Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Um estudo sobre declustering e uma proposta de algoritmo genético aplicada a tal situação

Gabriel F. Silva

Monografia apresentada como requisito parcial
para conclusão do Estudo em Sistemas de Informação

Orientador
Prof. Dr. Marcelo Ladeira

Coorientador
Prof. Dr. Claus Aranha

Brasília
2016

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Um estudo sobre declustering e uma proposta de algoritmo genético aplicada a tal situação

Gabriel F. Silva

Monografia apresentada como requisito parcial
para conclusão do Estudo em Sistemas de Informação

Prof. Dr. Marcelo Ladeira (Orientador)
CIC/UnB

Prof. Dr. Claus Aranha
Department of Computer Sciences/University of Tsukuba

Prof. Dr. Guilherme Ramos
CIC/UnB

Prof. Dr. Marcelo Ladeira
CIC/UnB

Prof. Dr. Rodrigo Bonifácio
Coordenador do Bacharelado em Ciência da Computação

Brasília, 8 de julho de 2016

Dedicatória

Dedico este trabalho ao *StackOverflow* e a todos que contribuem para sua excelência.

Agradecimentos

Agradeço a Deus, por ter me dado saúde, disposição e uma família fantástica.

Agradeço aos meus amigos e à minha família, pelo suporte constante.

Agradeço a Universidade de Brasília e ao Departamento de Ciência da Computação por terem um ambiente amigável e pela oportunidade de aprendizado com pessoas qualificadas.

Agradeço os conselhos do ótimo grupo do PIBIC formado por mim, Ícaro, Yuri, Khalil, Ladeira e Claus.

Obrigado a todos! Mostrarei a minha gratidão pagando um brigadeiro a quem estiver nessa lista.

Resumo

Declustering é o nome dado à atividade de classificar os terremotos em *mainshocks* (terremotos independentes) e *aftershocks* (terremotos ao menos parcialmente dependentes dos principais) e conduzir demais análises considerando apenas os *mainshocks*. Essa tarefa costuma ser desejável na geofísica, tendo aplicações na área de modelos de previsão de riscos de sismos. Nesse contexto, o presente trabalho analisa dois métodos já estabelecidos de *declustering* e propõe um método novo, baseado em algoritmos genéticos, o GADec. A pesquisa é feita com base em dados colhidos pela Agência Meteorológica Japonesa, durante os anos de 2000 à 2012. Para os métodos tradicionais, avalia-se se os *mainshocks* fornecidos seguem uma distribuição de Poisson no tempo, e se, usando apenas os terremotos principais indicados por tais métodos, pode-se melhorar a performance do GAModel, um modelo de previsão de riscos. Para o GADec, faz-se uma análise preliminar do algoritmo proposto, de modo a avaliar sua viabilidade computacional em termos de tempo e sua taxa de convergência. Os resultados obtidos para os métodos tradicionais de *declustering* foram medianos: para algumas regiões, após a aplicação dos métodos, foi possível negar a hipótese nula H_0 = "Os *mainshocks* seguem uma distribuição de Poisson", o que não deveria ocorrer em teoria. Por outro lado, foi possível mostrar que o uso de tais métodos realmente melhora o desempenho do GAModel. Por fim, os resultados preliminares obtidos para o GADec mostram que o algoritmo proposto é um fracasso. O tempo de execução do algoritmo mostrou-se inviável para catálogos com tamanho razoável, e o algoritmo não teve uma convergência razoável.

Palavras-chave: algoritmos genéticos, clustering, declustering

Abstract

Declustering is an important task in the geophysics area, with application on quake forecasting models. It consists of classifying quakes into mainshocks (independent quakes) and aftershock(quakes at least partially dependend of mainshocks). In this context, this work analyses two established declustering methods and proposes a new one (the GADec), based on genetic algorithms. The data used was colected and organized by the Japan Meteorological Agency, and covers the years from 2000 to 2012. For the study of the traditional methods, it's analysed whether the mainshocks follow a Poisson distribution in time and whether using only the mainshocks (instead of the whole catalog) results in an improvement in the performance of the GAModel, a quake forecasting model. For the GADec, it's done an initial analysis regarding its time efficiency and rate of convergence. The results obtained were both good and bad. On the bad size, after applying the techniques, it was possible, for some regions, to refute the null hyphotesis H_0 = "*Mainshocks* follow a Poisson distribution", which goes against the predicted in theory. On the good size, it was possible to show that using the established methods improves the perfomance of the GAModel. Finally, the preliminary results obtained by the GADec indicates that the proposed algorithm failed: the time of execution was impractible for catalogs with reasonable size and the algorithm didn't present a reasonable rate of convergence.

Keywords: genetic algorithms, clustering, declustering

Sumário

1 Introdução **1**

1.1 Resumo da Monografia 1

1.2 Definição do Problema 1

1.3 Objetivos 2

2 Fundamentação Teórica **3**

2.1 *Clustering* e *Declustering* 3

2.1.1 *Clustering* 3

2.1.2 *Declustering* 3

2.2 Métodos tradicionais de *declustering* 4

2.2.1 O Método da Janela 4

2.2.2 O Método Single Link Clustering 5

2.3 A distribuição de Poisson como métrica para avaliar a qualidade de *declustering* 6

2.4 Algoritmos genéticos 6

2.4.1 Analogias com a biologia 7

2.4.2 Principais operadores 7

2.4.3 Funcionamento de um GA 8

2.5 GADec - Uma proposta de algoritmo genético para *declustering* 9

2.5.1 Indivíduos 9

2.5.2 População Inicial 10

2.5.3 Função de *Fitness* 10

2.5.4 *Crossover*, Mutação, Seleção e Elitismo 10

3 Dados disponíveis e a divisão do catálogo em regiões **12**

3.1 O catálogo da JMA 12

3.2 A divisão em regiões 12

4 Metodologia **17**

4.1 Imagens com os *clusters* dos maiores terremotos 17

4.2 O teste de chi-quadrado para validar a hipótese de Poisson 18
4.3 O teste de Kruskal-Wallis para avaliar a performance do GAModel 18

5 Resultados Obtidos **20**
5.1 Análise da convergência do GADec 20
5.2 Quantidade de *clusters* encontrada em cada método 20
5.3 *Cluster* dos terremotos com maior magnitude 21
5.3.1 Método da janela 21
5.3.2 Método SLC 21
5.4 Resultados obtidos usando o teste de chi-quadrado 22
5.5 Performance do GAModel 24
5.5.1 Método da janela 24
5.5.2 Método SLC 27

6 Conclusão **29**
6.1 Análise dos resultados 29
6.1.1 Convergência do GADec 29
6.1.2 Imagens dos *clusters* obtidos 29
6.1.3 Testes de chi-quadrado 29
6.1.4 Testes de Kruskal-Wallis 30
6.2 Conclusão 30
6.3 Possíveis continuações para a pesquisa 30

Referências **32**

Lista de Figuras

2.1 Agrupamento de dados em clusters. 4
2.2 Crossover feito com a recombinação de partes das duas strings. 8

3.1 Mapa da Área referente a todo o catálogo. 13
3.2 Região de Kanto. 14
3.3 Região de Kansai. 14
3.4 Região do Leste do Japão. 15
3.5 Região de Tohoku. 16

