac{\text{SiO}_2}{\text{K}_2\text{O}}$, $\frac{\text{Al}_2\text{O}_3}{\text{K}_2\text{O}}$, $\frac{\text{Al}_2\text{O}_3}{\text{CaO}}$ e $\frac{\text{K}_2\text{O}}{\text{CaO}}$, as quais não foram registradas presenças em nenhuma das melhores soluções para o problema, demonstrando uma forte dissimilaridade entre as demais composições que caracterizam os materiais.

Figura 5.10: Menores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Químicas - Subconjunto Razões



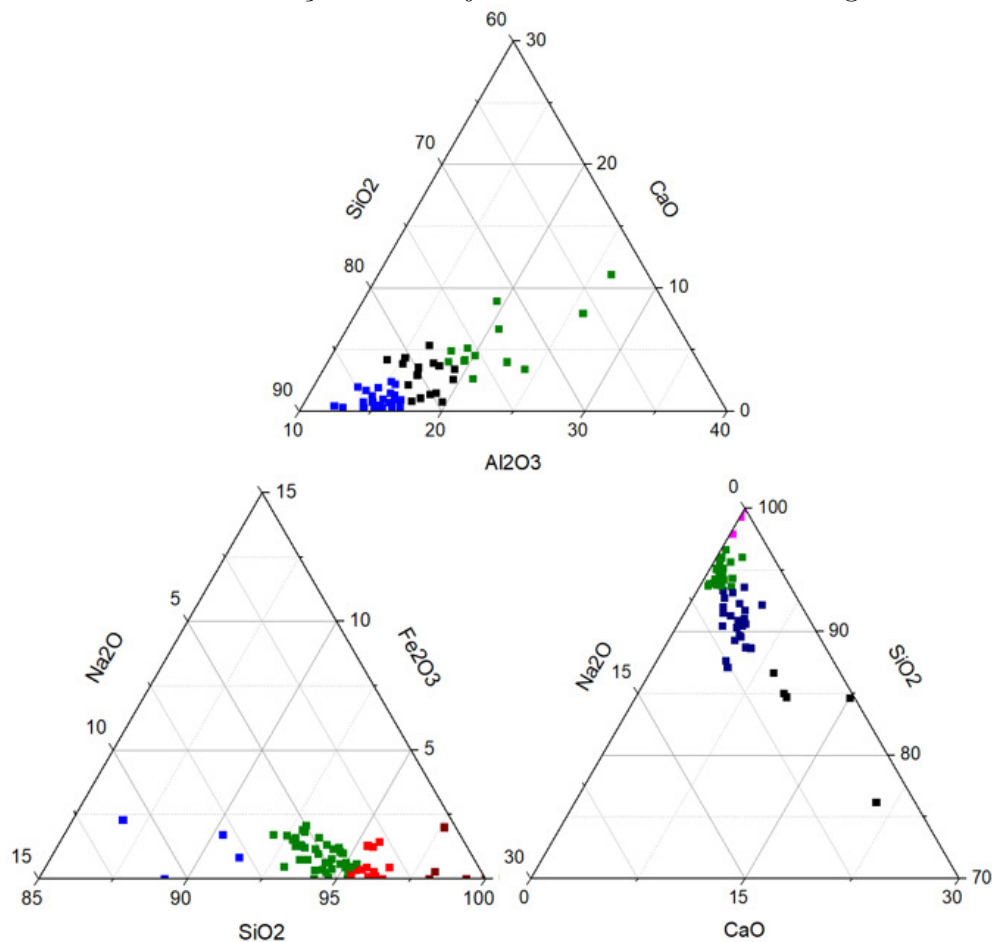
Fonte: Elaborado pelo autor (2020)

5.3 Resultados Computacionais - Instâncias Cristalográficas

Para os conjuntos de instâncias referentes ao banco de dados da análise cristalográfica, foram realizadas as mesmas análises relatadas anteriormente para os conjuntos referente ao banco de dados da análise química. Foram investigadas, inicialmente, as melhores soluções obtidas para cada conjunto de instâncias otimizado e as soluções de menor qualidade encontradas. Em um momento posterior, foi realizado um estudo acerca das frequências dos óxidos e razões dos óxidos nas melhores soluções obtidas, visando mapear os principais atributos caracterizadores no processo de clusterização.

Deste modo, a Figura 5.11 representa as melhores soluções alcançadas após a execução do método de otimização desenvolvido em todas as instâncias do conjunto de instâncias dos óxidos. As melhores soluções foram obtidas para as instâncias $[\text{Al}_2\text{O}_3, \text{CaO}, \text{SiO}_2]$, $[\text{SiO}_2, \text{Fe}_2\text{O}_3, \text{Na}_2\text{O}]$ e $[\text{CaO}, \text{SiO}_2, \text{Na}_2\text{O}]$, conforme ilustra a Figura.

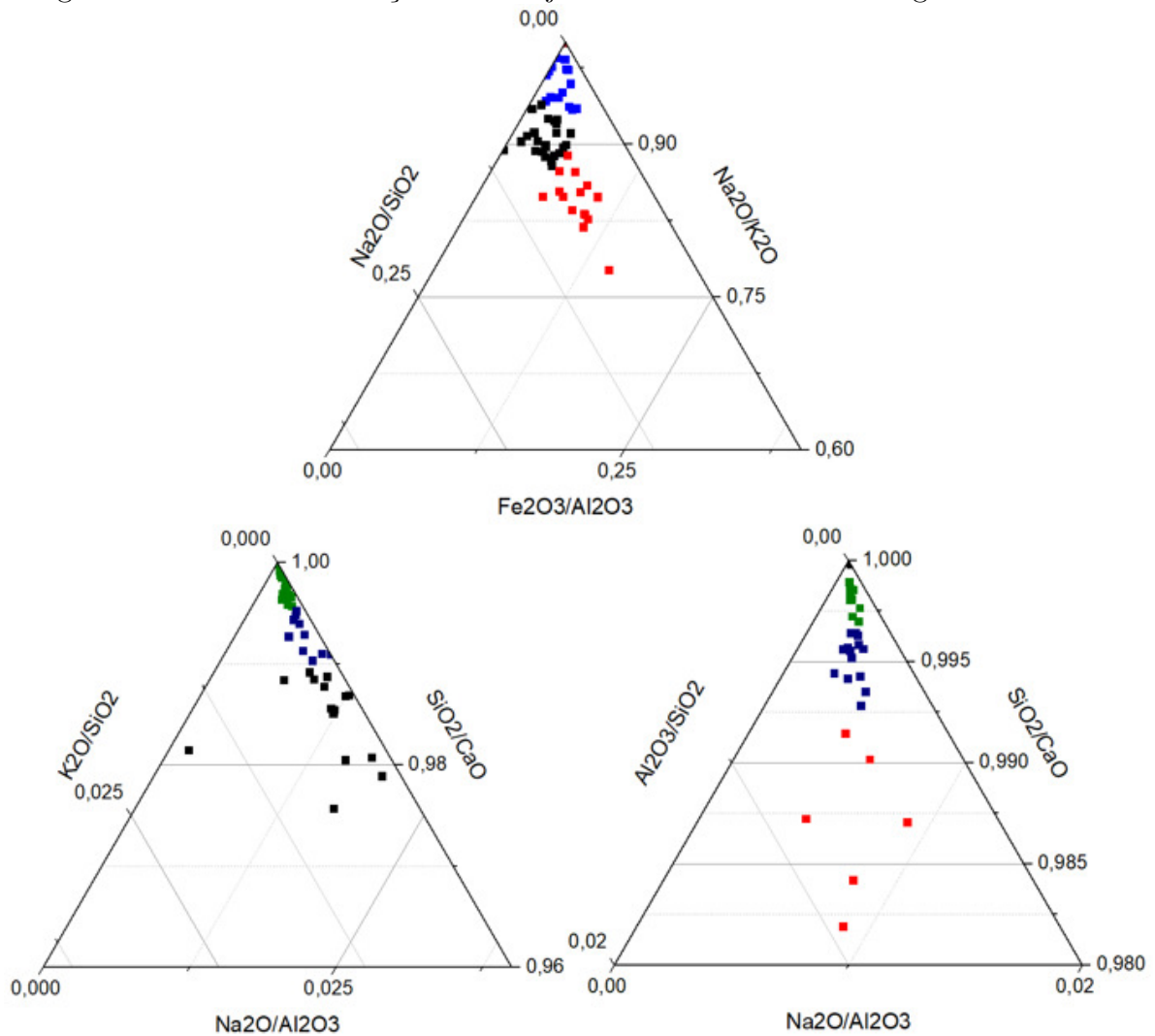
Figura 5.11: Melhores Soluções do Conjunto de Instâncias Cristalográficas - Óxidos



Fonte: Elaborado pelo autor (2020)

É possível observar, através das três melhores instâncias apresentadas, uma forte presença dos óxidos SiO_2 , Na_2O e CaO . Do mesmo modo, foi analisado os resultados obtidos para o conjunto de instâncias das razões de óxidos. As três melhores soluções alcançadas são relatadas na Figura 5.12 a seguir.

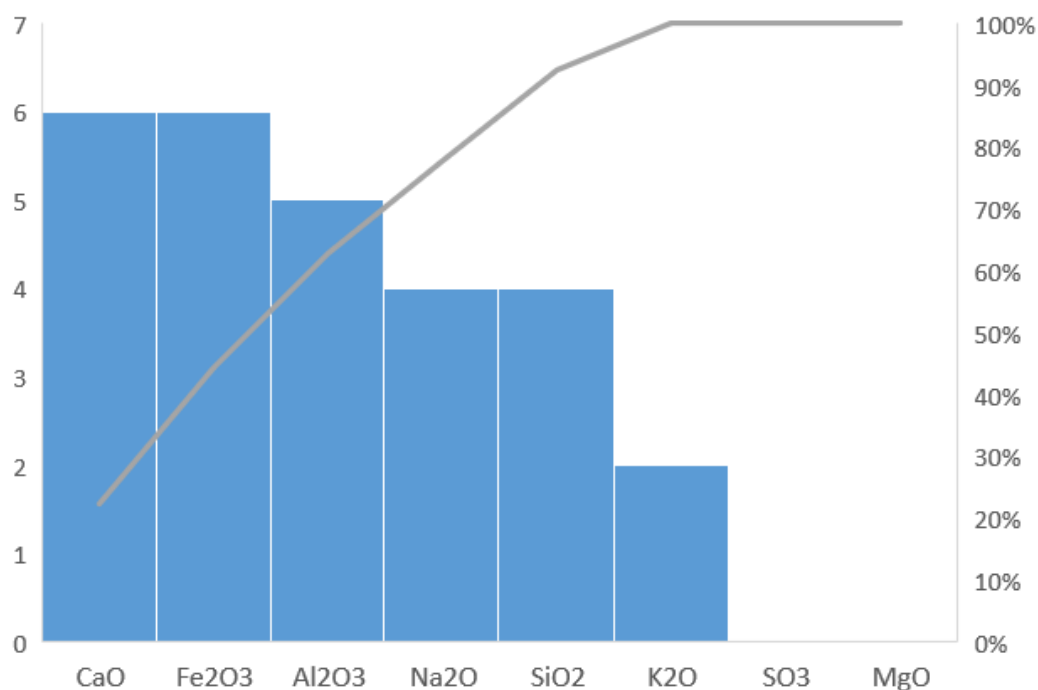
Figura 5.12: Melhores Soluções do Conjunto de Instâncias Cristalográficas - Razões



Fonte: Elaborado pelo autor (2020)

De forma análoga ao realizado nas instâncias do banco de dados advindo da análise química, devido ao grande volume de instâncias, foi realizado um mapeamento das maiores e menores frequências dos óxidos e de suas razões nas 20% e 5% melhores soluções obtidas, respectivamente, para cada um de seus conjuntos de instâncias. Os resultados alcançados no estudo permitiram que fossem auferidas os principais atributos responsáveis pelas melhores clusterizações realizadas dos dados do problema. A Figura 5.13 apresenta as melhores frequências obtidas para o conjunto de instâncias dos óxidos.

Figura 5.13: Frequências dos Óxidos nas 20% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Óxidos

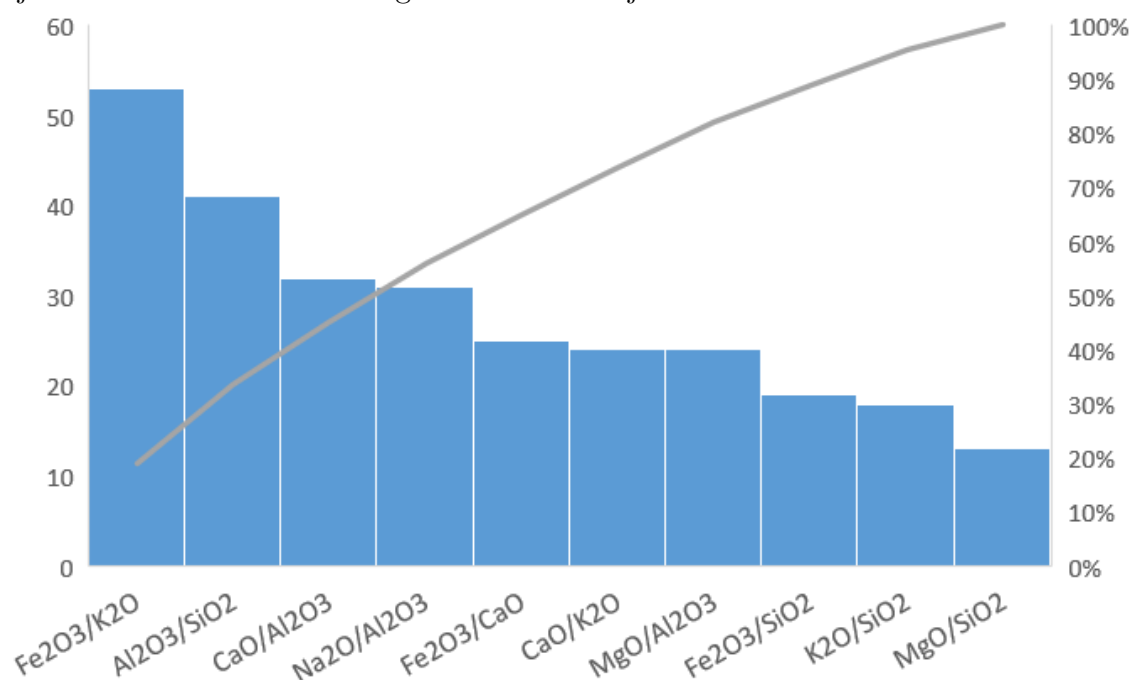


Fonte: Elaborado pelo autor (2020)

O estudo realizado permitiu observar que os óxidos que impactaram de forma mais positiva na qualidade das soluções obtidas foram o CaO e o Fe_2O_3 , enquanto K_2O e MgO foram os que obtiveram uma menor presença nas melhores soluções. Em especial ao caso do MgO, não houve nenhuma presença. Um outro insight interessante observado foi de que, apesar do SiO_2 estar presente nas três melhores soluções obtidas, considerando o conjunto das 20% melhores, como foi investigado, o óxido não está entre as que possuem uma maior frequência. A mesma análise foi realizada para o conjunto de instâncias das razões do óxidos, e os resultados são relatados na Figura 5.14.

