

# O uso de provérbios: Uma abordagem estatística

Mendes, Armando B.; Funk, Gabriela & Matthias e Sousa, Áurea

**Resumo:** Neste artigo, apresentam-se os resultados, ainda parciais, do projecto de análise de uma base de dados construída para identificação de provérbios. Esta base de dados integra informação sobre 25.000 expressões idiomáticas e mais de um milhar de respostas válidas a inquéritos de reconhecimento de provérbios. Os objectivos do projecto incluem várias actividades desde a compreensão do domínio e dos dados até à extracção de conhecimento que permita caracterizar utilizadores de grupos de provérbios e prever a região geográfica onde estes têm vivido. A metodologia CRISP-DM é utilizada e ilustrada, de acordo com estes objectivos. Dá-se especial ênfase às fases de preparação e pré-processamento, limpeza de dados e à modelação e análise de dados. Utilizam-se métodos de análise de dados como as árvores de regressão e classificação e a análise de *clusters*. Utilizam-se igualmente dados simbólicos tendo em conta a estrutura complexa e a existência de variabilidade interna. Os resultados já obtidos são comparados com o conhecimento de domínio paremiológico.

**Palavras-chave:** Extracção de Conhecimento; Metodologia CRISP\_DM, Preparação de Dados e Pré-Processamento; Provérbios; Análise de Dados Simbólicos.

## 1. Introdução

A extracção de conhecimento a partir de bases de dados relacionais é uma actividade com crescente interesse em diversas áreas da sociedade, desde o ambiente competitivo das empresas às actividades de investigação científica. Menos usual é a utilização de tais metodologias em áreas pouco quantitativas como o estudo da Cultura Popular de expressão oral. Supõe-se, neste trabalho, que os provérbios representam uma matriz conceptual dos diversos esquemas de pensamento, pelo que o estudo dos provérbios reconhecidos por habitantes de regiões geográficas distintas permite adquirir um conhecimento básico das correntes de pensamento colectivo (Mendes, Funk, & Funk, ).

Considerando válida esta hipótese, propomo-nos, neste trabalho, descobrir padrões úteis que possam gerar conhecimento segundo dois vectores essenciais:

- Caracterizar grupos de utilizadores de provérbios de acordo com o seu grau de reconhecimento;

- Classificar provérbios segundo os locais geográficos onde vivem ou viveram os seus utilizadores.

Tendo em conta estes objectivos, utilizam-se metodologias de *data mining* as quais permitem explorar dados e estimar modelos a partir de dados heterogéneos e de natureza complexa, expressos tanto em escalas quantitativas como não quantitativas (Hand, Mannila & Smyth, 2001 e Diday (2000). No entanto, para obter conhecimento dos dados organizados em bases de dados é necessário um processo longo e não trivial, envolvendo não apenas conhecimentos tecnológicos, mas também conhecimentos do domínio em que foram recolhidos os dados (Berry & Linoff, 1997).

Outros parâmetros relevantes em projectos de grande dimensão com muitos intervenientes são os factores humanos indispensáveis para gerir conflitos e equipas formadas por indivíduos com diferentes fundos culturais, capacidades técnicas, habilitações e hábitos, assim como as técnicas de gestão de projectos e controlo (Lavrač, Motoda *et al.*, 2004).

Vários autores têm sugerido metodologias para abordar problemas de extracção de conhecimento de bases de dados (*e.g.*: Berry & Linoff, 1997, Fayyad, Piatetsky-Shapiro, *et al.* 1996 e Hand, Mannila & Smyth, 2001). No entanto, uma metodologia tem-se destacado como padrão em aplicações de *data mining* para extracção de conhecimento de bases de dados. A metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) tem sido validada com vários projectos de grande dimensão e resulta do trabalho conjunto de especialistas em diferentes sectores no mercado de *data mining* com o objectivo de ser independente do sector e das aplicações em que seria utilizada (Clifton & Thuraisingham, 2001 e Lavrač, Motoda *et al.*, 2004).