5.1 *Fitness* do GADec, conforme o passar das gerações. 21
5.2 *Clusters* dos maiores terremotos ocorridos no Japão - método da janela. . . 22
5.3 *Clusters* dos maiores terremotos ocorridos no Leste do Japão - método da janela. 22
5.4 *Clusters* dos maiores terremotos ocorridos em Kansai - método da janela. . 23
5.5 *Clusters* dos maiores terremotos ocorridos no Leste do Japão - método SLC. 23
5.6 *Clusters* dos maiores terremotos ocorridos em Kansai - método SLC. 24

Lista de Tabelas

3.1 Quantidade de terremotos por região geográfica. 13

5.1 Quantidade de *clusters* por região geográfica e método. 21
5.2 Valores de p-value para uma magnitude maior ou igual a 3.8. 24
5.3 Valores de p-value para uma magnitude maior ou igual a 4.0. 25
5.4 Teste de Kruskal-Wallis para o método da janela, considerando a região do Leste do Japão. 25
5.5 Teste de Kruskal-Wallis para o método da janela, considerando a região de Kansai. 25
5.6 Teste de Kruskal-Wallis para o método da janela, considerando a região de Kanto. 26
5.7 Teste de Kruskal-Wallis para o método da janela, considerando a região de Tohoku. 26
5.8 Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região do Leste do Japão. 27
5.9 Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Kansai. 27
5.10 Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Kanto. 28
5.11 Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Tohoku. 28

Lista de Abreviaturas e Siglas

CIC Departamento de Ciência da Computação.

GA Algoritmos Genéticos.

JMA Agência Meteorológica Japonesa.

SLC Single Link Clustering.

UnB Universidade de Brasília.

Capítulo 1

Introdução

1.1 Resumo da Monografia

Neste primeiro capítulo é apresentado o problema em questão e o objetivo traçado para a pesquisa. Seguindo isso, no segundo capítulo, é descrita a fundamentação teórica necessária para compreender a pesquisa. No terceiro capítulo, explicam-se os dados disponíveis e como eles foram dividos, enquanto que no quarto capítulo detalha-se a metodologia que foi seguida durante o desenvolvimento da pesquisa. No quinto capítulo, mostram-se os resultados obtidos e, por fim, no sexto capítulo, analisam-se os resultados obtidos, apresenta-se a conclusão relativa ao trabalho feito e indicam-se possíveis caminhos de continuação da pesquisa.

1.2 Definição do Problema

O conhecimento acerca de terremotos é fundamental para minimizar os impactos causados por tais fenômenos, sendo obtido a partir da análise de dados registrados anteriormente, que ficam organizados em catálogos de terremotos [13].

A partir desse catálogo, pode-se separar os terremotos em *mainshocks* (terremotos independentes) e *aftershocks* (terremotos que são, ao menos parcialmente, dependentes de terremotos principais). Feito isso, é possível eliminar da análise os *aftershocks*, e trabalhar apenas com os *mainshocks* [13]. Tal processo é importante para as pesquisas na área de avaliação de riscos sísmicos e na área de modelos de previsão de terremotos, e recebe o nome de *declustering* [13].

1.3 Objetivos

Este projeto de pesquisa tem como objetivo aplicar métodos de *declustering* tradicionais ao catálogo de terremotos da Agência Meteorológica Japonesa (JMA, do inglês *Japan Meteorological Agency*), que contém os terremotos ocorridos no Japão de 2000 até 2012. Do resultado obtido será feita uma análise para ver se os *mainshocks* seguem propriedades previstas em teoria e se a utilização de tais métodos melhora a performance do GAModel [2], um modelo de previsão de risco de sismos. Por fim, uma tentativa de aplicar Algoritmos Genéticos (GA, do inglês *Genetic Algorithms*) para realizar o *declustering* é feita, e analisa-se se a modelagem proposta é viável.

Capítulo 2

Fundamentação Teórica

Neste capítulo são apresentados os fundamentos teóricos necessários para a compreensão da pesquisa. Inicialmente, é definida a tarefa de *clustering* e de *declustering* e comenta-se sobre os métodos tradicionais de *declustering* e quais métricas podem ser definidas para se avaliar os resultados obtidos. Por fim, aborda-se o tema de algoritmos genéticos e como podemos usá-los para a tarefa de *declustering*.

2.1 *Clustering* e *Declustering*

2.1.1 *Clustering*

Clustering é a tarefa de agrupar uma coleção de dados, de modo que dados pertencentes a um mesmo grupo sejam o mais similar possível (tal similaridade depende da aplicação em questão), e sejam diferentes dos dados pertencentes a outro grupo [1]. A Figura 2.1 ilustra a classificação de dados (representados como pequenos quadrados) em 3 *clusters* diferentes (representados pelas 3 cores diferentes).

Trata-se de uma campo de estudos com aplicações para as áreas de mineração de dados, aprendizagem de máquina, reconhecimento de padrões entre outras [1]. Como há vários modelos para *clustering*, não há um melhor algoritmo para a tarefa de *clustering* , sendo a escolha do algoritmo dependente do contexto.

2.1.2 *Declustering*

Declustering corresponde a técnica de, a partir de um cátalogo de sismos, separar os terremotos em duas classes: *mainshocks* e *aftershocks* [13]. Os *mainshocks* são terremotos independentes, enquanto que os *aftershocks*(também chamados de *foreshocks*, *offsprings* ou *triggered earthquakes*) são terremotos ao menos parcialmente dependentes de algum *mainshock* [13].

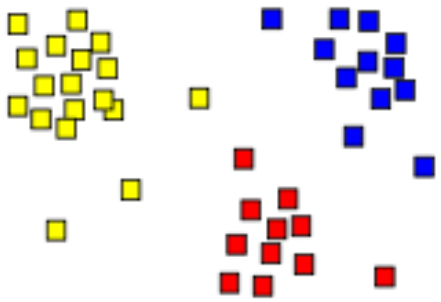

Figura 2.1: Agrupamento de dados em clusters.

Assim, o nome *declustering* vem da possibilidade de, após separarmos os *mainshocks* dos *aftershocks*, removermos os *aftershocks* e usarmos apenas os *mainshocks* em um determinado problema, evitando informações redundantes [13]. Percebe-se então que *declustering* está relacionado com *clustering*, visto que *declustering* pode ser visto como a tarefa de agrupar os dados de um catálogo, i.e *clustering*, e considerar, para cada agrupamento, apenas o *mainshock*. Este tratamento de dados é importante, tendo aplicações para as áreas de avaliação de riscos sísmicos e modelos de previsão de sismos [13].

2.2 Métodos tradicionais de *declustering*

Existem vários métodos da geofísica que realizam o *declustering* de um catálogo de terremotos, não existindo um consenso sobre qual método é melhor [13]. Entre esses métodos, pode-se destacar o método da janela [5], o método Single Link Clustering (SLC) [4] [3], o método de Reasenberg [11] e o método ETAS [14]. Nesse trabalho, estudou-se e implementou-se o método da janela e o método SLC, os quais são explicados a seguir.