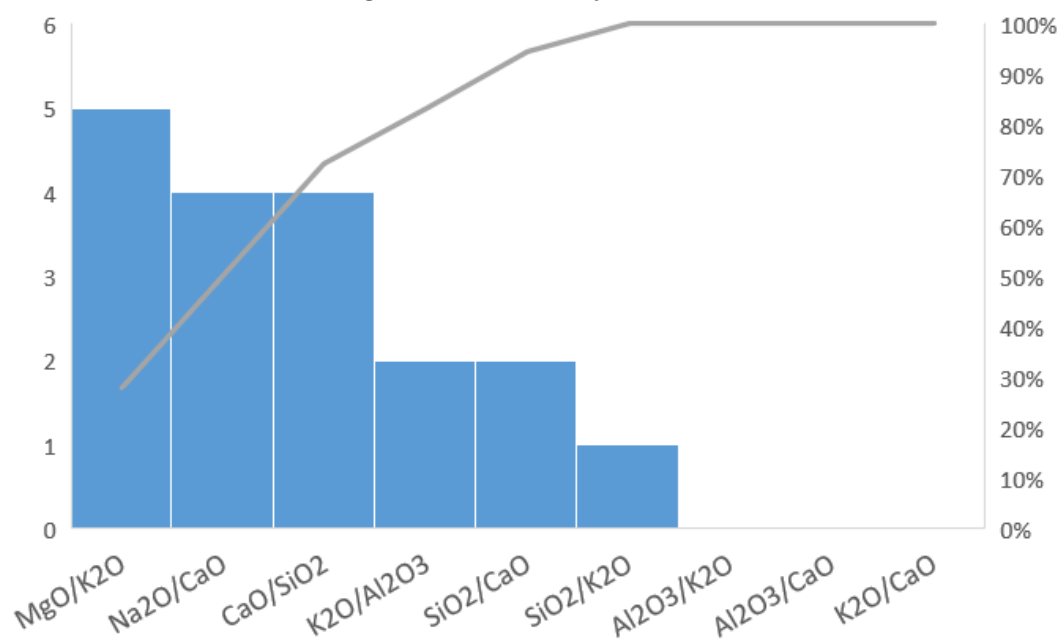
Os resultados do estudo permitiram verificar que uma presença significativa das razões $\frac{\text{Fe}_2\text{O}_3}{\text{K}_2\text{O}}$, $\frac{\text{Al}_2\text{O}_3}{\text{SiO}_2}$, $\frac{\text{CaO}}{\text{Al}_2\text{O}_3}$ e $\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}$ nas instâncias onde foram obtidas as melhores soluções para o problema de clusterização tratado. Por sua vez, foram verificadas também as razões que apresentaram menores presenças nas melhores instâncias em questão, seguindo o mesmo relato feito para o banco de dados da análise química. A análise permitiu identificar as razões com menores desempenho de clusterização, a exemplo das razões $\frac{\text{Al}_2\text{O}_3}{\text{CaO}}$ e $\frac{\text{K}_2\text{O}}{\text{CaO}}$, conforme apresentado na Figura 5.15. Por fim, de modo a proporcionar uma melhor visualização dos resultados, a Figura 5.16 apresenta uma visão geral das melhores soluções obtidas para cada um dos conjuntos de instâncias considerados como objetos de estudo na presente tese.

Figura 5.14: Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Razões



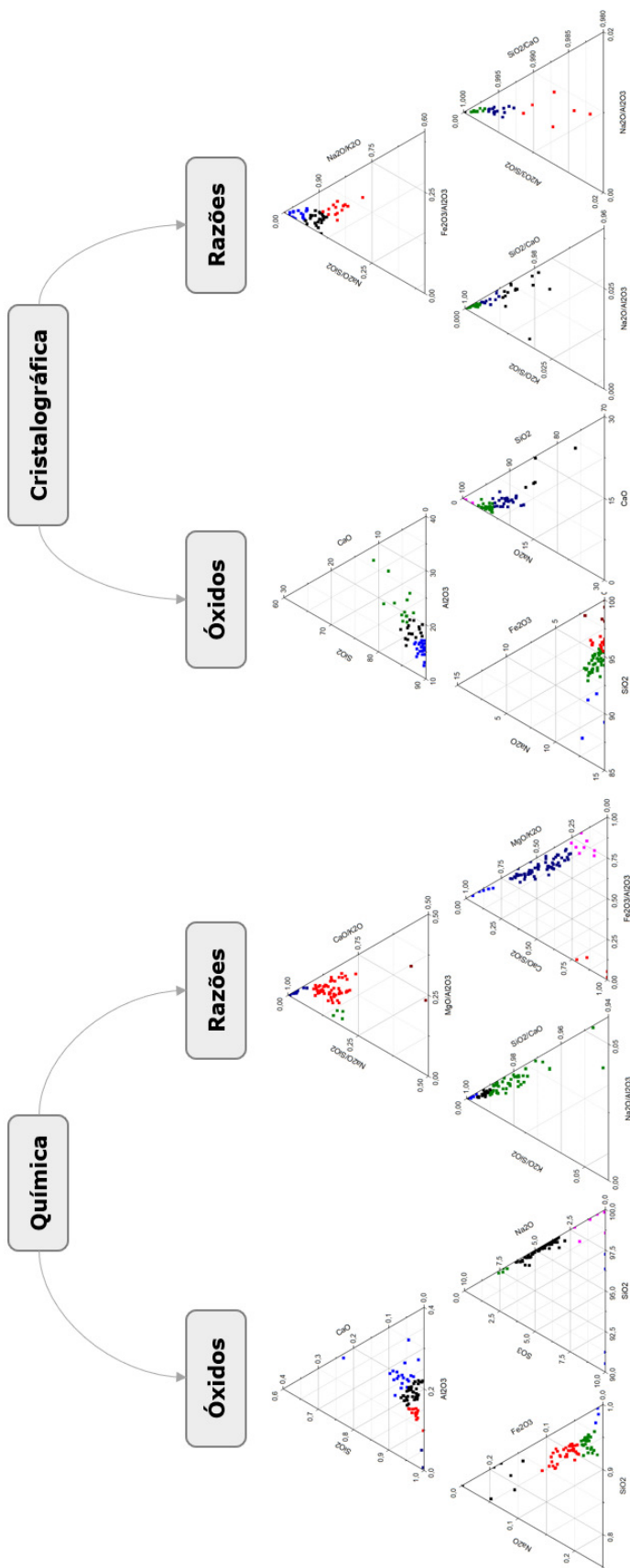
Fonte: Elaborado pelo autor (2020)

Figura 5.15: Menores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Conjunto de Instâncias Cristalográficas - Subconjunto Razões



Fonte: Elaborado pelo autor (2020)

Figura 5.16: Overview Melhores Soluções por Banco de Dados e Conjunto de Instâncias



Fonte: Elaborado pelo autor (2020)

Capítulo 6

Resultados e Discussões - Parte III: Análise e Comparação dos Resultados Alcançados por Conjunto de Instâncias

Aplicar o método desenvolvido e avaliar os resultados quanto à sua eficiência e desempenho matemático e computacional nas clusterizações alcançadas, bem como se há indicação da prevalência de algum elemento químico (ou conjunto de elementos) que determinem a formação das melhores clusterizações no ranqueamento.

O presente capítulo objetiva realizar uma análise comparativa entre os resultados obtidos pela aplicação do método de otimização nas instâncias dos dois bancos de dados (referentes às coletas realizadas por intermédio de análises químicas e cristalográficas), os quais foram devidamente apresentados nas Seções 5.2 e 5.3.

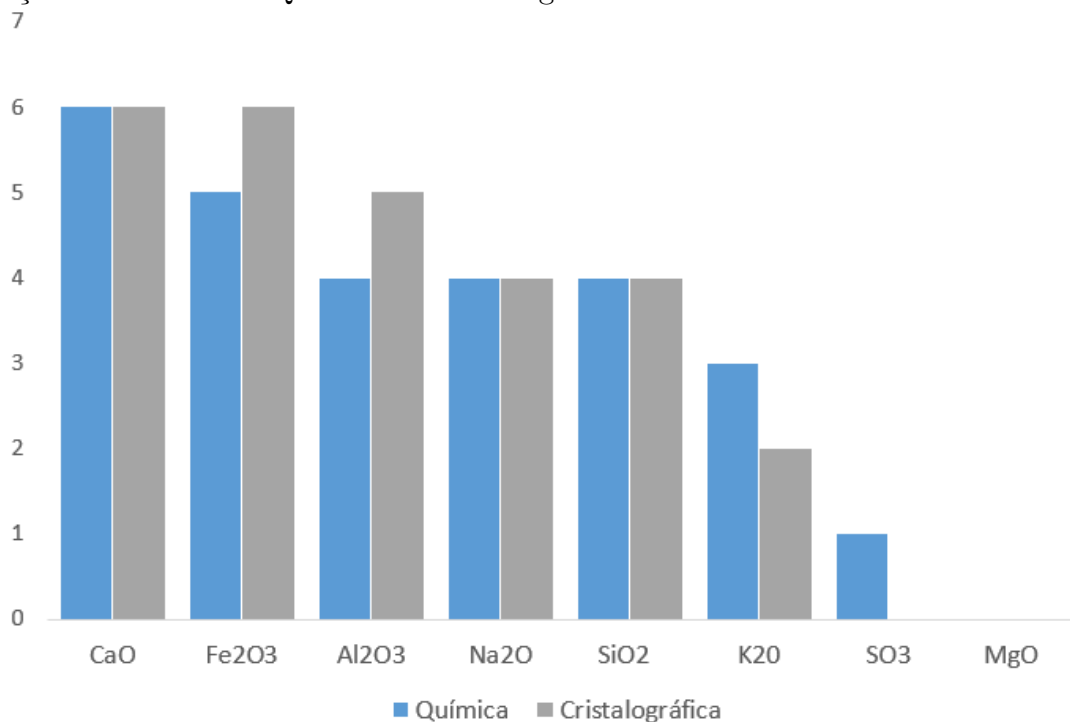
As análises realizadas foram divididas em duas etapas. A primeira, visando comparar de forma sistêmica e analítica os resultados de clusterização obtidos pela execução do método em cada uma das instâncias. E uma segunda etapa do estudo, que objetiva avaliar e destacar aspectos e tendências acerca da influência mineralógica nos agrupamentos formados durante o processo de otimização em clusterização.

6.1 Análise Sistêmica dos Resultados

De modo a avaliar a congruência entre os resultados obtidos para as instâncias com as mesmas configurações (atributos) dos bancos de dados distintos, foi realizada uma análise sistêmica das frequências obtidas para as melhores soluções encontradas. O principal objetivo desta análise é mensurar se há uma sinergia entre as frequências de óxidos (considerando os conjuntos de instâncias dos óxidos) e de razões de óxidos (conjuntos de razões) nas melhores soluções encontradas após realizados os procedimentos de otimização em todas as instâncias do problema.

Com relação as maiores frequências observadas nas 20% melhores soluções alcançadas para as instâncias dos conjuntos de óxidos, foi constatado que há uma sinergia significativa entre os óxidos com maiores frequências, assim como entre os óxidos de menores frequências. Nas instâncias dos dois bancos de dados, os óxidos CaO , Fe_2O_3 e Al_2O_3 se destacaram, nesta ordem, assumindo as posições que possuem maiores frequências. Por outro lado, os óxidos MgO , K_2O e SO_3 foram responsáveis pelas menores frequências. É importante observar ainda que, tanto no caso das instâncias referentes a análise química, como no caso das instâncias advindas da análise cristalográfica, o MgO não registrou aparições entre as instâncias em que foram obtidas os melhores desempenhos nas clusterizações alcançadas. A Figura 6.1 destaca os resultados da análise conjunta em questão.