Esta metodologia resulta num modelo processual hierárquico que estabelece seis fases, cada uma das quais compreendendo um conjunto de tarefas. A metodologia, actualmente em revisão, recomenda as seguintes fases:

1. Compreender o domínio e o problema. Compreender os objectivos do contexto onde surge o problema, avaliar o problema, estabelecer objectivos para as actividades de *data mining* e produzir um plano do projecto.
2. Explorar e compreender os dados. Recolher alguns dados iniciais, descrever e explorar os dados, verificar a qualidade dos dados.
3. Integração e pré-processamento dos dados. Seleccionar os dados, limpeza dos dados, definir uma estrutura para as tabelas relacionais, integrar os dados e formatar os dados.

4. Modelação e análise de dados. Seleccionar a(s) técnica(s) de modelação, desenhar o método de validação, estimar e validar o modelo.
5. Avaliação e validação dos modelos no contexto do domínio. Avaliar os resultados, rever o processo se necessário e determinar os passos seguintes.
6. Implementação e operação do sistema. Planeamento da implementação, planeamento da monitorização e manutenção, produção de um relatório final e revisão de todo o projecto.

Como frequentemente se verifica em metodologias para apoio à decisão, também esta apresenta ciclos de retorno entre as fases enumeradas, dependendo do grau de satisfação relativamente aos resultados obtidos em cada fase, a que alguns autores chamam a espiral de modelação e extracção de conhecimento (Lavrač, Motoda et al., 2004). Note-se que o próprio processo de apoio à decisão, frequentemente não termina quando a solução é identificada, podendo esta originar novos problemas mais específicos e iniciar novos projectos.

Neste artigo, descreve-se um projecto para a extracção de conhecimento de uma base de dados contendo registos de provérbios e algumas características das pessoas que os reconhecem. Tanto o projecto como o artigo seguem a metodologia CRISP-DM, iniciando-se com a descrição do problema e do domínio onde surge. Mais especificamente pretende-se identificar provérbios típicos de uma determinada zona geográfica, considerando como estratégias sociais bem sucedidas. Para estudos sociolinguísticos e, talvez, sócio-culturais, informações deste tipo representarão uma valiosa fonte.

## **2. Compreender o domínio e o problema**

As cinco características puramente linguísticas (proposição autónoma, breve, independente do contexto, formada por várias palavras e geralmente realizada por uma frase) dos provérbios podem-se resumir no conceito linguístico de texto mínimo. Os provérbios geralmente são considerados como regras sociais de conduta ou, mais simples, como estratégias de comportamento empregues num determinado contexto.

A estratégia subjacente é considerada como proverbial se uma grande faixa da sociedade a reconhece, usa ou adapta. Se for utilizada com frequência, podemos concluir que a mesma tem sucesso.

O uso frequente e o conhecimento do provérbio (de acordo com uma forma canónica) são duas faces da mesma moeda, o que nos leva a crer que o uso frequente pode ser comprovado pelo conhecimento geral e vice-versa.

No processo de recolha de provérbios, o seleccionador pede normalmente a amigos ou conhecidos para lhe indicarem frases que considerem como provérbios. Qualquer referência é incluída na selecção, como também cópias de exemplares oriundas de outras recolhas. Essa metodologia não é totalmente adequada aos critérios apresentados, uma vez que não basta a intuição individual para decidir sobre o conhecimento geral do provérbio. São necessários também dados estatísticos que o comprovem.

Nas três colectâneas Funk & Funk (2001a, 2001b, 2002), testaram-se inicialmente 25.000 exemplares, reunidos nas recolhas tradicionais mais importantes. Na primeira fase, foram excluídos cerca de 3.000 entradas que não satisfaziam a condição de texto mínimo. Numa segunda fase, verificou-se que, na ilha de São Miguel, só perto de 1/6 do referido *corpus* tinha uma taxa de reconhecimento de, pelo menos, 10%. Numa terceira fase, adicionou-se a esta fracção mais alguns elementos da recolha realizada nos Açores por Côrtes-Rodrigues (1982), atingindo, assim, cerca de 6.000 exemplares.

Este conjunto de provérbios foi apresentado a grupos de inquiridos em todas as ilhas do Arquipélago dos Açores e nas duas mais importantes zonas de emigração açoriana, nomeadamente na Califórnia e nos estados da Nova Inglaterra, na costa leste dos E.U.A. Foram identificados 1524 exemplares, que, numa ilha ou numa das zonas da emigração, reuniam uma taxa de reconhecimento de, pelo menos, 45%.