2.2.1 O Método da Janela

O método da janela é uma maneira simples de separar *mainshocks* de *aftershocks*. Para um determinado terremoto do catálogo, os outros terremotos são identificados como *aftershocks* caso eles tenham magnitude menor e ocorram próximos do primeiro terremoto, tanto em termos de tempo quanto em termos de distância física. Caso um segundo terremoto esteja próximo temporalmente do primeiro terremoto, diz-se que o segundo está na janela do tempo do primeiro e, caso o segundo terremoto tenha ocorrido próximo geograficamente do primeiro, diz-se que o segundo está na janela da distância do primeiro [13].

Essa noção de proximidade depende da magnitude do terremoto para o qual estamos analisando a janela, i.e., os intervalos de tempo e de distância são funções da magnitude M de um terremoto: t(M) e d(M) [13]. Os parâmetros que definem o comprimento da janela do tempo e da distância mudam de um estudo para outro e usualmente não são obtidos a partir de um processo de otimização [13]. Para a pesquisa em questão, considerou-se as janelas como detalhados por Gardner e Knopoff [5]:

$$d(M) = 10^{0.1238*M+0.983}[km] \tag{2.1}$$

$$t(M) = \begin{cases} 10^{0.032*M+2.7389}, & \text{se } M \geq 6.5 \\ 10^{0.5409*M-0.547}, & \text{caso contrário} \end{cases} [dias] \tag{2.2}$$

2.2.2 O Método Single Link Clustering

O método Single Link Clustering é um método de *clustering* também existente em outras áreas do conhecimento [4]. Trata-se de um método de *clustering* hierárquico e aglomerativo, ou seja, é construída uma hierarquia de *clusters*, começando com todos os terremotos em *clusters* separados e unindo *clusters* a cada etapa.

O algoritmo é descrito abaixo:

1. Inicialmente cada terremoto encontra-se em um *cluster* próprio.
2. Calcule uma distância máxima D, tal que *clusters* que estejam a uma distância maior do que D não possam ser unidos em uma único *cluster*.
3. Una cada *cluster* ao *cluster* mais próximo dele, obtendo assim um único *cluster*. Não realize essa ação apenas caso a distância entre esses dois *clusters* exceda D.
4. Caso tenha ocorrido alguma união entre *clusters* no passo 3, volte ao passo 3. Caso nenhuma união tenha ocorrido ou o número de *clusters* seja igual a 1 o algoritmo termina.

No algoritmo para *clustering* descrito, define-se a distância entre dois *clusters* como sendo igual a menor distância entre dois terremotos que pertençam a eles. De acordo com [4], para o cálculo da distância entre dois terremotos, utiliza-se a fórmula:

$$d_{st} = \sqrt{d^2 + C^2 * t^2} \tag{2.3}$$

onde d_{st} é a métrica de distância entre os terremotos, d é a distância geográfica entre os dois terremotos (medida em kilometros), t é o tempo decorrido entre os dois terremotos

(medido em dias) e C é uma constante que permite a comparação entre tempo e distância (tendo portanto $km * dia^{-1}$ como unidade de medida). Utilizou-se $C = 1$, conforme sugerido pelos autores [4].

Por fim, para o cálculo da distância máxima os autores propõe a fórmula:

$$D = 9.4 * \sqrt{S_1} - 25.2 \tag{2.4}$$

onde S_1 é a mediana de todas as distâncias entre os terremotos. Aplicando-se o algoritmo Single Link Clustering, obtêm-se um agrupamento de terremotos de modo que para finalizar a tarefa de *declustering*, basta selecionar de cada *cluster* o terremoto mais representativo. Escolheu-se selecionar aquele mais próximo do centróide do cluster.

2.3 A distribuição de Poisson como métrica para avaliar a qualidade de *declustering*

Embora não exista um método único de *declustering*, pode-se avaliar a qualidade dos resultados obtidos através de algumas métricas, que levam em conta conhecimentos da geofísica. Nessa seção é explicada uma das métricas utilizada nessa pesquisa: a distribuição de Poisson.

Alguns estudos ([5] e [13]) comprovam que a distribuição de terremotos em um determinado catálogo não segue uma distribuição de Poisson no tempo. Apesar disso, estudos mostraram que, para regiões tectônicas grandes o suficiente, os terremotos principais com magnitude maior que um dado limite (comumente em torno de 3.8 ou 4.0 [9]) são homogêneos no tempo, ou seja, seguem uma distribuição de Poisson [13]. Assim sendo, verificar se um catálogo de terremotos contendo apenas os *mainshocks* com magnitude maior que o limite segue uma distribuição de Poisson pode ser uma boa medida para avaliar a qualidade do método de *declustering*.

2.4 Algoritmos genéticos

Os algoritmos genéticos são um campo de estudos da área de inteligência artificial, com muitas aplicações para problemas de busca em um espaço de soluções [10]. Nessa abordagem, parte-se de um conjunto de soluções para um determinado problema e, com base na aplicação de operadores inspirados em ideias provenientes da biologia, uma melhor solução para o problema em questão é encontrada. Assim, pode-se considerar algoritmos genéticos um método de busca heurística, que procuram no conjunto de soluções uma melhor solução [6].

2.4.1 Analogias com a biologia

Na área de algoritmos genéticos faz-se referência a muitos conceitos da biologia. Assim, para uma melhor compreensão sobre o assunto, deve-se explicar tais referências, e é o que fazemos a seguir. Uma possível solução para o problema em questão é chamada de indivíduo [6]. Por consequência, um conjunto de possíveis soluções para um determinado problema é visto como um conjunto de indivíduos, recebendo então o nome de população [6].

A qualidade de uma solução para a aplicação traduz-se como o quão bem adaptado o indivíduo está, e o objetivo de encontrar a melhor solução tem por analogia encontrar o melhor indivíduo, i.e., o indivíduo mais adaptado, ou seja, aquele com maior *fitness* [6]. Os algoritmos genéticos são baseados na ideia biológica de que, assim como indivíduos podem evoluir e se tornar cada vez mais adaptados, pode-se aplicar operadores a um conjunto inicial de soluções para um problema (ou seja, uma população) de modo a obter soluções cada vez melhores (ou seja, indivíduos cada vez mais adaptados). Tais operadores são descritos a seguir.

2.4.2 Principais operadores

Os principais operadores para a área de GA são *crossover*, seleção e mutação [10]. Entretanto, antes de explicá-los, é necessário definir função de *fitness*. Conforme já dito, encontrar a melhor solução para o problema em estudo corresponde a encontrar o indivíduo mais bem adaptado. A função de *fitness* avalia justamente isso, associando a um indivíduo um valor numérico, que traduz o quão bem o indivíduo representa uma solução para a aplicação [10].

Crossover

O operador de *crossover* gera dois novos indivíduos a partir de dois indivíduos já existentes. Assim como na natureza o indivíduo gerado pelo cruzamento de dois pais tem em seu DNA uma combinação do DNA de seus genitores, em GA os indivíduos gerados pelo operador de *crossover* tem algumas características de cada um dos dois indivíduos que o produziram [6].