Figura 6.1: Comparação das Maiores Frequências dos Óxidos nas 20% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas

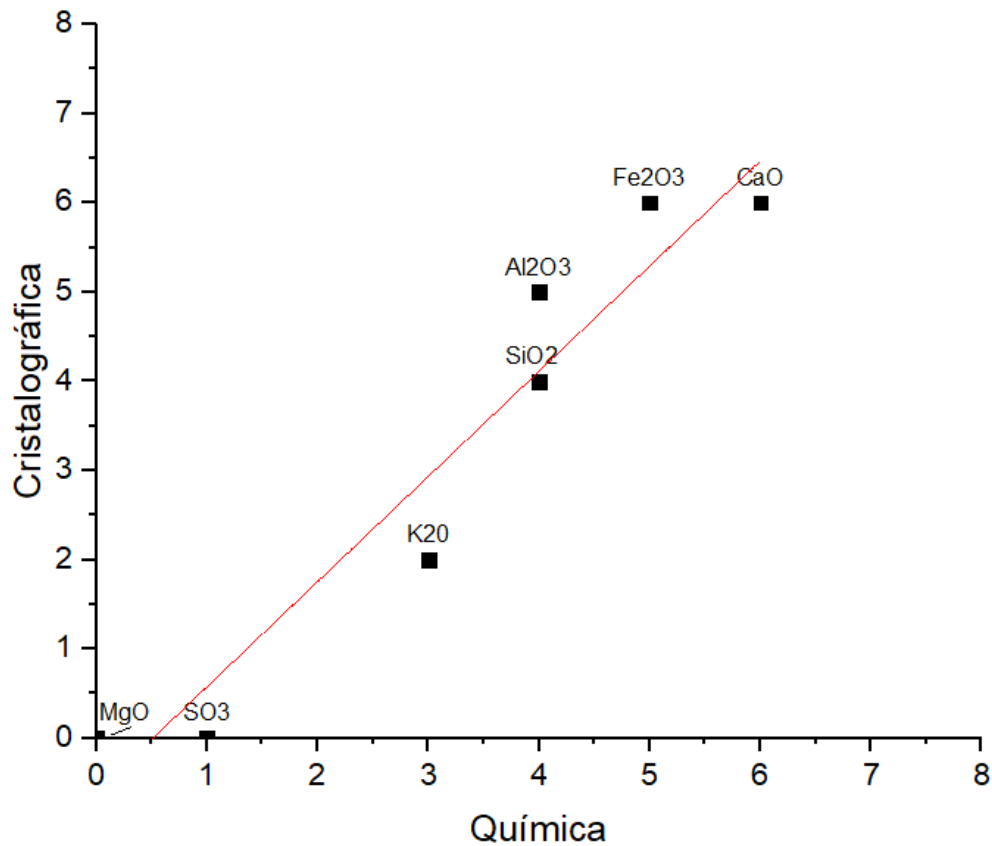


Fonte: Elaborado pelo autor (2020)

A análise dos resultados referente as instâncias das razões dos óxidos possibilitou inferir que há uma correlação significativa entre as melhores soluções obtidas. A Figura 6.2 destaca os resultados para a análise de correlação realizada, destacando a obtenção de 0,962 para o coeficiente de Pearson, que indica a presença de fortes indícios de similaridade entre as britas das melhores soluções encontradas nas instâncias dos dois conjuntos.

Por sua vez, observando o conjunto de instâncias das razões dos óxidos, apesar de pequenas variações entre as melhores soluções alcançadas, foi constatado que as frequências

Figura 6.2: Análise de Estatística de Correlação - Maiores Frequências dos Óxidos nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas



Fonte: Elaborado pelo autor (2020)

das dez razões mais presentes no conjunto de melhores soluções apresentam fortes indícios de similaridade, como ilustra a Figura 6.3.

Conforme pode ser observado, as maiores frequências nas melhores soluções obtidas de ambos os conjuntos avaliados foram registradas nas razões $\frac{\text{Fe}_2\text{O}_3}{\text{K}_2\text{O}}$, $\frac{\text{Al}_2\text{O}_3}{\text{SiO}_2}$, $\frac{\text{CaO}}{\text{Al}_2\text{O}_3}$ e $\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}$. Além disso, é possível perceber a forte presença dos óxidos Al_2O_3 , SiO_2 , CaO e Fe_2O_3 dentre as razões com maiores frequências, seja no numerador ou denominador. Os mesmos óxidos também foram constatados como mais presentes nas melhores soluções quando observados os resultados para as instâncias do conjunto de óxidos, demonstrando mais uma forte correlação existente entre as melhores clusterizações obtidas e os óxidos destacados.

De forma análoga a avaliação descritiva realizada nas melhores soluções do conjunto de instância dos óxidos, a Figura 6.4 a seguir apresenta a análise de correlação realizada para os resultados apresentados. Conforme pode ser observado, foi obtido um coeficiente de Pearson de aproximadamente 0,964, indicando que há uma similaridade significativa entre as melhores soluções obtidas para os conjuntos de soluções avaliados.

Figura 6.3: Comparação das Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas

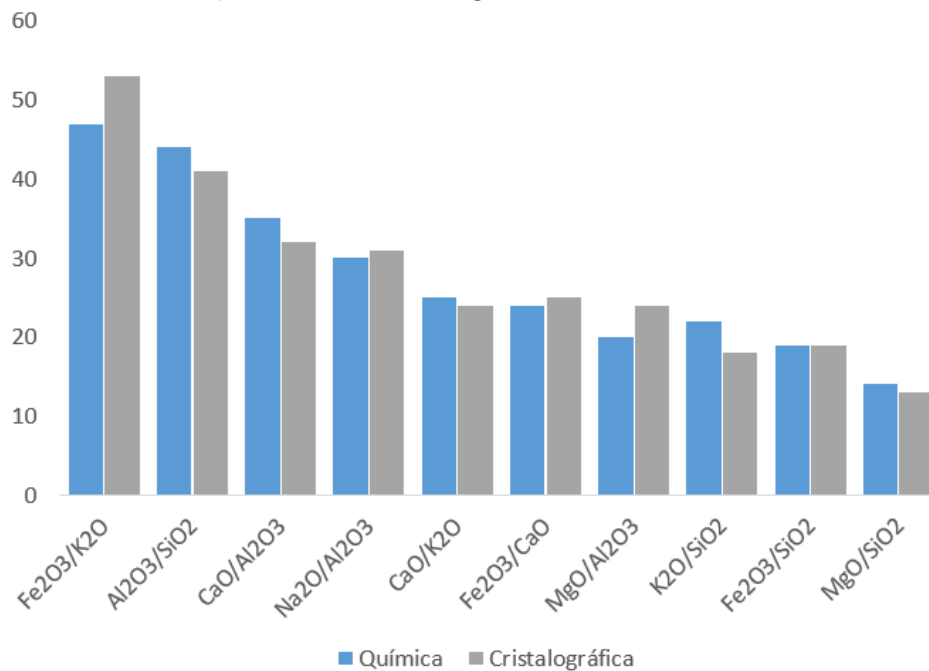
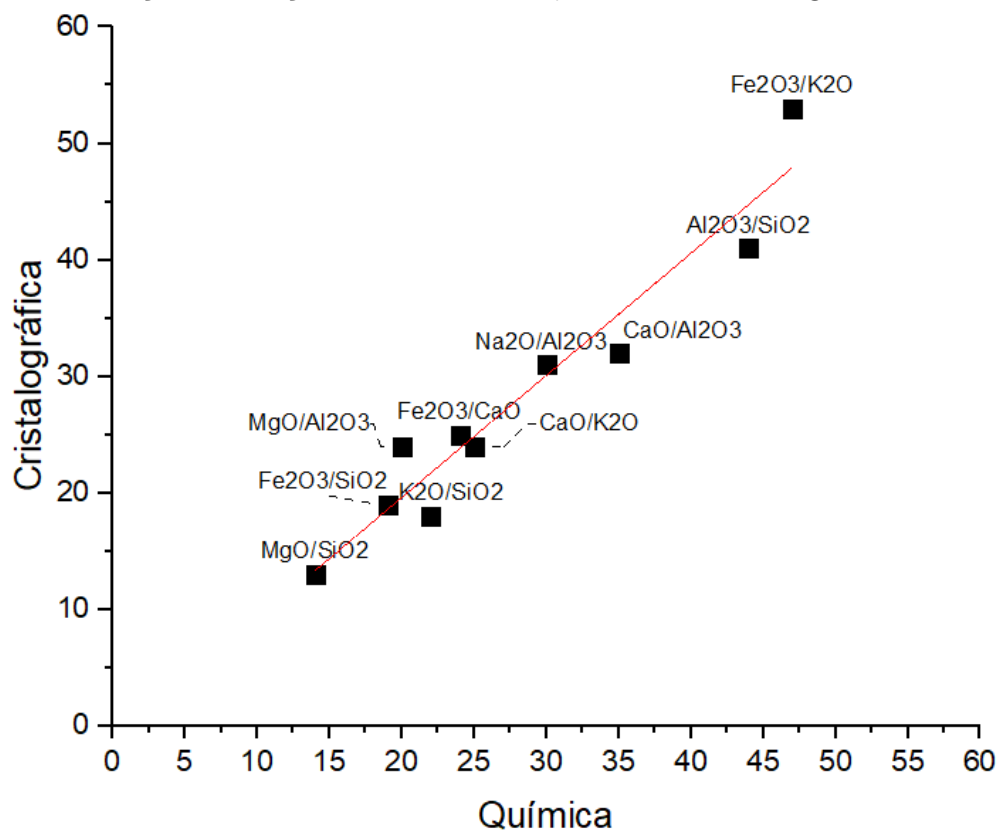
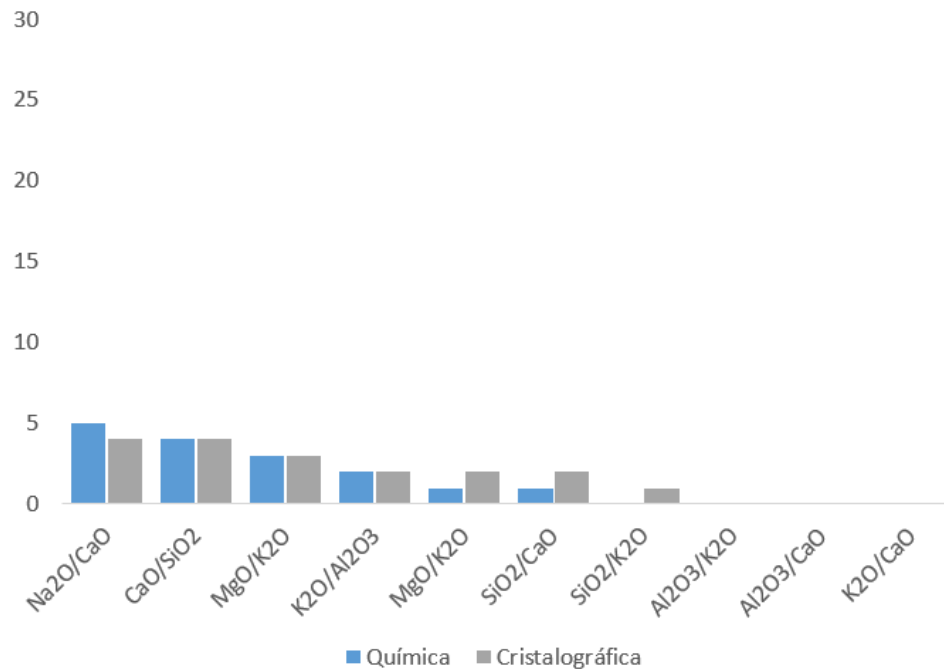


Figura 6.4: Análise de Estatística de Correlação - Maiores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas



Por fim, também foram realizadas análises referentes as razões com menores presenças nas instâncias onde as melhores soluções foram obtidas. Os resultados obtidos reforçam a existência de uma grande similaridade entre os resultados, de modo que há uma sinergia entre as razões constatadas com as menores frequências nas melhores soluções. A Figura 6.5 destaca o estudo.

Figura 6.5: Comparação das Menores Frequências das Razões nas 5% Melhores Soluções Alcançadas - Instâncias Químicas e Cristalográficas



6.2 Análise e Avaliação Mineralógica

Visando a obtenção de mais insights acerca dos melhores agrupamentos obtidos com o método proposto, foi realizado um estudo com o intuito de se investigar possíveis correlações entre os clusters ótimos formados durante o processo de clusterização, e suas composições mineralógicas. O objetivo principal é avaliar se, e de que forma, as presenças de determinados minerais nos agregados impactam e/ou justificam as clusterizações formadas ao longo dos processos de otimização.

Para uma melhor ilustração, a Tabela 6.1 apresenta um estrato referente as primeiras 40 britas (materiais), da melhor solução obtida após a execução do método proposto para a instância composta pelos óxidos Al_2O_3 , CaO e SiO_2 (advindas da análise cristalográfica). A ideia principal da abordagem de avaliação é estudar se a solução obtida, composta pela indicação da brita e seu respectivo cluster a qual pertence, possui correlações/justificativas

Devido ao grande volume de instâncias e dados do problema, para que tal análise fosse possível de ser realizada, foi necessário implementar, em linguagem de programação C++, um algoritmo de avaliação iterativo, responsável por relacionar e avaliar a melhor solução obtida para cada uma das instâncias com as quantificações mineralógicas das britas. Os resultados obtidos através das análises realizadas permitiram constatar que há relações significativas nos clusters formados através da clusterização química, com suas respectivas proporções mineralógicas originais. Em outras palavras, diversos dos clusters formados através dos atributos químicos (óxidos ou razões de óxidos), poderiam, também, serem justificadas através das características mineralógicas originais das britas.

O estudo permitiu demonstrar que as composições mineralógicas podem influenciar de maneira significativa as decisões do algoritmo de otimização no processo de clusterização, mesmo que indiretamente. Foi possível observar que tanto a presença de determinados minerais (e suas quantidades), como a completa ausência em alguns casos, estiveram diretamente relacionados com a formação dos clusters na imensa maioria das instâncias. Isto é, em 98,2% das instâncias objetos do processo de otimização, houve ao menos um dos clusters onde foi possível destacar uma tendência clara de correlação ligada as composições mineralógicas das britas.