### **3. Integração e pré-processamento dos dados**

Nesta fase da metodologia CRISP-DM, incluem-se todas as actividades que permitem obter como *output* uma tabela de dados que será utilizada pelas aplicações de modelação na fase seguinte. Note-se que esta fase é fortemente dependente dos objectivos da modelação. Assim, para cada um dos objectivos enunciados na Introdução é necessário produzir uma tabela de dados distinta. Para o primeiro objectivo os tuplos da tabela têm de estar referenciados ao indivíduo que reconhece os provérbios. Pelo contrário, o segundo objectivo necessita de uma tabela de dados referenciada aos cerca de 1.500 provérbios utilizados nos inquéritos.

Segundo Chen (2001), para objectivos preditivos e de modelação, a utilização de demasiados dados pode ser efectivamente prejudicial ao ser necessário acomodar demasiadas excepções às regras gerais, contribuindo para problemas de sobreajustamento. Nesses casos, a utilização de amostras permite obter regras de decisão mais generalizáveis. Outro fundamento para reduzir o número de linhas de uma tabela de dados será a redução dos casos atípicos ou *outliers* que não contenham informação relevante e que conduzam a níveis de ruído nos dados mais elevados.

Estas considerações justificam a selecção de provérbios efectuada, uma vez que utilizar todas as expressões idiomáticas (cerca de 6.000) introduziria elevados níveis de ruído nos dados. Outra das operações de redução de dados que foram implementadas neste caso foi a redução no número de valores possíveis para campos nominais e ordinais, conhecida na literatura como *smooth a feature*. Em Mendes, Funk, & Funk (2007) descrevem-se os procedimentos utilizados tanto para a redução dos dados como para a limpeza dos mesmos.

A tabela de dados a ser utilizada pelas aplicações de modelação terá de corresponder a uma única relação (*flat file*) cuja chave será um identificador do indivíduo para o primeiro objectivo e um identificador do provérbio para o segundo objectivo. Partindo de uma base de dados relacional e normalizada, a criação destas vistas ou tabelas de dados corresponde a uma desnormalização. Assim, é necessário nas relações  $1 \rightarrow n$  resumir a informação na tabela do lado  $n$ , de molde a não se perder informação relevante.

Deste modo, surgem variáveis resultantes de agregação de dados como médias, desvios padrão e frequências relativas ou percentagens. No caso do segundo objectivo este processo é especialmente complexo uma vez que as localidades onde as pessoas viveram mais de 5 anos têm uma relação  $n \rightarrow n$  com os provérbios reconhecidos pelos mesmos indivíduos. Assim, foi necessário proceder a dois níveis de agregação, um primeiro para os indivíduos que reconhecem os provérbios e um segundo para as localidades onde esses indivíduos viveram mais de 5 anos. Tal foi conseguido com dois níveis de *queries*. Para este caso obteve-se uma tabela de dados compreendo apenas informação quantitativa, já que as variáveis nominais como as localidades onde viveram mais de 5 anos foram agregadas segundo frequências relativas resultando em percentagens do número de indivíduos que reconheceram um provérbio específico e que viveram mais de 5 anos numa localidade.

#### **4. Modelação e análise dos dados**

Nesta secção descrevem-se os modelos e as técnicas de estimação usadas para responder aos objectivos do projecto de extracção de conhecimento de bases de dados. Como é frequente em projectos de *data mining* aplicam-se várias técnicas distintas de estimação e comparam-se resultados.

##### **4.1. Caracterização de utilizadores de provérbios**

A tabela de dados resultante do pré-processamento inclui uma variável *target* ou dependente, correspondente à percentagem de provérbios reconhecidos por cada

inquirido. Trata-se de uma variável quantitativa com valores entre 1,85 e 97,3, média 44,6 e desvio padrão 19,0, correspondendo a um total de 902 casos. As variáveis potencialmente explicativas da percentagem de provérbios reconhecidos incluem a idade, o género, habilitações literárias, local da entrevista, número de locais onde viveu mais de 5 anos, local mais recente onde viveu, primeiro local onde viveu e se os locais onde viveu são todos considerados apenas rurais, apenas urbanos ou ambos.