É importante perceber que ao gerarmos e avaliarmos um novo indivíduo estamos estudando uma nova solução no espaço de busca, e que o indivíduo gerado por *crossover* pode ter *fitness* maior que seus pais ou não. Na representação mais simples e comum de GA, indivíduos são representados como *strings* de *bits*. O *crossover* pode ser realizado então combinando partes das duas *strings*, conforme mostra a Figura 2.2, em que os indivíduos

c e d são descendentes dos indivíduos a e b. Tal técnica de *crossover* é conhecida por *one-point crossover*.

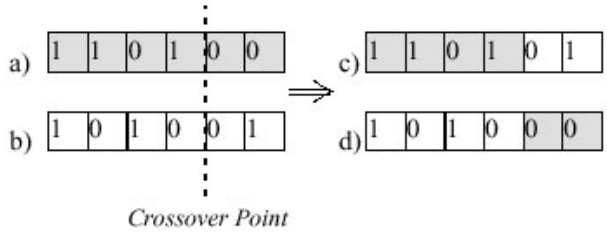

Figura 2.2: Crossover feito com a recombinação de partes das duas strings.

Seleção

Assim como na natureza os indivíduos mais bem adaptados tem maior probabilidade de gerar descendentes, os algoritmos genéticos devem valorizar e combinar as melhores soluções já encontradas, enquanto removem ou dão poucas chances às piores soluções, para que possam obter soluções cada vez melhores. O operador responsável por isso é o operador de seleção. Tal operador costuma selecionar os indivíduos de acordo com a sua *fitness*, i.e., ao quão bem eles resolvem o problema em questão [6].

Uma técnica comum para seleção é a *Deterministic Tournament Selection*, na qual selecionam-se n indivíduos aleatoriamente da amostra e, dentre eles, os k indivíduos mais aptos são escolhidos para *crossover*.

Mutação

Por fim, o operador de mutação deve fazer pequenas modificações em um indivíduo, na esperança de que essas pequenas alterações gerem um indivíduo com maior *fitness*. A mutação é um operador importante, já que aumenta a variabilidade da população e contribui para evitar soluções que são simplesmente ótimos locais [6].

2.4.3 Funcionamento de um GA

Cada indivíduo tem sua *fitness* avaliada, e o melhor indivíduo é guardado em uma lista. Em seguida, aplica-se o operador de seleção, para selecionar os individuos que irão gerar descendentes. Depois, há a operação probabilística do operador de *crossover*, gerando-se os descendentes da população. O operador de mutação pode (de acordo com uma determinada probabilidade) ser aplicado, aumentando a diferença (variabilidade) dos descendentes gerados em relação aos seus pais e contribuindo para evitar ótimos locais. Por fim, os descendentes gerados substituem (total ou parcialmente) a população que os gerou.

Há várias técnicas para realizar essa substituição geracional: pode-se substituir apenas os n piores indivíduos da população, onde n é um parâmetro a ser especificado ou pode-se garantir que o melhor indivíduo da população antiga não será nunca substituído (tal técnica se chama elitismo) ou pode-se substituir a população como um todo [6].

Tal processo é repetido várias vezes, para que várias soluções sejam exploradas. O número de vezes que o processo de gerar descentes a partir de uma população é repetido é conhecido como número de gerações, e fica tipicamente na casa das centenas. O melhor indivíduo gerado por todas as gerações é retornado, e corresponde a melhor solução encontrado pelo algoritmo genético[6].

Os algoritmos genéticos podem então ser vistos como um método de busca heurística. O fato de serem algoritmos estocásticos permite que busquem soluções variadas no espaço de busca a cada geração, enquanto que a heurística de combinar e valorizar as melhores soluções enquanto dão poucas chances a soluções piores torna-os computacionalmente eficientes para grandes volumes de dados. De fato, diversos estudos comprovam a riqueza de aplicações de GA [10].

2.5 GADec - Uma proposta de algoritmo genético para *declustering*

Explica-se a seguir o GADec, nome dado ao algoritmo genético para a tarefa de *declustering* proposto nessa pesquisa.

2.5.1 Indivíduos

No GADec, cada indivíduo é representado como um *array*, em que a quantidade de elementos do vetor é igual ao número de terremotos sendo analisado. Cada elemento do *array* pode assumir apenas dois valores: 0 ou 1. O valor 1 indica que o terremoto representado naquela posição é centróide do seu *cluster*, enquanto que um valor 0 indica que o terremoto correspondente à posição não é centróide do seu *cluster*. Para determinar a qual *cluster* um terremoto com valor 0 associado pertence, calcula-se a distância do terremoto a todos os centróides e escolhe-se o centróide com a menor distância. A métrica para calcular a distância entre dois terremotos usada é a mesma da descrita no método SLC.

Essa codificação do indivíduo significa que pode-se a partir do *array* do indivíduo obter todo o agrupamento de terremotos que o indivíduo representa. Há, entretanto, um custo computacional, já que terremotos que não são centróides devem ser associados ao centróide mais próximo.

2.5.2 População Inicial

Um indivíduo inicial é gerado através de um processo de dois passos:

1. Escolhe-se o número de centróides do indivíduo de modo aleatório, com a restrição ser uma quantidade entre o número mínimo de *clusters* encontrado nos métodos anteriores e o número máximo de *clusters* encontrado nos métodos anteriores.

2. Escolhe-se quais terremotos serão centróides para aquele indivíduo. Tal escolha é aleatória, respeitando a restrição de que o número de centróides escolhidos deve ser correspondente ao calculado no primeiro passo.

A população inicial é então obtida simplesmente repetindo o processo descrito anteriormente, até que a quantidade de indivíduous seja igual ao estabelecido previamente.

2.5.3 Função de *Fitness*

Para a função de *fitness*, utilizou-se a fórmula:

$$\sum_{1 \leq i \leq k} D_{inter}(C_i)w - D_{intra}(C_i) \tag{2.5}$$

onde $D_{inter}(C_i)$ é a distância intracluster para o *cluster* (C_i) e $D_{intra}(C_i)$ é a distância entre (C_i) aos demais *clusters*. k é o número de *clusters* e w é um parâmetro definido pelo usuário. Para o GADec, considerou-se w como sendo igual a $\frac{1}{2}$.

Escolheu-se tal fórmula por ela já ter sido usada em outro estudo sobre *clustering* com algoritmos genéticos [12]. Poderia-se assim verificar se uma fórmula já usada na literatura de GA e *clustering* é adequada ao domínio de agrupamento de sismos.

2.5.4 *Crossover*, Mutação, Seleção e Elitismo

O funcionamento do GADec é baseado na ação de 3 operadores principais: *crossover*, seleção e mutação. Além destes, aplicou-se técnicas de elitismo para que os melhores indivíduos fossem preservados durante a substituição geracional. Cada um desses tópicos é explicado a seguir.

Crossover

O operador de *crossover* do GADec é o *one-point crossover*, e a probabilidade escolhida para a aplicação de tal operador é de 0.9.

Mutação

O operador de mutação é o *flip-bit*, em que *bits* do *array* do indivíduo são modificados (se valem 1 passam a valer 0, se valem 0 passam a valer 1). A probabilidade escolhida para que o *flip-bit* fosse aplicado a uma posição do *array* foi de 0.01.

Seleção

O operador de seleção escolhido foi o *Deterministic Tournament Selection*, em que torneios com 5 indivíduos foram realizados e os 2 indivíduos mais aptos foram selecionados para *crossover*.