A Tabela 6.2 destaca cinco instâncias nas quais as soluções foram tomadas como exemplo para demonstrar a análise de correlação realizada pelo algoritmo proposto, todas pertencentes ao grupo daquelas onde foram obtidas as melhores soluções em clusterização. Em seguida, as Tabelas de 6.3 até 6.7 ilustram com mais clareza as análises realizadas e alguns dos resultados e insights obtidos nas correlações das clusterizações. É importante notar que, devido o tamanho das instâncias conter muitas britas, foi realizado um recorte do que seria a tabela completa de cada instância, de modo a ser possível sua apresentação em uma página no presente estudo.

Tabela 6.2: Instâncias de Exemplificação do Estudo de Análise Mineralógica das Clusterizações Obtidas em Suas Soluções

Instâncias
Al_2O_3 , CaO e SiO_2
CaO , SiO_2 e Na_2O
Al_2O_3 , CaO e Fe_2O_3
Na_2O , Al_2O_3 e CaO
$\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}$, $\frac{\text{SiO}_2}{\text{CaO}}$ e $\frac{\text{K}_2\text{O}}{\text{SiO}_2}$

Tabela 6.4: Análise Mineralógica da Clusterização da Instância [CaO, SiO₂, e Na₂O]

RIETVELD - QUANTIFICAÇÃO MINERALÓGICA																		
Instância		Sílica	Mica			Feldspato								Outros				
Brita	Cluster	Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita
20	1	8,9	8,9	5,8				39,2		12,0		9,5		15,9				
55	1	15,4	6,6			12,9	2,9				25,7		6,6	30,0				
56	1	4,9	8,5			14,3					10,6	23,7	12,2	25,8				
65	1	15,8					40,7	30,1					4,3	9,2				
2	2																100,0	
43	2																15,3	84,7
53	2	1,6															93,4	5,0
37	3	1,3														98,1		0,6
42	3	2,8														97,2		
50	3	3,3														96,7		
0	4	100,0																
3	4	71,5				15,5				13,0								
66	4	82,3		12,0		5,3									0,4			
1	5	33,6		5,3		29,5				21,1				10,4				
6	5	38,4	2,7			26,3				32,6								
11	5	41,5	4,1			32,5				21,6				0,4				
12	5	37,4		12,3		41,3				4,3					4,8			
13	5	25,3	3,3			40,6				18,5				12,4				
18	5	38,9	2,7	8,8		4,4				21,6	7,5	6,9	9,3					
21	5	27,0	2,3			30,5				32,6			7,6					
22	5	17,5				67,4							12,5		2,6			
23	5	40,0	10,1	15,8		21,6				7,2				4,7	0,6			
24	5	29,0	5,4	5,9		32,9				21,0				5,9				
26	5	33,6	1,0	4,7		25,8				34,8								
27	5	34,8	9,8	9,2		30,3						14,6			1,2			
28	5	33,4	3,0	16,8		38,8				7,2					0,8			
32	5	28,2	2,8	4,2		31,0				33,8								
33	5	32,3	4,7			28,7				19,0		12,8		2,6				
34	5	29,1	0,8	2,4		27,3				40,4								
35	5	26,9	5,4	2,9		34,3				28,8				1,4	0,2			
36	5	37,9	7,6			23,0				8,6			16,8	6,1				
40	5	31,1	4,3	10,2		31,0				23,3					0,1			
41	5	35,0	4,1	3,5		21,2				36,2								
44	5	26,5	6,0			22,1				29,9			15,6					
49	5	39,2	4,2			10,8					6,0	23,3	16,6					
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
17	6	24,9	3,8			24,4				28,3	11,7			6,9				
19	6	15,1	6,6			12,0				18,8	10,9		21,9	14,7				
29	6	18,6	7,4	8,5		12,4				26,3	11,7			14,6	0,5			
30	6	22,2	7,5			28,6				29,0				11,7	1,1			
38	6	23,2	7,3		6,1	14,9			11,3				16,3	20,7	0,3			
39	6	16,9	8,3	3,6		24,3				9,1	4,3		12,8	20,7				
47	6	11,4	7,9			17,1						30,1	15,0	18,6				
54	6	19,4	3,7	1,4		10,7	40,4							24,4				
59	6	23,2	5,1			15,3					11,9	18,3	20,5	5,7				
60	6	20,7	11,9			16,0						16,2	20,9	14,3				
63	6	29,7	5,2	2,5		19,7	18,1			1,9				22,9				

Tabela 6.7: Análise Mineralógica da Clusterização da Instância $\left[\begin{smallmatrix} \text{Na}_2\text{O} \\ \text{Al}_2\text{O}_3 \end{smallmatrix}, \begin{smallmatrix} \text{SiO}_2 \\ \text{CaO} \end{smallmatrix}, \begin{smallmatrix} \text{K}_2\text{O} \\ \text{SiO}_2 \end{smallmatrix} \right]$

RIETVELD - QUANTIFICAÇÃO MINERALÓGICA																		
Instância		Sílica	Mica			Feldspato								Outros				
Brita	Cluster	Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita
6	0	38,4	2,7			26,3				32,6								
11	0	41,5	4,1			32,5				21,6				0,4				
12	0	37,4		12,3		41,3				4,3					4,8			
26	0	33,6	1,0	4,7		25,8				34,8								
27	0	34,8	9,8	9,2		30,3						14,6			1,2			
28	0	33,4	3,0	16,8		38,8				7,2					0,8			
32	0	28,2	2,8	4,2		31,0				33,8								
10	1	34,7	15,8			42,5											7,0	
16	1	28,2		14,1		16,1									10,7	16,9		14,1
19	1	15,1	6,6			12,0				18,8	10,9		21,9	14,7				
20	1	8,9	8,9	5,8				39,2		12,0		9,5		15,9				
25	1	27,4	11,0	11,6		20,5		10,6		17,7					1,2			
29	1	18,6	7,4	8,5		12,4				26,3	11,7			14,6	0,5			
37	1	1,3														98,1		0,6
38	1	23,2	7,3		6,1	14,9			11,3				16,3	20,7	0,3			
39	1	16,9	8,3	3,6		24,3				9,1	4,3		12,8	20,7				
42	1	2,8														97,2		
43	1																15,3	84,7
45	1	29,1	8,1			9,1					14,3		26,5	12,9				
46	1	14,4	12,4			25,1			10,7		15,0		13,8	8,6				
47	1	11,4	7,9			17,1						30,1	15,0	18,6				
48	1	19,1	7,7			19,6					19,0	31,3		3,3				
50	1	3,3														96,7		
53	1	1,6															93,4	5,0
54	1	19,4	3,7	1,4		10,7	40,4							24,4				
55	1	15,4	6,6			12,9	2,9				25,7		6,6	30,0				
56	1	4,9	8,5			14,3					10,6	23,7	12,2	25,8				
57	1	13,1	5,4	8,6		22,0			24,9		20,0			6,0				
62	1	23,0	7,7			15,6			14,8		16,1	13,6		9,3				
63	1	29,7	5,2	2,5		19,7	18,1			1,9				22,9				
65	1	15,8					40,7	30,1					4,3	9,2				
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
33	5	32,3	4,7			28,7				19,0		12,8		2,6				
35	5	26,9	5,4	2,9		34,3				28,8				1,4	0,2			
44	5	26,5	6,0			22,1				29,9			15,6					
51	5	30,7				37,8				21,9		7,6		2,0				
58	5	28,8	4,2	4,8		17,0				29,9			15,3					
61	5	31,6	3,4	5,8		18,8				29,8			10,7		0,0			
1	6	33,6		5,3		29,5				21,1				10,4				
4	6	24,9	3,6	8,4		9,0				25,2			29,0	0,1				
8	6	17,3	10,9			36,1				21,3				10,2	4,2			
13	6	25,3	3,3			40,6				18,5				12,4				
18	6	38,9	2,7	8,8		4,4				21,6	7,5	6,9	9,3					
24	6	29,0	5,4	5,9		32,9				21,0				5,9				
30	6	22,2	7,5			28,6				29,0				11,7	1,1			
31	6	22,8	8,9	7,7		29,8				26,4				4,5				
36	6	37,9	7,6			23,0				8,6			16,8	6,1				

Através das análises reportadas referentes às soluções obtidas para as cinco instâncias objetos de exemplificação do estudo, foi possível constatar que há elementos de correlação significativos entre as composições mineralógicas e os atributos químicos considerados nos processos de clusterização. Na análise mineralógica referente a solução da instância $[\text{Al}_2\text{O}_3, \text{CaO} \text{ e } \text{SiO}_2]$, por exemplo, foi constatado uma forte influência da Biotita, minerais do grupo Feldspato, Calcita, Calcita Mag. e Dolomita.

Tomando como base o Cluster 1 é possível constatar que todos os elementos não possuem Biotita em sua composição, assim como são os elementos que possuem o maior percentual de Quartzo. Por outro lado, observando o Cluster 6, nota-se que todas os agregados não possuem presença de nenhum mineral do grupo Feldspato e que são as únicas britas que possuem composições relacionadas a Calcita, Calcita Mag. e Dolomita.

Com relação a análise da solução obtida para a instância $[\text{CaO}, \text{SiO}_2, \text{ e } \text{Na}_2\text{O}]$, foi possível mapear ainda mais correlações relacionadas aos minerais presentes nas britas. Todas as britas que possuem Calcita Mag. em sua composição foram agrupadas exclusivamente no Cluster 2, assim como todas as que possuem Calcita foram agrupadas unicamente no Cluster 3. Assim como ocorrido na solução da instância anteriormente avaliada, o Cluster 4 foi responsável por clusterizar as três britas que possuem maior percentual de Quartzo, na medida em que não possuem composição alguma de Biotita. Por outro lado, dentre as demais britas que possuem composições significativas em Quartzo e Biotita, foi observado que as que pertencem ao Cluster 5 não possuem ou apresentam percentuais significativamente baixos de Hornblenda em relação aos demais clusters.

Análise da solução referente a instância $[\text{Al}_2\text{O}_3, \text{CaO} \text{ e } \text{Fe}_2\text{O}_3]$ demonstrou a geração de diversos insights. O Cluster 3, por exemplo, é formado contempla todas as britas que possuem Labradorita em sua composição mineralógica. Além disso, constatou que o Cluster 2 é formado pelas britas que possuem altos níveis de Quartzo e não possuem o mineral Biotita, enquanto os Clusters 1 e 6 são formados exclusivamente pelas britas que possuem Calcita Mag. e Calcita, respectivamente, em suas composições. Por fim, o Cluster 0 é formado pelas duas únicas britas que possuem os minerais Anortita e Hornblenda simultaneamente, demonstrando o potencial de clusterização de tais minerais.

Por sua vez, a clusterização da instância $[\text{Na}_2\text{O}, \text{Al}_2\text{O}_3 \text{ e } \text{CaO}]$ foi realizada em uma quantidade menor de clusters. Entretanto, ainda assim foram observadas correlações com as composições mineralógicas. Todas as britas que possuem, ao mesmo tempo, os minerais Albita e Biotita em suas composições, foram agrupados nos Clusters 4 e 5. Entre os dois clusters, é possível perceber com clareza que o fator primordial na diferenciação

foi o percentual de Quartzo em suas composições. As britas com percentuais mais leves, foram agrupadas no Cluster 5, enquanto as demais formaram o Cluster 4.