Tendo em conta o objectivo de caracterização dos utilizadores de provérbios, técnicas estatísticas como a análise de regressão ou de aprendizagem automática, como as árvores de regressão, parecem adequadas. A análise de regressão tem dificuldade em lidar com variáveis em escalas não numéricas ou quantitativas, como é o caso das ordinais e nominais. Note-se que este tipo de escalas é muito comum na maioria de bases de dados. Por esta razão, as árvores de regressão são muitas vezes preferidas.

Os algoritmos de árvores de regressão incluem uma evolução do AID, o CHAID – *CHI-square Automatic Interaction Detection* (Biggs, de Ville & Suen, 1991 e Kass, 1980) e o CART – *Classification And Regression Trees* (Breiman, Friedman, *et al.*, 1984). O método utilizado por estes algoritmos consiste na divisão recursiva do conjunto de observações em subgrupos filhos, construindo uma árvore da raiz para as folhas. Em cada passo, o algoritmo determina uma regra de classificação, seleccionando uma variável e um ponto de corte nos valores dessa variável que minimize uma medida de impureza (CART) ou que maximize a distinção estatística dos filhos relativamente à variável dependente (CHAID). O objectivo consiste em obter divisões dos dados que permitam definir grupos tão homogéneos quanto possível, relativamente à variável dependente (*target variable*). Este processo é repetido até que uma regra de paragem seja atingida, como, por exemplo, a incapacidade de encontrar novas variáveis que permitam divisões dos dados estatisticamente significativas ou simplesmente um nível máximo de dimensão da árvore. Alguns algoritmos, como o CART, permitem ainda a poda da árvore, ao efectuarem uma revisão da árvore obtida e ao removerem ramos considerados pouco eficientes na previsão da variável dependente.

A qualidade dos resultados está associada a factores como o número de observações, de variáveis disponíveis, graus de liberdade, e às técnicas de amostragem utilizadas. Estes métodos são especialmente adequados quando é necessário utilizar um elevado número de observações e variáveis explicativas em escalas de medida não quantitativas. Note-se que estes algoritmos não garantem a optimalidade das soluções, dado que são heurísticos. No entanto, a utilização de diversos algoritmos sobre os mesmos dados e com parametrizações distintas permite obter um número elevado de árvores que devem ser posteriormente comparadas e analisadas.

No caso presente, utilizaram-se os dois algoritmos anteriormente descritos com diferentes critérios de paragem, com poda e sem poda, obtendo-se árvores muito semelhantes. As árvores obtidas conduzem basicamente às mesmas conclusões quanto à importância das variáveis explicativas e ordem pela qual impõem partições aos dados. Destas árvores é possível concluir que os inquiridos em ilhas menores ou destinos de emigração conhecem menos provérbios, resultante da primeira partição segundo o local onde foi efectuada a entrevista.

Verifica-se, igualmente, uma correlação significativa entre a idade do informante e o grau de reconhecimento de provérbios, para aqueles que foram inquiridos em ilhas de maior dimensão. Esta correlação foi confirmada calculando o coeficiente de correlação de Pearson entre as duas variáveis numéricas para a totalidade dos dados.

#### **4.2. Utilização de objectos simbólicos para classificação**

Numa tabela de dados simbólicos as linhas correspondem a  $N$  objectos simbólicos e as colunas a  $p$  variáveis simbólicas  $\{Y_1, \dots, Y_j, \dots, Y_p\}$ , as quais podem assumir valores tais como subconjuntos de categorias, intervalos da recta real, ou distribuições de frequências (Diday (2000)). As tabelas de dados simbólicos podem, pois, descrever dados heterogéneos e as suas células podem não conter apenas um único valor, como seria de esperar da primeira forma normal, mas diversos valores que podem estar ponderados e ligados por regras lógicas e por taxonomias. Surge, assim, a necessidade de estender os métodos da análise de dados clássica ao caso de dados de natureza complexa, também designados por dados simbólicos. A Análise de Dados Simbólicos (ADS) define-se, assim, como uma extensão da análise de dados clássica a essas tabelas (Diday (1988)).