Elitismo

Foi feito elitismo, já que para cada geração, os dois indivíduos mais aptos são colocados como indivíduos da geração seguinte.

Capítulo 3

Dados disponíveis e a divisão do catálogo em regiões

O presente capítulo descreve o formato dos dados disponíveis e aborda a divisão de tais dados em 4 regiões.

3.1 O catálogo da JMA

O catálogo utilizado durante a pesquisa contém dados acerca de terremotos coletados pela JMA, começando no início do ano 2000 e terminando no fim do ano de 2012. Para cada um dos terremotos, é conhecida sua latitude, longitude, magnitude, profundidade, data e momento de ocorrência.

3.2 A divisão em regiões

Um dos objetivos da pesquisa era o estudo de como métodos de *declustering* poderiam melhorar a performance do GAModel, um modelo de previsão de risco de sismos. Teve-se portanto de considerar as regiões usadas na pesquisa que implementa o GAModel [2], as quais são descritas a seguir:

- **Kanto**: região com latitude entre 34.8 e 37.05 e com longitude entre 138.8 e 141.05.
- **Kansai**: região com latitude entre 34.0 e 36.0 e com longitude entre 134.5 e 136.5.
- **Tohoku**: região com latitude entre 37.0 e 41.0 e com longitude entre 139.8 e 141.8.
- **Leste do Japão**: região com latitude entre 37.0 e 41.0 e com longitude entre 140.0 e 144.0.

Tabela 3.1: Quantidade de terremotos por região geográfica.

Região Geográfica	**Quantidade de terremotos**
Todo o Catálogo	220 195
Kanto	15 694
Kansai	1 869
Tohoku	14 072
Leste do Japão	43 561

Como o número de terremotos que ocorreram em cada região impacta a viabilidade computacional em termos de tempo dos algoritmos estudados, mostra-se na Tabela 3.1 a quantidade de terremotos por região.

Finalmente, por finalidades ilustrativas, mostra-se a seguir os mapas para as 4 regiões.

Figura 3.1: Mapa da Área referente a todo o catálogo.

Figura 3.2: Região de Kanto.

Figura 3.3: Região de Kansai.

Figura 3.4: Região do Leste do Japão.

Figura 3.5: Região de Tohoku.

Capítulo 4

Metodologia

O presente capítulo aborda a metodologia seguida durante o projeto de pesquisa. São detalhados os testes feitos para se chegar nos resultados e o porquê de tais testes terem sido escolhidos. Planejava-se realizar tais testes para as 4 regiões do Japão mencionadas previamente e também para o catálogo como um todo, mas o tempo que seria gasto por alguns dos algoritmos em algumas regiões não permitiu isso para todos os métodos.

Para o método da janela, foi possível fazer tal análise para todas as áreas mencionadas. Já para o método Single Link Clustering, o cálculo da distância máxima D conforme a fórmula 2.3 não foi possível (fez-se tal cálculo e obteve-se um valor negativo). Decidiu-se então escolher o valor de D como sendo igual a 12, pois tal escolha já havia sido feita em estudos anteriores [3]. Embora para tal método não tenha sido viável executá-lo para o catálogo como um todo, foi possível a sua execução para cada uma das 4 regiões definidas no GAModel.

Finalmente, para o método GADec não foi possível executá-lo para a grande maioria das regiões estudadas, devido ao tempo que seria gasto (o cálculo da função de *fitness* mostrou-se muito lento para catálogos com tamanho razoável), tendo sido portanto implementado apenas para a região de Kansai. Para tal região, executou-se o algoritmo 10 vezes (devido ao fato do GADec ser estocástico, deve-se executá-lo mais de uma vez) com 50 indivíduos por população e 100 gerações. Fez-se uma análise da convergência de tal algoritmo, obtendo-se resultados preliminares ruins. Devido a isso, não foram feitas as imagens dos *clusters* para tal método, os testes de Poisson e os testes de Kruskal-Wallis.

4.1 Imagens com os *clusters* dos maiores terremotos

Com a finalidade de ilustrar a pesquisa em questão, foram feitas imagens mostrando os *clusters* dos terrremotos com maior magnitude. Analisou-se o catálogo como um todo e também as regiões do Leste do Japão e de Kansai (não foram feitas imagens para as

demais regiões para que esse documento não ficasse tão longo). Para que as imagens não ficassem muito carregadas, mostram-se apenas os *clusters* dos 10 terremotos com maior magnitude para as regiões e os *clusters* dos 15 maiores terremotos para o catálogo como um todo. Para cada *cluster*, são mostradas a data de ocorrência do *mainshock* e a maior magnitude dos sismos que compõem o agrupamento.

4.2 O teste de chi-quadrado para validar a hipótese de Poisson

Para cada um dos métodos, e para cada uma das regiões estudadas em um determinado método, avaliou-se caso os *mainshocks* obtidos seguiam a distribuição de Poisson ou não. Além disso, avaliou-se também se o catálogo original seguia uma distribuição de Poisson ou não. Para isso, são feitos dois testes de chi-quadrado, em que a hipótese nula é: H_0 = "Os terremotos seguem uma distribuição de Poisson no tempo" (ou seja, os terremotos são homogêneos no tempo).

No primeiro teste, consideram-se apenas *clusters* com terremotos com magnitude maior ou igual a 3.8, enquanto que no segundo teste apenas *clusters* com terremotos com magnitude maior ou igual a 4.0 são levados em conta. Em ambos os testes, analisou-se a frequência de terremotos para intervalos de 10 dias. A escolha de tais valores ocorreu pela sua utilização prévia em estudos anteriores sobre o tema [5].

Funciona da seguinte maneira: é feito um teste de chi-quadrado, obtendo-se assim um valor de p-value. Caso esse valor seja menor que o nível de significância α pré-estabelecido, pode-se rejeitar a hipótese nula H_0 com confiabilidade 1 - α [7]. Para a pesquisa em questão, considerou-se um valor de α igual a 0.05, por tratar-se de um valor comumente usado em pesquisas [7].

4.3 O teste de Kruskal-Wallis para avaliar a performance do GAModel

O GAModel é um modelo de previsão de risco de sismos que, a partir de um registro contendo dados de terremotos de anos anteriores, tenta prever onde ocorrerão os terremotos de um determinado ano [2]. Para avaliar o impacto dos métodos de *declustering* na performance do GAModel procedeu-se da seguinte maneira para cada um dos métodos e cada uma das 4 regiões:

1. Sendo Algoritmos Genéticos estocásticos, foram executadas 10 simulações do GAModel quando este recebia como entrada o catálogo original. Para este caso, o GA-

Model faz um pequeno filtro, mantendo apenas terremotos com magnitude maior do que 3.0 e com profundidade menor que 25km.

2. Foram executadas 10 simulações do GAModel quando este recebia como entrada o catálogo declusterizado, ou seja, contendo apenas os dados relativos aos *mainshocks*.

3. Compara-se os desempenhos do GAModel com e sem o uso de *declustering* através de um teste de Kruskal-Wallis, permitindo assim verificar se é possível rejeitar a hipótese nula $H_0 =$ "A média de performance dos dois grupos é igual".