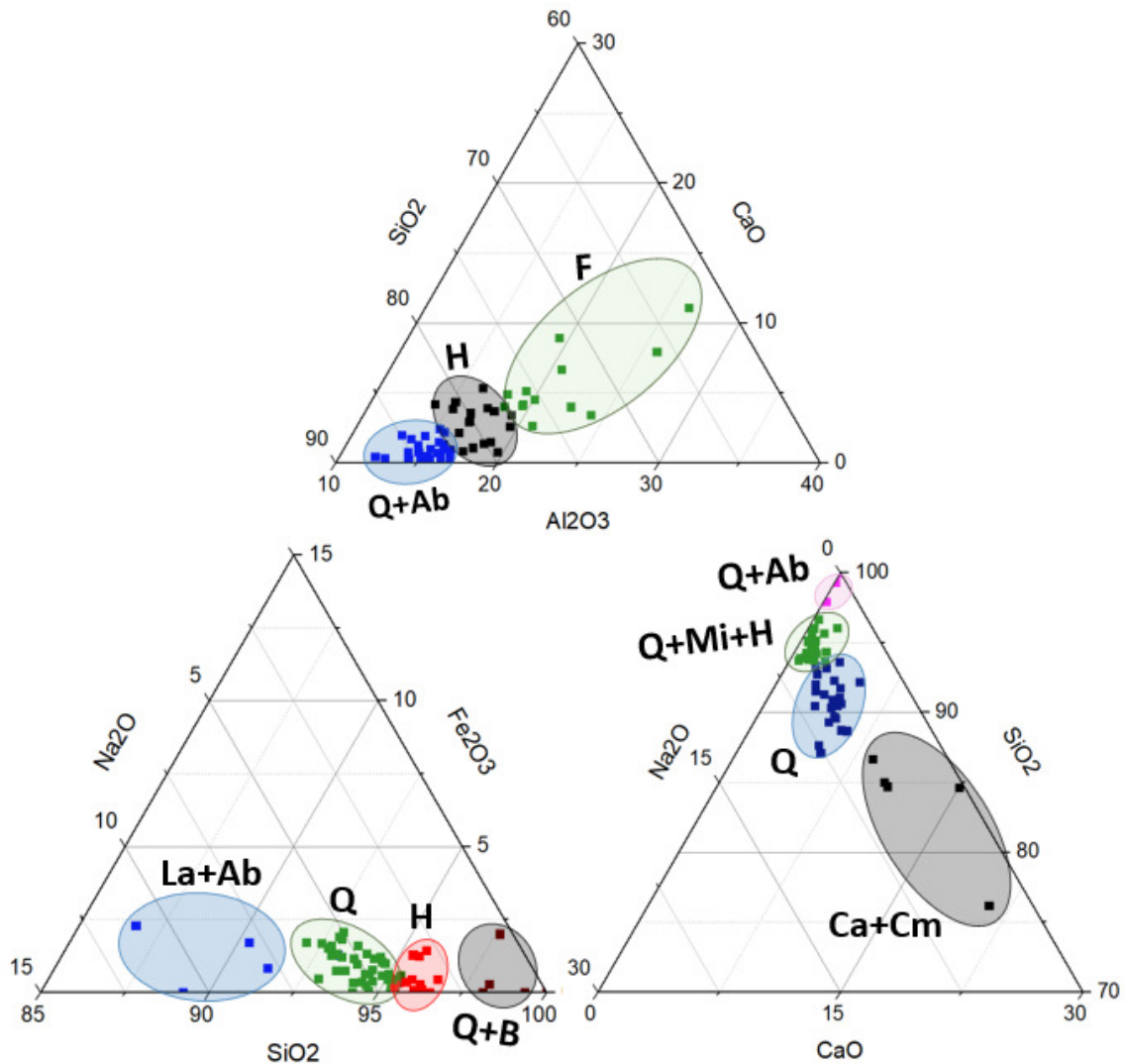
A título de exemplificação, foi também apresentada a clusterização obtida para uma instância do conjunto das razões de óxidos. A análise realizada para a solução da instância $\left[\frac{\text{Na}_2\text{O}}{\text{Al}_2\text{O}_3}, \frac{\text{SiO}_2}{\text{CaO}}, \frac{\text{K}_2\text{O}}{\text{SiO}_2}\right]$ demonstrou um insight bastante específico. Todas as britas que possuem algum percentual dos minerais Andesina, Anortita, Anortoclase, Calcita, Calcita Mag. ou Dolomita em suas composições, foram agrupados no Cluster 1. Isto é, todos os demais clusters foram formados por britas que não possuem os minerais relatados em suas composições mineralógicas.

No estudo implementado, foi levado em consideração um estrato referente as 5% instâncias onde foram alcançadas as melhores soluções de clusterização, assim como as 5% onde foram alcançados os piores desempenhos na qualidade na solução. Com isso, foi possível analisar as frequências onde os clusters que foram justificados pelas composições mineralógicas foram formados com o mesmo padrão.

Tal análise permitiu determinar os minerais (ou conjuntos de minerais) mais influenciadores no processo de clusterização dos atributos das britas, assim como detectar aqueles que, mesmo considerando instâncias com diversos atributos (óxidos e razões de óxidos) distintos, conseguiu persistir com a mesmo impacto nas decisões do algoritmo de clusterização.

Além disso, o estudo foi complementado com uma análise gráfica acerca das correlações existentes entre as clusterizações alcançadas com o método de otimização e as composições mineralógicas presentes nos clusters. A Figura 6.6 a seguir ilustra as clusterizações realizadas e os respectivos minerais (ou grupo de minerais) que justificam os agrupamentos realizados, conforme os resultados obtidos pelo estudo de correlação mineralógica realizado. Note que, devido o grande volume de instâncias, são demonstradas para as três instâncias onde foram obtidas melhores soluções no processo de otimização no conjunto de instâncias dos óxidos, relatadas na Seção 5.3.

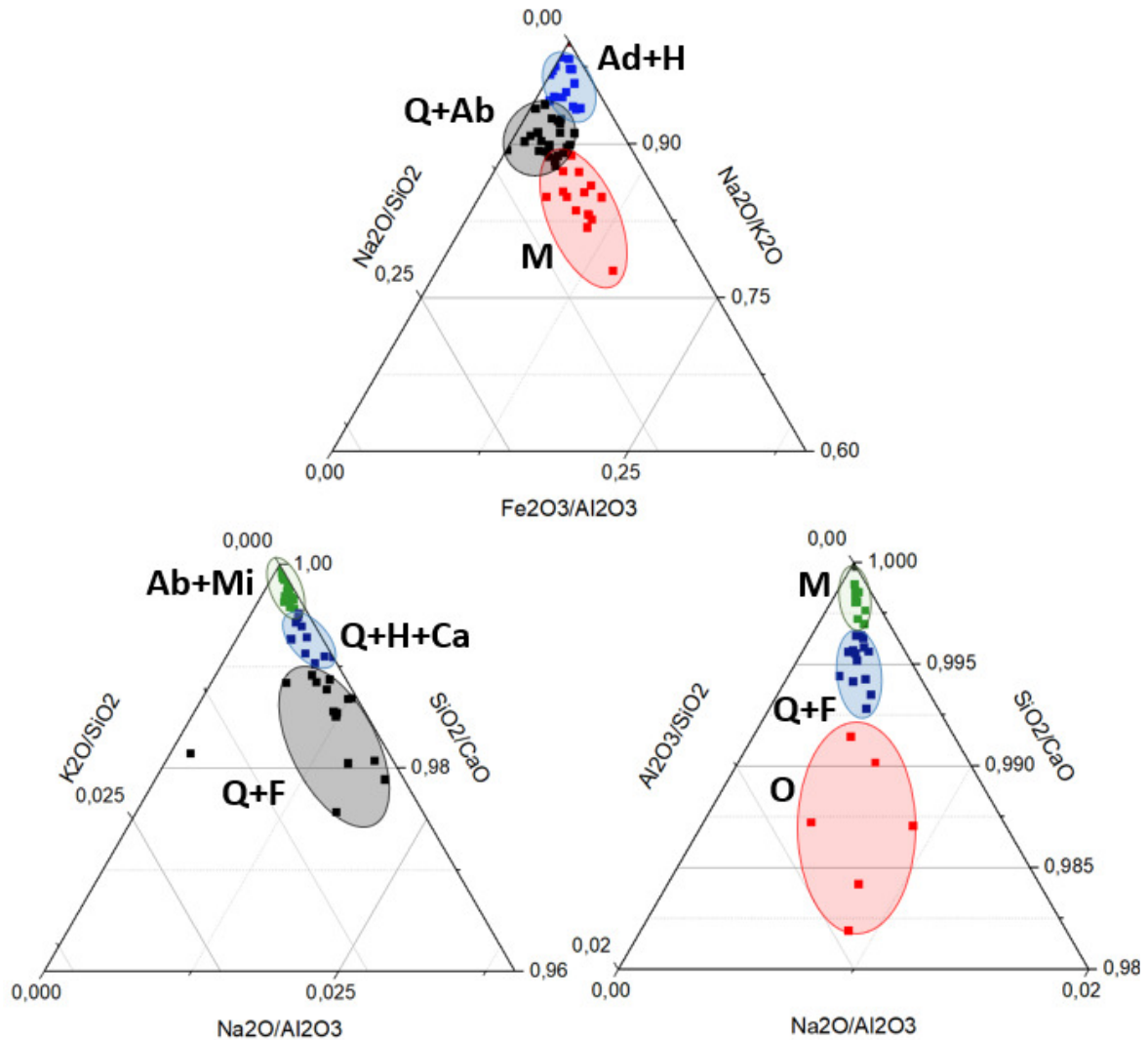
Figura 6.6: Análise Mineralógica das Clusterizações - Melhores Soluções - Óxidos



Fonte: Elaborado pelo autor (2020)

Conforme pode ser observado, é destacada a forte influência do Quartzo (Q), Mica (M), Feldspato (F), Outros (O), bem como, por vezes, os próprios minerais especificamente, como a Albita (Ab), Biotita (B), Calcita (Ca), Calcita Mag. (Cm), Hornblenda (H), Labradorita (La), Microclina (Mi), dentre outros. A correlação dos minerais e seus grupos com os cluters formados durante o processo de otimização demonstra que há uma correlação direta (tanto em aspectos químicos, como mineralógicos) entre as britas agrupadas. O mesmo mapeamento com os resultados do estudo feito foi realizado também para as instâncias do conjunto de instâncias referente as razões dos óxidos. A Figura 6.7 ilustra a análise para as instâncias onde foram obtidas as maiores qualidades de solução na clusterização.

Figura 6.7: Análise Mineralógica das Clusterizações - Melhores Soluções - Razões



Fonte: Elaborado pelo autor (2020)

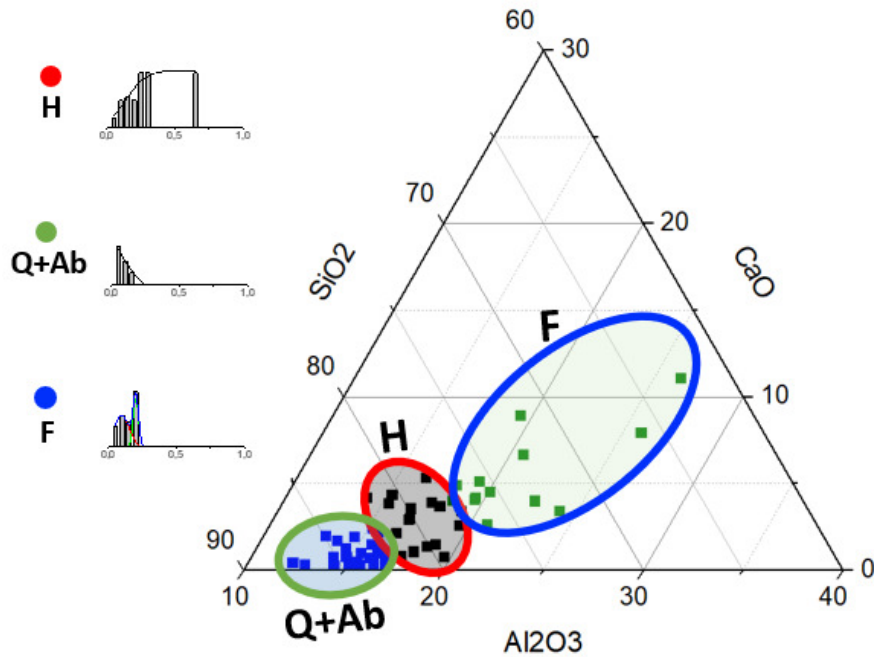
6.3 Análise de Expansão nas Clusterizações

Visando auferir a correlação entre as probabilidades de expansão das britas e as composições mineralógicas das britas, foi realizado um estudo analítico que levando em consideração os clusters obtidos pelo método de otimização para cada uma das melhores soluções das instâncias tratadas no problema. A análise permitiu determinar os minerais (e/ou grupos de minerais) que mais influenciam nas expansões, bem como determinar uma distribuição de probabilidade associada aos clusters.

A análise busca possibilitar a mensuração da probabilidade e intensidade de expansão de uma determinada brita qualquer, baseado, fundamentalmente, no cluster da qual faz parte ou em sua composição mineralógica. De modo a ilustrar os resultados obtidos com

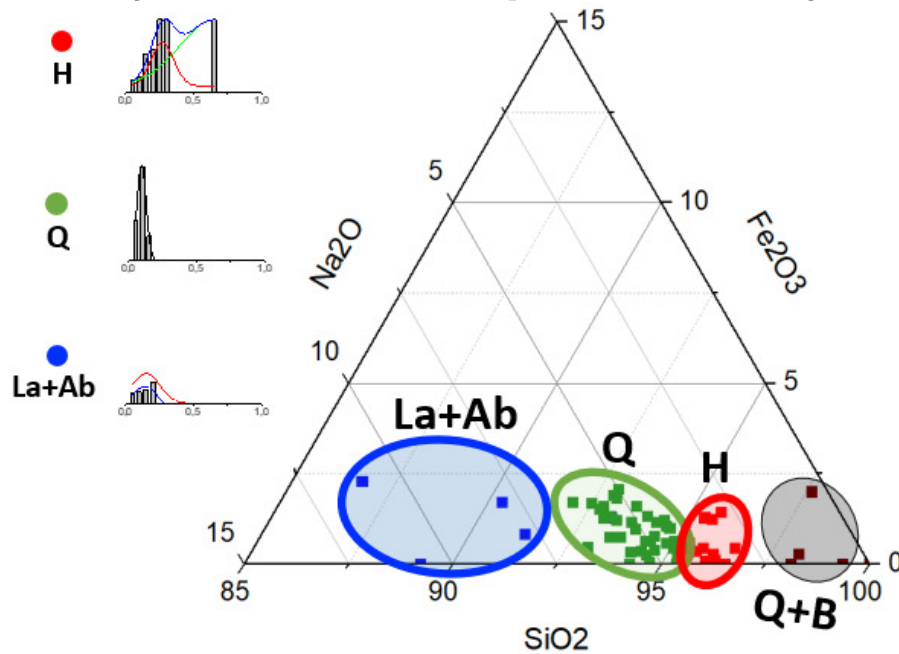
a análise, as Figuras de 6.8 a 6.13 demonstram os resultados alcançados através do estudo para um pequeno estrato das melhores britas obtidas por conjunto de instância da base de dados cristalográfica.