A classificação de dados simbólicos pode basear-se em medidas de comparação entre elementos, clássicas ou probabilísticas, usando algoritmos hierárquicos aglomerativos clássicos ou probabilísticos (Gowda & Diday, 1991). Em particular, foi utilizada a abordagem probabilística de Análise Classificatória Hierárquica Ascendente (ACHA) designada por metodologia VL - Validade da Ligação (Sousa, 2005; Bacelar-Nicolau, 2000).

Os dados provenientes da base de dados de provérbios são intrinsecamente complexos, uma vez que envolvem três entidades com relações  $I \rightarrow n$ , nomeadamente: os provérbios, os inquiridos que os reconheceram e as localidades onde viveram mais de 5 anos. Assim, as técnicas para análise de objectos simbólicos implementadas no software SODAS são particularmente adequadas. Deste modo, utilizam-se neste trabalho métodos de Análise Classificatória Hierárquica Ascendente (ACHA) das unidades de dados simbólicos utilizando o coeficiente de afinidade generalizado ponderado (Bacelar-Nicolau, 2000) e critérios de agregação clássicos e probabilísticos (Nicolau & Bacelar-Nicolau, 1999; Bacelar-Nicolau, 1988).

## **5. Conclusões e conhecimento consolidado**

A aplicação da metodologia CRISP-DM tem-se revelado útil e prática, em especial no planeamento das actividades mais interdisciplinares, permitindo a definição de um esquema e linguagem comuns. Assim, tanto na fase de compreensão do domínio e dos dados como na fase de avaliação dos resultados dos modelos utilizaram-se reuniões e *workshops* de discussão. O projecto descrito neste artigo tem envolvido especialistas no domínio da paremiologia, da informática e vários especialistas na utilização de diferentes tipos de modelos. Foram implementadas regras de colaboração, de que é exemplo o presente artigo, redigido por todos os intervenientes.

Os resultados obtidos a partir da aplicação das diferentes técnicas de modelação foram confrontados com conhecimento de domínio paremiológico para validação dos resultados e consolidação do conhecimento que representam. Deste modo, confirmaram-se as suspeitas da relação significativa e positiva entre a competência proverbial e a idade, nomeadamente, o maior grau de reconhecimento de provérbios para inquiridos com idades superiores a 40-45 anos.

Confirmou-se, igualmente, que com excepção das ilhas do Corvo e de Santa Maria, cujo reportório de expressões proverbiais não foi explicitamente incluído nos dados utilizados, as restantes ilhas do arquipélago apresentam uma taxa de reconhecimento superior às zonas de emigração. Esta observação pode reflectir o facto de, nestes últimos locais, a assimilação de uma cultura anglo-saxónica reduzir o reportório proverbial em Português, apesar de existir evidência de que alguns provérbios são preservados como relíquias culturais.

Questões muito frequentemente debatidas na literatura paremiológica surgiram, neste estudo, com resultados inesperados, como a correlação entre as habilitações literárias e a extensão do reportório, que, ao contrário do esperado, se revelou inexistente para a quase totalidade do universo. Mesmo os indivíduos iletrados não revelaram reconhecer mais provérbios do que os restantes. A única excepção surgiu com os inquiridos nos destinos de emigração, onde a correlação se revelou negativa. Este resultado levanta a hipótese de existirem diferenças entre os inquiridos com formação no arquipélago e os que frequentaram o ensino no destino de emigração, hipótese que não pode ser avaliada com os dados disponíveis.

Da mesma forma, surpreendente foi a verificação de que não existe uma correlação entre a ruralidade das localidades onde o informante habitou mais de 5 anos e o grau de reconhecimento de provérbios. Tal era esperado, uma vez que alguns autores consideram os provérbios como elementos da cultura rural. Uma hipótese para explicar esta divergência seria os fluxos populacionais entre os ambientes rurais e urbanos, actualmente muito intensos.

Perante estes resultados promissores, pretende-se dar continuidade ao presente com a utilização de outros modelos e métodos, como é o caso dos métodos utilizados em *market basket* para associar provérbios que são reconhecidos em conjunto e a utilização de objectos simbólicos na análise de dados e extracção de conhecimento.