Capítulo 5

Resultados Obtidos

O capítulo atual apresenta os resultados obtidos. Para começar, mostra-se a quantidade de *clusters* encontrada em cada método por cada algoritmo. Depois, apresentam-se as imagens dos *clusters* e os mapas de calor obtidos. Em seguida, é mostrada a tabela com o valor de p-value obtido para cada um dos métodos no teste de chi-quadrado. Por fim, apresenta-se a média de performance do GAModel para cada um dos métodos, comparando-a com a média de performance do GAModel sem o uso de técnicas de *declustering* e apresenta-se o valor de p-value para o teste de Kruskal-Wallis.

5.1 Análise da convergência do GADec

Para a análise da convergência do GADec, fez-se um gráfico (Figura 5.1) mostrando para cada geração a média da *fitness* (das 10 execuções) do melhor indivíduo daquela geração. Na Figura 5.1, é mostrada também quanto a maior *fitness* e a menor *fitness* (das 10 execuções) do melhor indivíduo de um geração diferem da média da *fitness*, através de barras de erro.

5.2 Quantidade de *clusters* encontrada em cada método

Na Tabela 5.1 é mostrada a quantidade de terremotos contida por região e a quantidade de *clusters* encontrada por cada método. Para uma combinação entre região e método que não foi estudada, asteriscos foram colocados na devida célula da tabela.

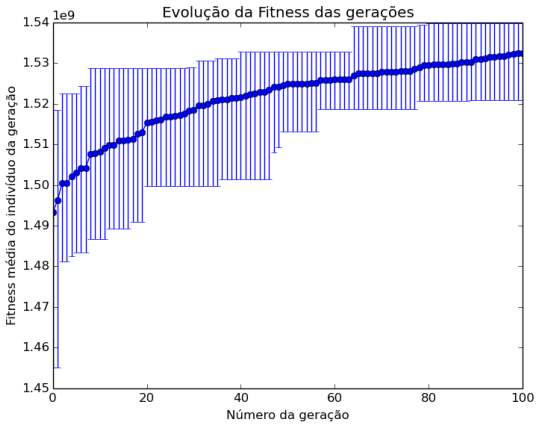

Figura 5.1: *Fitness* do GADec, conforme o passar das gerações.

Tabela 5.1: Quantidade de *clusters* por região geográfica e método.

Região	**Nº de sismos**	**Clusters-Janela**	**Clusters-SLC**
Todo o catálogo	220 195	32 947	*
Kanto	15 694	667	3805
Kansai	1 869	818	1343
Tohoku	14 072	699	2809
Leste do Japão	43 561	1156	9223

5.3 *Cluster* dos terremotos com maior magnitude

5.3.1 Método da janela

Para o método da janela são apresentados as imagens referentes ao catálogo como um todo (Figura 5.2), à região do Leste do Japão (Figura 5.3) e à região de Kansai (Figura 5.4).

5.3.2 Método SLC

Para o método SLC, são apresentadas as imagens referentes ao Leste do Japão (Figura 5.5) e à região de Kansai (Figura 5.6).

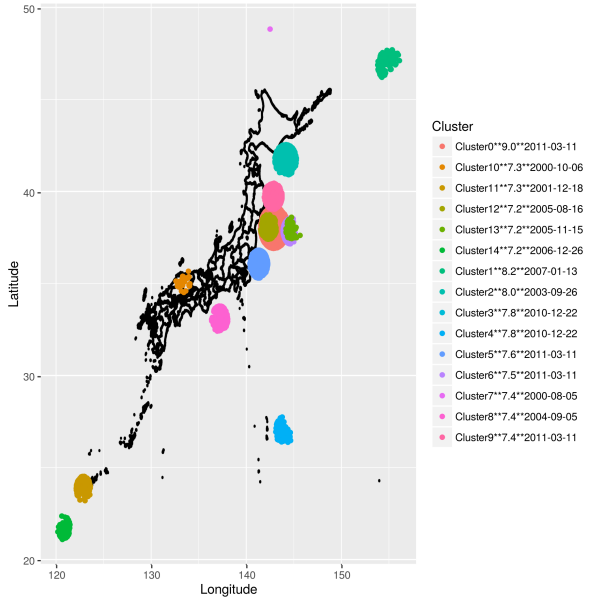

Figura 5.2: *Clusters* dos maiores terremotos ocorridos no Japão - método da janela.

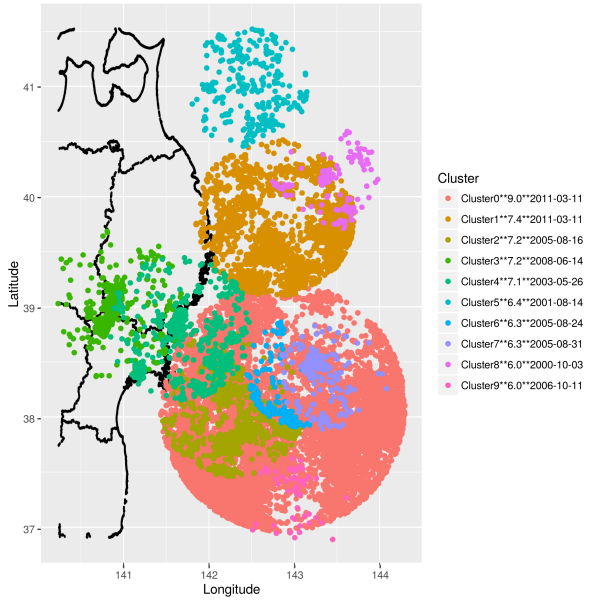

Figura 5.3: *Clusters* dos maiores terremotos ocorridos no Leste do Japão - método da janela.

5.4 Resultados obtidos usando o teste de chi-quadrado

Nesta seção, mostram-se os valores de p-value obtidos no teste de chi-quadrado. Para uma combinação entre região e método que não foi estudada, asteriscos foram colocados na devida célula da tabela. Na Tabela 5.2 temos os resultados considerando uma magnitude

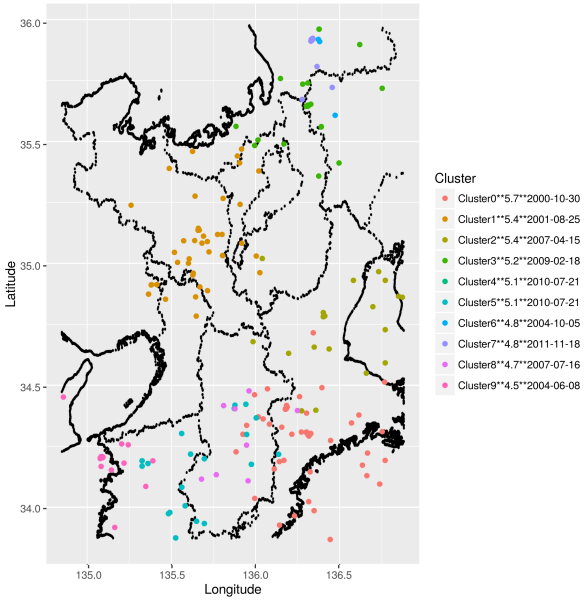

Figura 5.4: *Clusters* dos maiores terremotos ocorridos em Kansai - método da janela.