Figura 6.8: Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 1



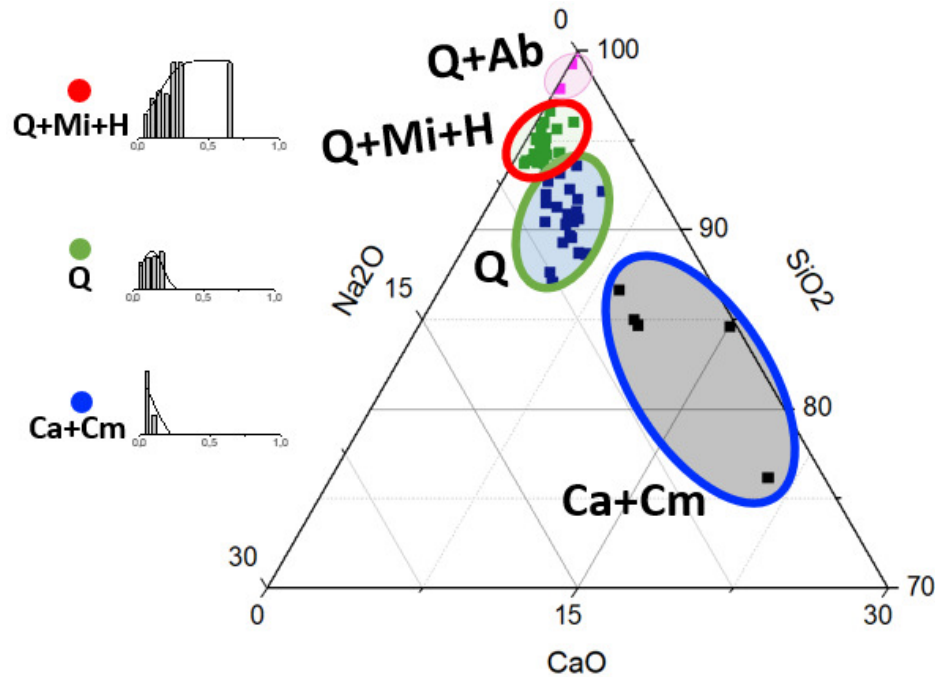
Fonte: Elaborado pelo autor (2020)

Figura 6.9: Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 2



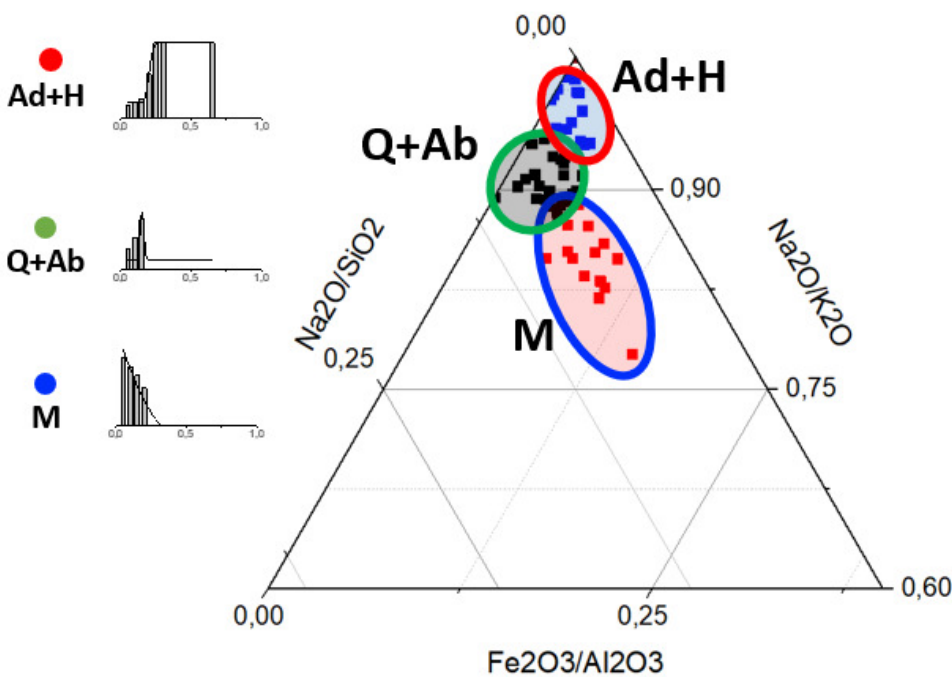
Fonte: Elaborado pelo autor (2020)

Figura 6.10: Avaliação da Probabilidade de Expansão das Clusterizações - Óxidos - 3



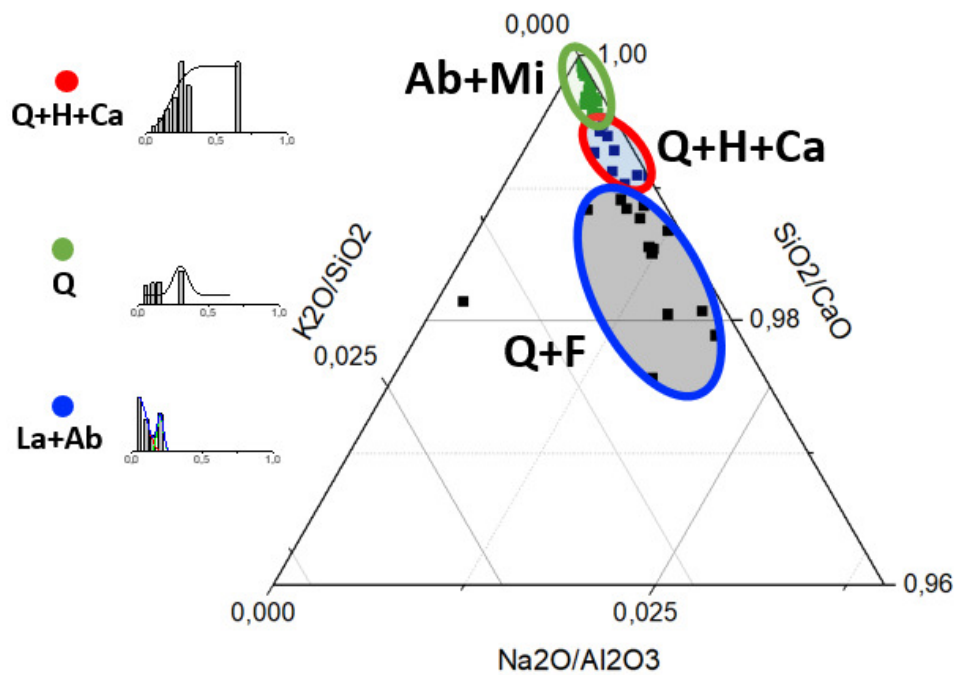
Fonte: Elaborado pelo autor (2020)

Figura 6.11: Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 1



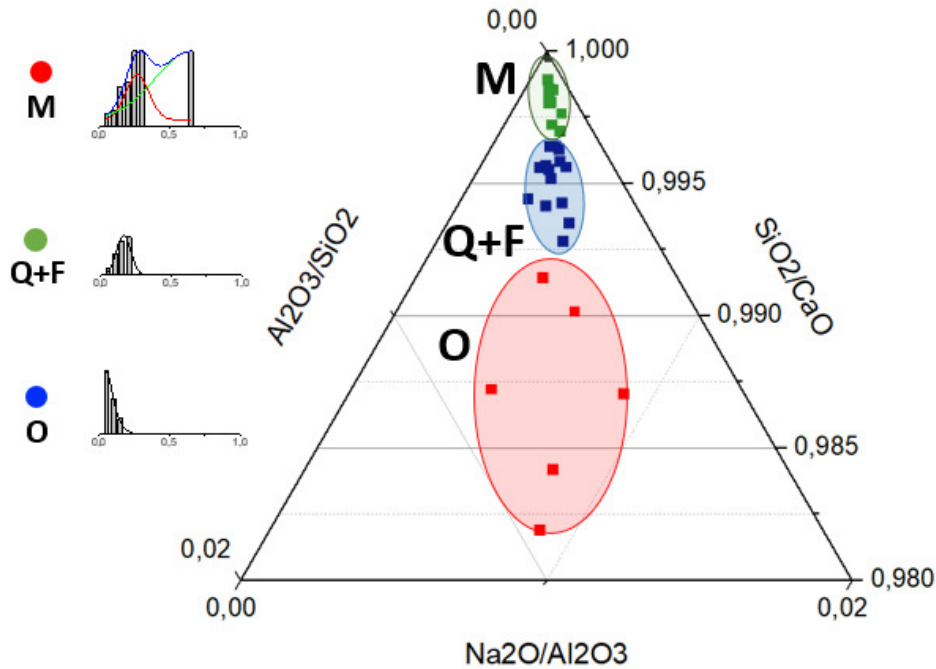
Fonte: Elaborado pelo autor (2020)

Figura 6.12: Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 2



Fonte: Elaborado pelo autor (2020)

Figura 6.13: Avaliação da Probabilidade de Expansão das Clusterizações - Razões - 3



Fonte: Elaborado pelo autor (2020)

Conforme demonstrado através do estudo realizado, foi possível observar que há uma série de minerais, bem como de grupos mineralógicos, que impactam diretamente no processo de clusterização. Não obstante, foi constatado que a probabilidade de expansão das britas possui uma associação significativa com os clusters definidos através do método de otimização. Tais aspectos demonstram com clareza que há uma influência significativa de determinadas composições mineralógicas nas possibilidades e condições de expansão das britas durante o processo produtivo.

Com os resultados obtidos através da análise de expansão nas clusterizações otimizadas, é possível atuar de maneira preditiva na tomada de decisão industrial, especialmente em problemáticas que dependem diretamente de avaliação e seleção de matérias-primas, onde a otimização de múltiplos critérios (incluindo suas restrições e objetivos) possui um impacto significativo no desempenho operacional e nos requisitos de qualidade desejados nos produtos finais.

Capítulo 7

Considerações Finais e Trabalhos Futuros

O presente trabalho propôs um método de clusterização com base na integração entre a metaheurística *Particle Swarm Optimization* e o algoritmo K-means que permitiu a determinação de agrupamento de componentes e seus respectivos atributos, de modo que atenderem a múltiplos critérios de otimização. Como principais critérios adotados no estudo, foram consideradas as composições químicas e mineralógicas de rochas utilizadas como agregados para concreto na região Nordeste do Brasil. Os resultados obtidos no trabalho demonstraram que os objetivos traçados foram atingidos, possibilitando alcançar sólidas contribuições para a literatura das temáticas abordadas, isto é, tanto para as áreas de engenharias, como para a área de ciência de dados e otimização.

Com relação as contribuições para as engenharias, os resultados do estudo podem proporcionar informações de impacto significativo para a tomada de decisão em ambientes industriais, assim como insights considerados inovadores, os quais envolvem: As melhores clusterizações obtidas pela aplicação do método de otimização nas instâncias do problema; Investigação e mapeamento das frequências de óxidos (e razões) nas melhores e piores soluções, as quais determinam quais os critérios mais e menos impactantes nas reações; Estudo de correlação mineralógica, envolvendo os clusters formados pelo método de otimização e as composições mineralógicas intracuster; Análise de distribuição de probabilidade de expansão nos clusters otimizados, possibilitando tanto uma avaliação preditiva sistêmica, como um mapeamento de correlações envolvendo as probabilidades de expansão em todos os clusters formados em todas as instâncias otimizadas.

Do ponto de vista da otimização e ciências dos dados, também foram alcançadas contribuições relevantes, tanto através da eficiência e inovação inerente a formulação e aplicação do método híbrido proposto, os quais demonstraram alcançar soluções de qualidade significativa, como pelos mecanismos e ferramentas propostas no desenvolvimento

do método. Desse modo, as principais contribuições no âmbito da otimização e ciência dos dados, foram: A proposição de uma metaheurística híbrida de clusterização automática em uma aplicação inovadora; Formulação matemática inédita de uma métrica de avaliação de desempenho eficiente para a problemática abordada; O desenvolvimento de um método de clusterização automática em abordagem híbrida envolvendo *Particle Swarm Optimization* e *K-Means*; A implementação de mecanismos eficientes no que tange as demais funções que envolvem a busca iterativa e de desempenho computacional do método proposto.

Além disso, é importante destacar a ausência de registros de estudos na literatura que abordem a implementação de métodos de otimização em problemas de clusterização em materiais da natureza abordada, a qual acentua o ineditismo do estudo realizado e suas possíveis contribuições para a literatura. Desse modo, diante dos resultados apresentados e do overview das principais contribuições relatadas, pode-se concluir, em relação aos objetivos específicos, que:

- A formulação matemática proposta para a métrica de avaliação da qualidade (desempenho) das soluções de clusterização no problema demonstraram, levando em consideração os critérios químicos e mineralógicos associados ao problema, retrata com eficiência o problema abordado, assim como foi fundamental para subsidiar uma escolha automática inteligente e efetiva pelo método de otimização proposto.
- O método de otimização desenvolvido, fundamentado em metaheurística híbrida evolucionária de *Particle Swarm Optimization* e *K-Means*, permitiu determinar os melhores atributos e características para a clusterização dos dados abordados no problema, baseado nas composições químicas e mineralógicas associadas três a três, com e sem repetição de óxidos e de suas razões.
- O método de otimização proposto permitiu, através dos mecanismos e ferramentas desenvolvidos na metaheurística, uma clusterização automática e eficiente, tanto da quantidade ideal de clusters, quanto dos melhores arranjos dos agregados presentes nas instâncias, para cada uma das configurações de três a três composições (óxidos ou razões).
- Dentre as principais variáveis químicas e mineralógicas, foi possível identificar suas frequências de ocorrência nas melhores soluções de clusterização alcançadas, permitindo identificar quais variáveis/características são mais representativas na formação dos agrupamentos;

- O análise envolvendo a correlação e avaliação mineralógica realizada permitiu demonstrar que há evidências de similaridades significativas entre os clusters formados pelo método de otimização e as composições mineralógicas dos materiais e, ainda, quais clusters podem ter seu processo de formação justificados pela ocorrência acentuada e/ou exclusiva de um determinado mineral (ou grupo de minerais).
- Com relação ao estudo envolvendo a expansão dos materiais, foram analisadas as correlações existentes entre os clusters formados em cada uma das instâncias e os coeficientes de expansão dos materiais. Desse modo, foi possível realizar um mapeamento da distribuição de probabilidade e intensidade de expansão dos materiais em cada cluster de cada instância dos conjuntos de instâncias.