## Referências

- Bacelar-Nicolau, 1988. Two probabilistic models for classification of variables in frequency tables. In Bock, H. H. (Ed.) *Classification and Related Methods of Data Analysis*. Amsterdam: North Holland. 181-186.
- Bacelar-Nicolau, 2000 The affinity coefficient. In Bock, H.-H. & Diday, E. (Eds.) *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data*. Heidelberg: Springer-Verlag. 160-165. ISBN: 3-540-66619-2.
- Berry, M. J. A. & Linoff, G. (1997). *Data Mining Techniques: For marketing, sales, and customer support*. USA: John Wiley & Sons. ISBN: 0-471-17980-9.
- Biggs, D.B. de Ville & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, 18, 49-62.
- Breiman, Leo; Friedman, Jerome H.; Olshen, Richard A. e Stone, Charles J. (1984). *Classification and Regression Trees*. California, USA: Wadsworth International. ISBN: 0-534-98053-8.
- Clifton, C. & Thuraisingham, B. (2001). Emerging standards for data mining. *Computer Standards and Interfaces*, 23 (3 July), 187-193.
- Chen, Zhengxin (2001). *Intelligent Data Warehousing: From data preparation to data mining*. Boca Raton, USA: CRC Press. ISBN: 0-849-31204-3.
- Côrtes-Rodrigues, A. (1982). *Adagiário Popular Açoriano*. Angra do Heroísmo: Antília, Secretaria Regional da Educação e Cultura. Volumes 1 e 2.
- Diday (1988). The symbolic approach in clustering and related methods of data analysis: The basic choices. In Bock, H.H. (Ed.) *Classification and Related Methods of Data Analysis*. Amsterdam: North Holland. 673-683.
- Diday (2000). Symbolic data analysis and the SODAS project: Purpose, history, perspective. In Bock, H.-H. & Diday, E. (Eds.) *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data*. Heidelberg: Springer-Verlag. 1-23. ISBN: 3-540-66619-2.
- Funk, G. & Funk M. (2001a). *Pérolas da Sabedoria Popular Portuguesa – Os Provérbios de S. Miguel*. Lisboa: Salamandra Edições. ISBN: 972-6891-80-9.

- Funk, G. & Funk M. (2001b). *Pérolas da Sabedoria Popular Portuguesa – Os Provérbios Açorianos nos EUA*. Lisboa: Salamandra Edições. ISBN: 972-689-189-2.
- Funk, G. & Funk M. (2002). *Pérolas da Sabedoria Popular Portuguesa – Os Provérbios das Ilhas do Grupo Central dos Açores (Faial, Graciosa, Pico, São Jorge e Terceira)*. Lisboa: Salamandra Edições. ISBN: 972-689-203-1.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (Eds.) (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, USA: MIT Press. ISBN: 0-262-56097-6.
- Gowda, K. C. & Diday, E. (1991). Symbolic clustering using a new dissimilarity measure. *Pattern Recognition*, 24 (6), 567-578.
- Hand, D. J., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*. Cambridge, USA: MIT Press. ISBN: 0-262-08290-X.
- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29 (2), 119-127.
- Lavrač, N., Motoda, H., Fawcett, T., Holte, R., Langley, P. & Adriaans, P. (2004). Introduction: Lessons learned from data mining applications and collaborative problem solving. *Machine Learning*, 57 (1-2), 13-34.
- Mendes, A., Funk, G., & Funk, M. (2007): Extrair Conhecimento de Provérbios. Submetido para publicação em Temas em Métodos Quantitativos VI. Lisboa: Sílabo.
- Nicolau & Bacelar-Nicolau, 1999. Clustering symbolic objects associated to frequency or probability laws by the weighted affinity coefficient. In Bacelar-Nicolau, H., Nicolau, F. C. & Janssen, J. (Eds.) *Applied Stochastic Models and Data Analysis: Quantitative methods in business and industry society*, 9<sup>a</sup> ed. Lisboa: INE. 155-158. ISBN: 972-98391-0-7.
- Sousa, Á. (2005). Contribuições à Metodologia VL e Índices de Validação para Dados de Natureza Complexa. Ponta Delgada, Portugal: Univ. dos Açores.