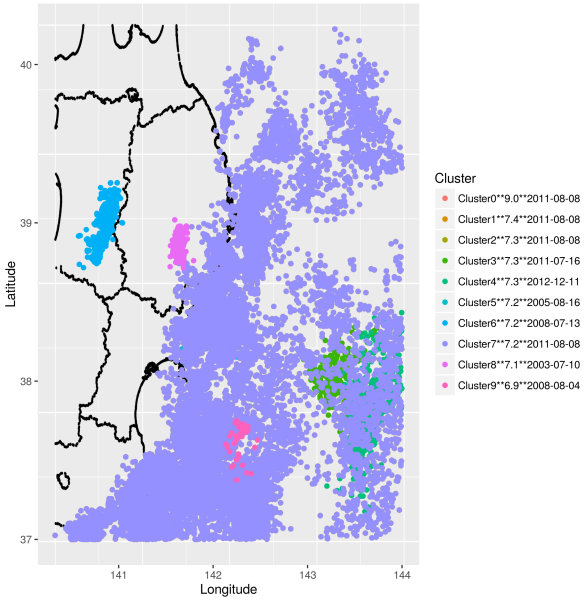

Figura 5.5: *Clusters* dos maiores terremotos ocorridos no Leste do Japão - método SLC.

limite de 3.8.

Na Tabela 5.3 temos os resultados considerando uma magnitude limite de 4.0.

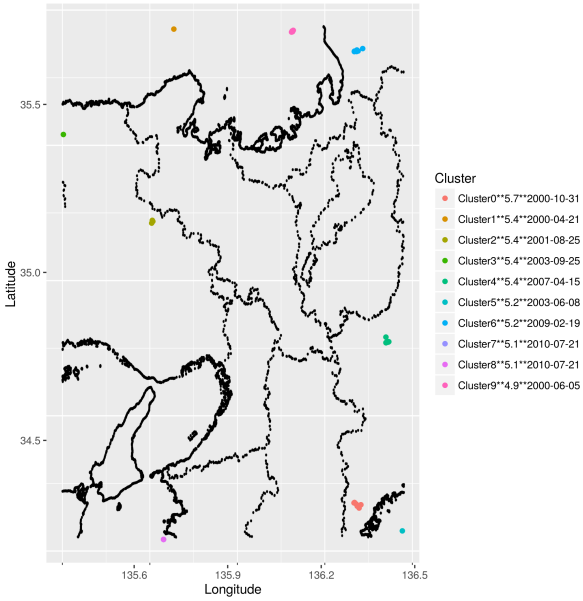

Figura 5.6: *Clusters* dos maiores terremotos ocorridos em Kansai - método SLC.

Tabela 5.2: Valores de p-value para uma magnitude maior ou igual a 3.8.

Região	**Sem *declustering***	**Janela**	**SLC**
Todo o catálogo	0.0	3e-37	*
Kanto	0.0	0.11	0.62
Kansai	1e-54	4e-6	0.66
Tohoku	0.0	1.43e-6	0.30
Leste do Japão	0.0	0.01	2e-28

5.5 Performance do GAModel

Nesta seção, mostra-se a comparação do resultado da performance do GAModel quando métodos de *declustering* são usados com a performance do GAModel sem a aplicação dos métodos de *declustering*. Para isso, cada tabela comparativa contém a média da performance do GAModel nas 10 execuções (com e sem o uso de *declustering*) e o valor de p-value para o teste de Kruskal-Wallis.

5.5.1 Método da janela

Temos os resultados obtidos no teste de Kruskal-Wallis para o método da janela nas tabelas: Tabela 5.4, Tabela 5.5, Tabela 5.6, Tabela 5.7.

Tabela 5.3: Valores de p-value para uma magnitude maior ou igual a 4.0.

Região	**Sem *declustering***	**Janela**	**SLC**
Todo o catálogo	0.0	2e-32	*
Kanto	0.0	0.10	0.56
Kansai	4e-18	2e-4	0.77
Tohoku	0.0	9e-9	0.19
Leste do Japão	4e-6	0.01	7e-12

Tabela 5.4: Teste de Kruskal-Wallis para o método da janela, considerando a região do Leste do Japão.

ano	**sem *declustering***	**com método da janela**	**p-value**
2005	-2131.5	-2063.8	1e-4
2006	-1921.3	-1869.3	1e-4
2007	-1842.2	-1776.9	e-5
2008	-3939.5	-3929.5	0.41
2009	-2010.2	-1922.0	1e-4
2010	-2299.5	-2204.3	1e-4

Tabela 5.5: Teste de Kruskal-Wallis para o método da janela, considerando a região de Kansai.

ano	**sem *declustering***	**com método da janela**	**p-value**
2005	-1624.0	-1607.9	1e-4
2006	-1619.1	-1608.9	3e-5
2007	-1616.7	-1614.7	3e-4
2008	-1615.2	-1612.5	1e-4
2009	-1620.3	-1615.6	1e-4
2010	-1615.8	-1612.1	1e-4

Tabela 5.6: Teste de Kruskal-Wallis para o método da janela, considerando a região de Kanto.

ano	**sem *declustering***	**com método da janela**	**p-value**
2005	-2204.8	-2203.0	3e-4
2006	-2171.9	-2170.0	3e-5
2007	-2103.8	-2099.4	1e-4
2008	-2101.0	-2095.2	7e-5
2009	-2280.6	-2276.4	9e-5
2010	-2109.7	-2097.9	7e-5

Tabela 5.7: Teste de Kruskal-Wallis para o método da janela, considerando a região de Tohoku.

ano	**sem *declustering***	**com método da janela**	**p-value**
2005	-962.0	-938.8	9e-5
2006	-892.3	870.9	1e-4
2007	-904.4	-868.0	5e-5
2008	-2863.1	-2839.6	2e-5
2009	-996.6	-976.9	1e-4
2010	-1136.2	-1100.8	5e-5

Tabela 5.8: Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região do Leste do Japão.

ano	**sem *declustering***	**com método SLC**	**p-value**
2005	-2131.5	-2072.8	1e-4
2006	-1921.3	-1870.5	1e-4
2007	-1842.2	-1781.2	1e-4
2008	-3939.5	-3934.5	0.87
2009	-2010.2	-1928.8	1e-4
2010	-2299.5	-2207.7	1e-4

Tabela 5.9: Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Kansai.

ano	**sem *declustering***	**com método SLC**	**p-value**
2005	-1624.0	-1609.4	1e-4
2006	-1619.1	-1609.0	4e-5
2007	-1616.7	-1615.0	0.008
2008	-1615.2	-1612.5	1e-4
2009	-1620.3	-1617.0	1e-4
2010	-1615.8	-1612.8	1e-4

5.5.2 Método SLC

Temos os resultados obtidos no teste de Kruskal-Wallis para o método SLC nas tabelas: Tabela 5.8, Tabela 5.9, Tabela 5.10, Tabela 5.11.