Com isso, em virtude de haver uma correlação significativa entre os critérios químicos e mineralógicos juntamente com a existência de minerais que definem a formação dos clusters e a probabilidade de expansão por reação álcali-agregado, é possível, através de método de otimização em clusterização proposto, bem como dos estudos mineralógicos e de expansão realizados, estabelecer critérios múltiplos que permitem a escolha menos suscetível ao fenômeno deletério em questão. Dessa forma, a abordagem de otimização desenvolvida serve de ferramenta para subsidiar a tomada de decisão quanto a seleção de agregados para a produção de concretos mais duráveis, bem como para outras aplicações industriais, como com relação à logística associada à disponibilidade das jazidas e a localização de obras de infraestrutura, possibilitando uma redução de custos operacionais e logísticos significativos.

Referências

- ADINETZ, A. et al. Gpumafia: Efficient subspace clustering with mafia on gpus. In: SPRINGER. *European Conference on Parallel Processing*. [S.l.], 2013. p. 838–849.
- AGGARWAL, C. C.; YU, P. S. Redefining clustering for high-dimensional applications. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 14, n. 2, p. 210–225, 2002.
- AGRAWAL, R. et al. Automatic subspace clustering of high dimensional data for data mining applications. In: *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*. [S.l.: s.n.], 1998. p. 94–105.
- AIDONG, Z. *Advanced analysis of gene expression microarray data*. [S.l.]: World Scientific Publishing Company, 2006.
- ÁNGELES, G. et al. Crystal structure of low magnesium-content alite: Application to rietveld quantitative phase analysis. *Cement and Concrete Research*, Elsevier, v. 38, n. 11, p. 1261–1269, 2008.
- ANKERST, M. et al. Optics: ordering points to identify the clustering structure. *ACM Sigmod record*, ACM New York, NY, USA, v. 28, n. 2, p. 49–60, 1999.
- BAKER, F. B.; HUBERT, L. J. Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 70, n. 349, p. 31–38, 1975.
- BENABDELLAH, A. C.; BENGHABRIT, A.; BOUHADDOU, I. A survey of clustering algorithms for an industrial context. *Procedia Computer Science*, Elsevier B.V., v. 148, p. 291–302, 2019. ISSN 18770509. Disponível em: <<https://doi.org/10.1016/j.procs.2019.01.022>>.
- BIRGIN, E. G.; MARTINEZ, J.; RONCONI, D. P. Optimizing the packing of cylinders into a rectangular container: A nonlinear approach. *European Journal of Operational Research*, Elsevier, v. 160, n. 1, p. 19–33, 2005.
- BLEKAS, K.; LAGARIS, I. E. Newtonian clustering: An approach based on molecular dynamics and global optimization. *Pattern Recognition*, Elsevier, v. 40, n. 6, p. 1734–1744, 2007.
- CALIŃSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, Taylor & Francis, v. 3, n. 1, p. 1–27, 1974.
- CHOU, C.-H.; SU, M.-C.; LAI, E. A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications*, Springer, v. 7, n. 2, p. 205–220, 2004.

- DAS, S.; ABRAHAM, A.; KONAR, A. Automatic kernel clustering with a Multi-Elitist Particle Swarm Optimization Algorithm. *Pattern Recognition Letters*, v. 29, n. 5, p. 688–699, 2008. ISSN 01678655.
- DAVENPORT, T. H.; PRUSAK, L. Working knowledge: How organizations manage what they know. *Ubiquity*, ACM, v. 2000, n. August, p. 6, 2000.
- DENG, G.-F.; LIN, W.-T.; LO, C.-C. Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization. *Expert Systems with Applications*, Elsevier, v. 39, n. 4, p. 4558–4566, 2012.
- DIMITRIADOU, E.; DOLNIČAR, S.; WEINGESSEL, A. An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, Springer, v. 67, n. 1, p. 137–159, 2002.
- EBERHART, R. C.; SHI, Y. Comparing inertia weights and constriction factors in particle swarm optimization. In: IEEE. *Proceedings of the 2000 congress on evolutionary computation. CEC00 (Cat. No. 00TH8512)*. [S.l.], 2000. v. 1, p. 84–88.
- ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*. [S.l.: s.n.], 1996. v. 96, n. 34, p. 226–231.
- FAHAD, A. et al. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, IEEE, v. 2, n. 3, p. 267–279, 2014.
- FERNANDES, W. *Effect of minor Constituents in Alite Polymorphs of Industrial Clinker Using Bond Valence Theory*. Tese (Doutorado) — Federal University of Paraíba, 2019.
- GAN, M. et al. Predicting packing characteristics of particles of arbitrary shapes. *KONA Powder and Particle Journal*, Hosokawa Powder Technology Foundation, v. 22, p. 82–93, 2004.
- GARZA-FABRE, M.; HANDL, J.; KNOWLES, J. An improved and more scalable evolutionary approach to multiobjective clustering. *IEEE Transactions on Evolutionary Computation*, IEEE, v. 22, n. 4, p. 515–535, 2017.
- GOMEZ, J.; DASGUPTA, D.; NASRAOUI, O. A new gravitational clustering algorithm. In: SIAM. *Proceedings of the 2003 SIAM International Conference on Data Mining*. [S.l.], 2003. p. 83–94.
- GONZALEZ, T. F. On the computational complexity of clustering and related problems. In: *System modeling and optimization*. [S.l.]: Springer, 1982. p. 174–182.
- GUANGYOU, Y. A modified particle swarm optimizer algorithm. In: IEEE. *2007 8th International Conference on Electronic Measurement and Instruments*. [S.l.], 2007. p. 2–675.
- GUHA, S.; RASTOGI, R.; SHIM, K. Cure: an efficient clustering algorithm for large databases. *ACM Sigmod record*, ACM New York, NY, USA, v. 27, n. 2, p. 73–84, 1998.

- HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, v. 17, n. 3, p. 107–145, 2011. ISSN 00138703.
- HALKIDI, M.; VAZIRGIANNIS, M.; BATISTAKIS, Y. Quality scheme assessment in the clustering process. In: SPRINGER. *European Conference on Principles of Data Mining and Knowledge Discovery*. [S.l.], 2000. p. 265–276.
- HIFI, M.; YOUSEF, L. A hybrid algorithm for packing identical spheres into a container. *Expert Systems with Applications*, 2018.
- JAEGER, H. M.; NAGEL, S. R.; BEHRINGER, R. P. Granular solids, liquids and gases. *Reviews of Modern Physics*, Springer, v. 68, n. 4, p. 1259, 1996.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM computing surveys (CSUR)*, Acm New York, NY, USA, v. 31, n. 3, p. 264–323, 1999.
- JASKOWIAK, P. A. et al. On strategies for building effective ensembles of relative clustering validity criteria. *Knowledge and Information Systems*, Springer London, v. 47, n. 2, p. 329–354, 2016. ISSN 02193116.
- JOHNSON, S. C. Hierarchical clustering schemes. *Psychometrika*, Springer, v. 32, n. 3, p. 241–254, 1967.
- JUNLIN, L.; HONGGUANG, F. Molecular dynamics-like data clustering approach. *Pattern Recognition*, Elsevier, v. 44, n. 8, p. 1721–1737, 2011.
- KARYPIS, G.; HAN, E.-H.; KUMAR, V. Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, IEEE, v. 32, n. 8, p. 68–75, 1999.
- KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. [S.l.]: John Wiley & Sons, 2009.
- KUO, R. J.; ZULVIA, F. E. Automatic Clustering Using an Improved Particle Swarm Optimization. *Journal of Industrial and Intelligent Information*, v. 1, n. 1, p. 46–51, 2013. ISSN 23013745.
- MANZANO, H. et al. Impact of chemical impurities on the crystalline cement clinker phases determined by atomistic simulations. *Crystal Growth & Design*, ACS Publications, v. 11, n. 7, p. 2964–2972, 2011.
- MISHRA, S. K.; PANDA, G.; MAJHI, R. A comparative performance assessment of a set of multiobjective algorithms for constrained portfolio assets selection. *Swarm and Evolutionary Computation*, Elsevier, v. 16, p. 38–51, 2014.
- NG, A. Y.; JORDAN, M. I.; WEISS, Y. On spectral clustering: Analysis and an algorithm. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2002. p. 849–856.
- NGUYEN, H. L.; WOON, Y. K.; NG, W. K. A survey on data stream clustering and classification. *Knowledge and Information Systems*, v. 45, n. 3, p. 535–569, 2015. ISSN 02193116.

- OLIVEIRA, T. B. S. de. *Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada*. 95 p. Tese (Doutorado) — Universidade de São Paulo, 2008.
- PARK, H.-S.; JUN, C.-H. A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, Elsevier, v. 36, n. 2, p. 3336–3341, 2009.
- REN, X.; ZHANG, W.; YE, J. Ftir study on the polymorphic structure of tricalcium silicate. *Cement and Concrete Research*, Elsevier, v. 99, p. 129–136, 2017.
- SAJANA, T.; RANI, C. M. S.; NARAYANA, K. V. A survey on clustering techniques for big data mining. *Indian Journal of Science and Technology*, v. 9, n. 3, p. 1–12, 2016. ISSN 09745645.
- SANCHEZ, M. A. et al. Fuzzy granular gravitational clustering algorithm for multivariate data. *Information Sciences*, Elsevier, v. 279, p. 498–511, 2014.
- SAXENA, A. et al. A review of clustering techniques and developments. *Neurocomputing*, Elsevier, v. 267, p. 664–681, 2017.
- SCHIKUTA, E.; ERHART, M. The bang-clustering system: Grid-based data analysis. In: SPRINGER. *International Symposium on Intelligent Data Analysis*. [S.l.], 1997. p. 513–524.
- SHI, J.; MALIK, J. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, Ieee, v. 22, n. 8, p. 888–905, 2000.
- SHI, Y.; EBERHART, R. A modified particle swarm optimizer. In: IEEE. *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)*. [S.l.], 1998. p. 69–73.
- SILVA, J. de A.; HRUSCHKA, E. R.; GAMA, J. An evolutionary algorithm for clustering data streams with a variable number of clusters. *Expert Systems with Applications*, Elsevier, v. 67, p. 228–238, 2017.
- SILVA, Y. L. T.; HERTHEL, A. B.; SUBRAMANIAN, A. A multi-objective evolutionary algorithm for a class of mean-variance portfolio selection problems. *Expert Systems with Applications*, Elsevier, v. 133, p. 225–241, 2019.
- SOBOLEV, K. The development of a new method for the proportioning of high-performance concrete mixtures. *Cement and Concrete Composites*, v. 26, n. 7, p. 901–907, 2004. ISSN 09589465.
- SOBOLEV, K.; AMIRJANOV, A. Application of genetic algorithm for modeling of dense packing of concrete aggregates. *Construction and Building Materials*, 2010. ISSN 09500618.
- SOUSA, G. *Estudo De Algoritmos Híbridos Para Clusterização De Dados Usando PSO*. Tese (Doutorado) — Universidade Federal de Lavras, 2010. Disponível em: <<http://www.bcc.ufla.br/monografias/2010/GustavoHenrique.pdf>>.
- STOYAN, Y.; CHUGAY, A. Packing cylinders and rectangular parallelepipeds with distances between them into a given region. *European Journal of Operational Research*, Elsevier, v. 197, n. 2, p. 446–455, 2009.

- WANG, J. Packing of unequal spheres and automated radiosurgical treatment planning. *Journal of Combinatorial Optimization*, Springer, v. 3, n. 4, p. 453–463, 1999.
- WANG, W. et al. Sting: A statistical information grid approach to spatial data mining. In: *VLDB*. [S.l.: s.n.], 1997. v. 97, p. 186–195.
- WEBBER, W.; MOFFAT, A.; ZOBEL, J. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)*, ACM New York, NY, USA, v. 28, n. 4, p. 1–38, 2010.
- WILLIAMS, S.; PHILIPSE, A. Random packings of spheres and spherocylinders simulated by mechanical contraction. *Physical Review E*, APS, v. 67, n. 5, p. 051301, 2003.
- WONG, K. C. A short survey on data clustering algorithms. *Proceedings - 2015 2nd International Conference on Soft Computing and Machine Intelligence, ISCMi 2015*, p. 64–68, 2016.
- WRIGHT, W. E. Gravitational clustering. *Pattern recognition*, Elsevier, v. 9, n. 3, p. 151–166, 1977.
- XING, G. et al. Integrated coverage and connectivity configuration for energy conservation in sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, ACM New York, NY, USA, v. 1, n. 1, p. 36–72, 2005.
- XIONG, H.; WU, J.; CHEN, J. K-means clustering versus validation measures: a data-distribution perspective. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, IEEE, v. 39, n. 2, p. 318–331, 2008.
- ZHANG, T.; RAMAKRISHNAN, R.; LIVNY, M. Birch: an efficient data clustering method for very large databases. *ACM Sigmod Record*, ACM New York, NY, USA, v. 25, n. 2, p. 103–114, 1996.