Tabela 5.10: Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Kanto.

ano	**sem *declustering***	**com método SLC**	**p-value**
2005	-2204.8	-2203.0	2e-4
2006	-2171.9	-2170.0	4e-5
2007	-2103.8	-2099.2	1e-4
2008	-2101.0	-2095.5	1e-4
2009	-2280.6	-2276.4	9e-5
2010	-2109.7	-2098.1	1e-4

Tabela 5.11: Teste de Kruskal-Wallis para o método Single Link Clustering, considerando a região de Tohoku.

ano	**sem *declustering***	**com método SLC**	**p-value**
2005	-962.0	-938.6	5e-5
2006	-892.3	870.9	5e-5
2007	-904.4	-868.0	5e-5
2008	-2863.1	-2839.5	1e-5
2009	-996.6	-976.6	5e-5
2010	-1136.2	-1100.8	5e-5

Capítulo 6

Conclusão

Neste capítulo analisam-se os resultados obtidos e indicam-se possíveis caminhos para continuação dessa pesquisa.

6.1 Análise dos resultados

6.1.1 Convergência do GADec

A Figura 5.1 indica que o GADec apresenta problemas. Observando as barras de erro, percebe-se que não houve convergência significativa. Além disso, observando a *fitness* média do melhor indivíduo na geração 0 e comparando com a *fitness* média do melhor indivíduo na última geração, percebe-se que a melhora na *fitness* foi muito baixa (menor que 3%).

6.1.2 Imagens dos *clusters* obtidos

Embora a imagem dos *clusters* tenha sido feita com propósitos ilustrativos, uma observação interessante é a de que no método da janela são mais comuns *clusters* esféricos, enquanto que no método SLC não há tal característica de modo tão acentuado.

6.1.3 Testes de chi-quadrado

Para o catálogo como um todo, i.e. sem a aplicação de nenhum método de *declustering*, foi possível rejeitar a hipótese nula H_0 = "Os terremotos seguem uma distribuição de Poisson", para todas as regiões analisadas, com confiabilidade igual a 95%, o que está de acordo com o previsto em teoria [13].

Já para o método da janela, foi possível rejeitar a hipótese nula para quatro das 5 regiões analisadas. A única exceção foi na região de Kanto. Em teoria, após a aplicação

de tal método, não deveria ser possível refutar a hipótese nula com o grau de confiabilidade estabelecido para nenhuma região [5]. Assim, tal resultado divergiu do previsto em teoria. Por último, é importante ressaltar que os resultados para tal método foram consistentes para os dois valores de magnitude estudados: 3.8 e 4.0.

Para o método SLC os resultados do teste de chi-quadrado também divergiram do previsto em teoria, já que foi possível rejeitar H_0 para a região do Leste do Japão. Conforme já dito, em teoria, não deveria ser possível refutar a hipótese nula para nenhuma das regiões analisadas [13].

6.1.4 Testes de Kruskal-Wallis

Para o método da janela e para o método Single Link Clustering, os resultados do teste de Kruskal-Wallis foram animadores. A performance média do GAModel foi melhor quando tais métodos de *declustering* foram aplicados ao catálogo do que quando nenhum *declustering* foi feito (ou seja, quando todos os terremotos com magnitude maior que 3.0 e profundidade menor que 25km foram considerados). Isso ocorreu para todas as regiões e para todos os anos considerados. Além disso, para o teste de Kruskal-Wallis, foi possível rejeitar a hipótese nula H_0 = "A média de performance dos dois grupos é igual" com confiabilidade de 95% para todas as regiões e para todos os anos considerados (a única exceção foi o Leste do Japão, no ano de 2008), mostrando assim que realmente faz diferença a aplicação de métodos de *declustering*.

6.2 Conclusão

Do trabalho apresentado, pode-se concluir que os métodos de *declustering* tradicionais implementados (método da janela e método SLC) melhoram a performance do GAModel, mesmo que o tempo de ocorrência entre os *mainshocks* nao siga uma distribuição de Poisson. Assim, os resultados ratificam que pode haver vantagens no uso de métodos de *declustering*. Por fim, o estudo propõe um algoritmo genético para a tarefa de *declustering* (o GADec), faz um estudo preliminar deste e conclui que a modelagem proposta parece não convergir para um bom resultado, além de não ser viável para catálogos com tamanho razoável, devido ao tempo necessário para a execução.

6.3 Possíveis continuações para a pesquisa

O trabalho apresentado não cobre várias questões que podem ser pensadas como possíveis caminhos de pesquisa. Poderia-se por exemplo estudar outros métodos tradicionais de *de-*

clustering, como o método de Reasenberg [11] e o método ETAS [14]. Outra possibilidade seria estudar o método da janela e o método Single Link Clustering para parâmetros diferentes dos considerados no estudo. Poderia-se também investigar como *declustering* afeta um outro método de previsão de riscos de sismos, que nao o GAModel.

Outro caminho de pesquisa seria investigar o porquê do teste de chi-quadrado não ter correspondido ao previsto em teoria. Por fim, o estudo falha em mostrar que é possível usar Algoritmos Genéticos no domínio de *declustering*, mas isso não significa que tenha sido mostrada a impossibilidade de combinar GA com *declustering*. Como existem muitas outras técnicas que combinam GA com *clustering* [8], uma continuação natural da pesquisa seria a investigação da viabilidade ou performance de uma dessas outras técnicas para a tarefa de *declustering*.

Referências

[1] Charu C Aggarwal e Chandan K Reddy. *Data clustering: algorithms and applications.* CRC Press, 2013. 3

[2] Aranha C., Lavinas Y., Ladeira M., e Enescu B. Is it possible to use generate good earthquake risk models using genetic algorithms. *Proceedings of the International Conference on Evolutionary Computation Theory and Applications (ECTA-2014)*, 2014. 2, 12, 18

[3] Scott D Davis e Cliff Frohlich. Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification. *Geophysical Journal International*, 104(2):289–306, 1991. 4, 17

[4] Cliff Frohlich e Scott D Davis. Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. *Geophysical Journal International*, 100(1):19–32, 1990. 4, 5, 6

[5] JK Gardner e L Knopoff. Is the sequence of earthquakes in southern california, with aftershocks removed, poissonian. *Bull. Seismol. Soc. Am*, 64(5):1363–1367, 1974. 4, 5, 6, 18, 30

[6] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989. 6, 7, 8, 9

[7] Dawn Griffiths. *Head first statistics.* "O'Reilly Media, Inc.", 2008. 18

[8] Eduardo Raul Hruschka, Ricardo JGB Campello, Alex A Freitas, et al. A survey of evolutionary algorithms for clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 39(2):133–155, 2009. 31

[9] Brad Luen e Philip B Stark. Poisson tests of declustered catalogues. *Geophysical Journal International*, 189(1):691–700, 2012. 6

[10] Melanie Mitchell. *An introduction to genetic algorithms.* MIT press, 1998. 6, 7, 9

[11] Paul Reasenberg. Second-order moment of central california seismicity, 1969–1982. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 90(B7):5479–5495, 1985. 4, 31

[12] Lin Yu Tseng e Shiueng Bien Yang. A genetic approach to the automatic clustering problem. *Pattern Recognition*, 34(2):415–424, 2001. 10

[13] van Stiphout, T., J. Zhuang, e D. Marsan. Seismicity declustering, 2012. 1, 3, 4, 5, 6, 29, 30

[14] Jiancang Zhuang, Yosihiko Ogata, e David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002. 4, 31