APÊNDICE A - Banco de Dados das Britas - Análise Química - Florescência de Raios X

Tabela A.1: Composição Química das Britas - Florescência de Raios X

Brita	Composição Química							
	Al ₂ O ₃	CaO	K ₂ O	SiO ₂	Fe ₂ O ₃	MgO	Na ₂ O	SO ₃
1	0,49	0,30	0,04	97,64	0,97	0,13	0,08	0,14
2	15,00	5,18	4,96	62,24	5,39	1,65	2,82	0,15
3	3,45	82,01	0,67	10,14	2,23	1,00	0,00	0,15
4	4,37	0,73	1,87	90,64	0,78	0,12	1,01	0,07
5	14,45	2,92	5,17	70,61	2,30	0,64	2,72	0,06
6	21,76	1,35	6,07	59,63	2,31	0,08	7,67	0,09
7	14,50	1,75	5,75	71,97	2,05	0,38	2,78	0,07
8	14,33	2,13	5,62	70,70	2,28	0,76	3,17	0,06
9	14,40	4,19	4,14	66,37	4,80	1,02	3,34	0,12
10	15,26	4,13	5,07	65,56	4,47	1,01	2,96	0,14
11	13,60	2,42	5,22	69,50	4,60	0,56	2,62	0,16
12	15,02	2,77	5,89	67,22	4,21	0,67	2,57	0,11
13	14,56	2,13	3,50	70,47	3,52	1,71	3,05	0,03
14	14,30	2,11	5,71	70,67	2,26	0,79	3,13	0,06
15	15,42	3,72	4,34	66,13	3,17	1,62	4,04	0,05
16	15,22	3,11	5,98	66,59	3,62	0,74	2,82	0,10
17	13,90	19,73	2,61	52,53	4,92	3,21	1,57	0,27
18	9,86	40,19	1,09	35,40	5,96	3,95	0,67	0,39
19	15,33	0,66	6,79	73,38	1,00	0,82	1,65	0,00
20	17,04	4,13	4,22	64,52	3,90	1,34	3,04	0,12
21	16,68	7,36	2,27	59,50	6,26	3,18	2,37	0,15
22	18,01	1,41	5,95	68,68	1,61	0,42	2,87	0,30
23	18,08	4,20	0,29	54,22	10,74	3,89	3,66	0,04
24	18,14	2,56	3,76	63,70	5,80	2,30	2,27	0,20
25	15,70	1,97	4,93	70,87	2,09	0,58	2,94	0,13
26	16,58	3,18	4,56	68,95	2,28	0,60	2,63	0,08
27	15,69	1,42	6,81	71,48	1,16	0,07	2,55	0,07
28	18,20	1,41	5,39	66,15	3,98	1,24	2,43	0,05
29	18,81	3,33	4,42	66,07	2,76	0,92	2,63	0,03
30	16,11	4,84	4,92	60,85	6,12	3,17	2,10	0,04

Tabela A.2: Composição Química das Britas - Método de Florescência de Raios X

...	...	...	...	...	...	...	...	...
31	16,40	3,69	5,17	63,23	4,93	2,36	2,43	0,05
32	16,87	2,53	6,27	66,01	3,30	0,78	2,82	0,04
33	17,85	2,28	5,45	67,15	2,52	0,62	2,74	0,06
34	16,75	1,95	5,74	70,02	1,90	0,42	2,60	0,06
35	17,89	1,49	6,74	68,40	1,92	0,34	2,45	0,03
36	17,43	2,25	6,25	65,48	4,13	0,64	2,38	0,04
37	15,78	3,71	3,29	67,03	4,90	1,38	2,51	0,11
38	0,94	94,07	0,19	2,51	0,60	11,97	0,00	0,24
39	13,93	7,17	2,61	52,63	12,09	3,31	3,03	0,41
40	15,44	5,81	3,65	58,22	8,88	2,36	2,87	0,08
41	15,08	2,62	5,25	67,91	3,90	0,75	3,17	0,07
42	14,49	1,83	6,31	68,81	4,13	0,34	2,78	0,06
43	0,57	95,39	0,08	1,75	0,47	0,00	0,19	0,10
44	0,52	71,24	0,03	2,61	0,67	24,57	0,00	0,07
45	15,25	3,09	4,90	66,24	4,36	1,05	3,05	0,06
46	14,48	5,95	2,52	57,06	11,51	2,39	1,96	0,42
47	13,22	6,35	3,30	54,23	14,23	2,28	3,30	0,33
48	13,13	5,06	3,33	59,37	10,79	2,91	3,33	0,17
49	14,35	3,69	3,67	64,52	8,23	1,24	3,67	0,19
50	12,01	1,99	4,38	71,91	4,19	0,68	3,57	0,07
51	0,26	95,02	0,03	2,04	0,65	0,00	0,00	0,08
52	14,65	1,66	6,17	70,45	2,40	0,23	3,48	0,08
53	15,42	1,61	5,67	71,15	1,92	0,21	3,33	0,06
54	0,85	93,62	0,16	2,17	0,72	1,97	0,00	0,23
55	10,96	6,34	0,45	59,61	11,44	3,73	5,74	0,12
56	16,64	6,37	0,88	59,20	8,87	2,42	3,71	0,16
57	16,13	5,01	4,40	58,62	7,03	2,76	3,21	0,10
58	11,87	4,62	2,20	60,91	9,81	3,57	4,76	0,00
59	12,06	1,53	4,84	73,57	2,79	0,52	3,64	0,07
60	12,28	4,02	2,85	66,88	6,25	1,26	3,98	0,12
61	12,50	4,71	3,20	64,23	7,16	2,56	3,55	0,09
62	15,98	0,31	3,08	57,23	16,55	0,73	2,99	0,00
63	12,58	4,18	3,19	65,29	6,73	1,85	3,82	0,13
64	12,33	5,57	0,67	66,96	6,65	2,22	4,33	0,15
65	13,08	3,86	3,22	65,85	6,44	1,10	4,25	0,12
66	18,61	2,77	0,11	41,09	28,69	1,66	3,53	0,00
67	7,88	0,21	0,19	73,95	14,65	0,43	0,45	0,00
68	17,98	0,56	5,65	65,70	3,72	0,94	4,46	0,00
69	19,23	0,90	2,40	70,52	0,79	0,21	5,56	0,05
70	11,22	2,19	3,76	74,77	1,56	0,43	5,34	0,06

Tabela A.3: Composição Química das Britas - Método de Florescência de Raios X

...	...	...	...	...	...	...	...	...
71	10,20	1,96	2,48	77,63	1,74	0,76	4,71	0,06
72	10,90	3,05	2,27	74,03	2,84	1,28	5,01	0,07
73	11,20	3,57	3,22	71,75	3,66	1,40	4,20	0,10
74	11,13	3,83	2,89	71,13	3,80	1,64	3,82	0,07
75	10,73	1,68	2,55	77,47	1,37	0,44	5,27	0,03
76	10,76	2,06	3,50	75,40	2,31	0,80	4,36	0,06
77	11,35	2,07	2,62	74,84	2,17	1,20	5,01	0,05
78	11,72	4,69	2,43	65,91	6,89	3,51	3,27	0,12
79	11,14	2,67	4,05	69,30	5,93	1,79	3,30	0,07
80	11,29	1,87	3,47	75,11	1,79	0,90	4,73	0,14
81	11,83	3,72	3,01	68,61	5,19	2,59	3,64	0,09
82	11,55	1,76	4,56	73,94	2,99	0,71	3,47	0,00
83	11,87	2,07	1,49	77,17	2,68	0,73	2,96	0,00
84	11,46	4,08	3,14	68,96	4,97	1,51	3,76	0,08

APÊNDICE B - Composição Mineralógica das Britas

Tabela B.1: Composição Mineralógica de Extrato do Banco de Dados

RIETVELD - QUANTIFICAÇÃO MINERALÓGICA (%)																	
	SÍL.	MICA			FELDSPATO								OUTROS				
Amostra	Quartzo	Biotita	Muscovita	Flogopita	Albita	Andesina	Anortita	Anortoclase	Microclina	Labradorita	Ortoclase	Oligoclase	Hornblenda	Caulinita	Calcita	Calcita Mag.	Dolomita
1	38,4	2,7	0,0	0,0	26,3	0,0	0,0	0,0	32,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
2	18,4	4,3	0,0	0,0	30,5	0,0	0,0	0,0	43,9	0,0	0,0	0,0	2,9	0,0	0,0	0,0	0,0
3	17,3	10,9	0,0	0,0	36,1	0,0	0,0	0,0	21,3	0,0	0,0	0,0	10,2	4,2	0,0	0,0	0,0
4	1,6	19,4	0,0	0,0	61,5	0,0	0,0	0,0	0,0	17,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	34,7	15,8	0,0	0,0	42,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	7,0	0,0
6	41,5	4,1	0,0	0,0	32,5	0,0	0,0	0,0	21,6	0,0	0,0	0,0	0,4	0,0	0,0	0,0	0,0
7	37,4	0,0	12,3	0,0	41,3	0,0	0,0	0,0	4,3	0,0	0,0	0,0	0,0	4,8	0,0	0,0	0,0
8	25,3	3,3	0,0	0,0	40,6	0,0	0,0	0,0	18,5	0,0	0,0	0,0	12,4	0,0	0,0	0,0	0,0
9	11,6	0,0	10,0	0,0	41,6	0,0	0,0	0,0	23,3	0,0	0,0	0,0	13,6	0,0	0,0	0,0	0,0
10	16,7	0,0	5,6	8,8	20,6	0,0	0,0	0,0	37,6	0,0	0,0	10,7	0,0	0,0	0,0	0,0	0,0
11	28,2	0,0	14,1	0,0	16,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	10,7	16,9	0,0	14,1
12	24,9	3,8	0,0	0,0	24,4	0,0	0,0	0,0	28,3	11,7	0,0	0,0	6,9	0,0	0,0	0,0	0,0
13	38,9	2,7	8,8	0,0	4,4	0,0	0,0	0,0	21,6	7,5	6,9	9,3	0,0	0,0	0,0	0,0	0,0
14	15,1	6,6	0,0	0,0	12,0	0,0	0,0	0,0	18,8	10,9	0,0	21,9	14,7	0,0	0,0	0,0	0,0
15	8,9	8,9	5,8	0,0	0,0	0,0	39,2	0,0	12,0	0,0	9,5	0,0	15,9	0,0	0,0	0,0	0,0
16	27,0	2,3	0,0	0,0	30,5	0,0	0,0	0,0	32,6	0,0	0,0	7,6	0,0	0,0	0,0	0,0	0,0
17	17,5	0,0	0,0	0,0	67,4	0,0	0,0	0,0	0,0	0,0	0,0	12,5	0,0	2,6	0,0	0,0	0,0
18	40,0	10,1	15,8	0,0	21,6	0,0	0,0	0,0	7,2	0,0	0,0	0,0	4,7	0,6	0,0	0,0	0,0
19	29,0	5,4	5,9	0,0	32,9	0,0	0,0	0,0	21,0	0,0	0,0	0,0	5,9	0,0	0,0	0,0	0,0
20	27,4	11,0	11,6	0,0	20,5	0,0	10,6	0,0	17,7	0,0	0,0	0,0	0,0	1,2	0,0	0,0	0,0
21	33,6	1,0	4,7	0,0	25,8	0,0	0,0	0,0	34,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
22	34,8	9,8	9,2	0,0	30,3	0,0	0,0	0,0	0,0	0,0	14,6	0,0	0,0	1,2	0,0	0,0	0,0
23	33,4	3,0	16,8	0,0	38,8	0,0	0,0	0,0	7,2	0,0	0,0	0,0	0,0	0,8	0,0	0,0	0,0
24	18,6	7,4	8,5	0,0	12,4	0,0	0,0	0,0	26,3	11,7	0,0	0,0	14,6	0,5	0,0	0,0	0,0
25	22,2	7,5	0,0	0,0	28,6	0,0	0,0	0,0	29,0	0,0	0,0	0,0	11,7	1,1	0,0	0,0	0,0
26	22,8	8,9	7,7	0,0	29,8	0,0	0,0	0,0	26,4	0,0	0,0	0,0	4,5	0,0	0,0	0,0	0,0
27	28,2	2,8	4,2	0,0	31,0	0,0	0,0	0,0	33,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
28	32,3	4,7	0,0	0,0	28,7	0,0	0,0	0,0	19,0	0,0	12,8	0,0	2,6	0,0	0,0	0,0	0,0
29	29,1	0,8	2,4	0,0	27,3	0,0	0,0	0,0	40,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
30	29,1	8,1	0,0	0,0	9,1	0,0	0,0	0,0	0,0	14,3	0,0	26,5	12,9	0,0	0,0	0,0	0,0
31	14,4	12,4	0,0	0,0	25,1	0,0	0,0	10,7	0,0	15,0	0,0	13,8	8,6	0,0	0,0	0,0	0,0
32	11,4	7,9	0,0	0,0	17,1	0,0	0,0	0,0	0,0	0,0	30,1	15,0	18,6	0,0	0,0	0,0	0,0
33	19,1	7,7	0,0	0,0	19,6	0,0	0,0	0,0	0,0	19,0	31,3	0,0	3,3	0,0	0,0	0,0	0,0
34	39,2	4,2	0,0	0,0	10,8	0,0	0,0	0,0	0,0	6,0	23,3	16,6	0,0	0,0	0,0	0,0	0,0
35	3,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	96,7	0,0	0,0
36	30,7	0,0	0,0	0,0	37,8	0,0	0,0	0,0	21,9	0,0	7,6	0,0	2,0	0,0	0,0	0,0	0,0
37	37,6	1,2	3,3	0,0	14,8	0,0	0,0	0,0	21,7	0,0	0,0	20,7	0,8	0,0	0,0	0,0	0,0
38	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	93,4	5,0
39	19,4	3,7	1,4	0,0	10,7	40,4	0,0	0,0	0,0	0,0	0,0	0,0	24,4	0,0	0,0	0,0	0,0
40	15,4	6,6	0,0	0,0	12,9	2,9	0,0	0,0	0,0	25,7	0,0	6,6	30,0	0,0	0,0	0,0	0,0