IDENTIFICAÇÃO DO NÚMERO MÍNIMO-IDEAL DE INDICADORES ECONÔMICOS POR GERAÇÃO DE AGRUPAMENTOS DE SÉRIES TEMPORAIS

Capitani, L.M.
Divisão de Engenharia Mecânica e Aeronáutica
Instituto Tecnológico de Aeronáutica-ITA – email: leandro.capitani@gmail.com

Scarpel, R.A.
Divisão de Engenharia Mecânica e Aeronáutica
Instituto Tecnológico de Aeronáutica-ITA – email: rodrigo@ita.br

Dotti, E.
Divisão de Engenharia Mecânica e Aeronáutica
Instituto Tecnológico de Aeronáutica-ITA – email: dotti@ita.br

Resumo

O foco deste trabalho foi desenvolver um método baseado em agrupamentos de séries temporais que fosse capaz de identificar grupos de indicadores econômicos mostrando quais eram, de fato, relevantes para classificar a situação econômica atual do país . Escolheu-se como métrica de dissimilaridade a DTW (*Dynamic Time Warping*) e como método de geração de agrupamentos foi adotado o método hierárquico do centróide que gerou agrupamentos significativamente interpretáveis. Como resultado percebeu-se que apenas quatro índices são suficientes para determinar a situação econômica do país, e eles são: Índice de Produção intensiva em capital, Índice de consumo, Índice de Insumos para a Produção e Índice de produção para consumo.
Palavras chave: geração de agrupamentos de séries temporais; indicadores econômicos; DTW; *dynamic time warping*.

Abstract

The focus of this work was to develop a method based on clustering time series that was capable to identify groups of economical indicators in fact showing which they were, relevant to classify the current economical situation of the country. It was chosen as metric of dissimilarity the DTW (Dynamic Time Warping) and as method of generation of groupings it was adopted the hierarchical method of the centróide that generated groupings significantly you interpreted. As result was noticed that only four indexes are enough to determine the economical situation of the country, and they are: Index of intensive Production in capital, consumption Index, Index of Inputs for the Production and production Index for consumption.
Keywords: *clustering of time series data; economic indicators; dynamic time warping; DTW*

1.0 Introdução

Métodos de geração de agrupamentos vêm ganhando espaço em mineração de dados, sendo aplicados a diversas áreas do conhecimento, desde medicina em ressonâncias magnéticas até em geologia em previsão de terremotos.
Existem três grandes grupos de métodos de geração de agrupamentos: os que trabalham os dados crus, sem modificar as series, lembrando que padronização não é considerada uma modificação, existe um segundo grande grupo que constrói agrupamentos baseados em características e o terceiro apóia a construção dos agrupamentos em modelos como ARMA, ARIMA entre outros. Nesse trabalho é aplicado o conceito do primeiro grupo, os que trabalham com os dados não padronizados.
No que diz respeito a métodos capazes de gerar agrupamentos baseado em dados não padronizados, muitos fatores podem mudar de um método para o outro, cada método é construído par atender séries de tamanhos iguais ou desiguais, alguns métodos como aqueles que se utilizam da dissimilaridade por DTW permitem tanto serie de tamanhos iguais quanto desiguais. A escolha da medida de dissimilaridade é um fator que diferencia os modelos, no *paper* do Liao ele apresenta nove medidas de dissimilaridade embora exista um numero ainda maior. Os métodos de geração de algoritmos também variam, podendo ser usado métodos hierárquicos, fuzzy, c-médias, k-médias e redes neurais. Por fim, o critério de avaliação de homogeneidade do cluster é o ultimo fator que diferencia um modelo do outro, alguns se utilizam deles e outros não. Nota-se uma grande gama de possibilidades para a construção de modelos, ainda que só falasse de modelos baseados nos dados crus.
Liao faz uma revisão de métodos de geração de agrupamentos em seu *paper* e mostra as diversas aplicações que estão sendo desenvolvidas utilizando essas técnicas. A tabela que apresenta as aplicações utilizando dados crus foi retirada do Liao e pode ser vista abaixo.

2.0 Motivação

De acordo com Caiado et al. (2006), em economia, pode-se estar interessado em classificar a situação econômica de um país apenas observando alguns indicadores de séries temporais, tal como o Produto Interno Bruto(PIB), Gastos com Investimentos, Gastos com Consumo, ou até mesmo, Taxa de Desemprego. Contudo escolher o número mínimo-ideal de indicadores e escolher indicadores que não sejam redundantes não é uma tarefa trivial.

3.0 Objetivos

O trabalho tem como objetivos principais Identificar quais indicadores são relevantes para classificação da situação do país e identificar o número mínimo-ideal de indicadores. O trabalho tem como objetivos específicos: descobrir e visualizar padrões em Séries Temporais, identificar Séries semelhantes (seqüências que combinam), detectar Anomalias e gerar agrupamentos por métodos hierárquicos.

4.0 Metodologia

Para iniciar o estudo adquiriu-se a base de dados, foram coletadas trinta e sete séries temporais do "Sistema Gerenciador de Séries Temporais (SGS)" do Banco Central, no site https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries.

No SGS cada série temporal possui um número associado a ela, os números e as respectivas séries coletadas são apresentados abaixo entre parênteses, elas estão também classificadas por tipo.
Séries: Indicadores de antecedentes para nível de atividade.
Produção Industrial: S1- Bens de Capital (11.067); S2 - Bens Intermediários (11.068); S3 - Bens de Consumo Duráveis (11.070); S4 - Bens de Consumo Não-Duráveis e Semiduráveis (11.071).
Produto Interno Bruto: S10 - Mensal (4.380); S11 - Acumulado em 12 meses (4.382).
Volume de Vendas no Varejo: S29 - Combustíveis e Lubrificantes (1.483); S33 - Hipermercados e Supermercados (1.561); S30 - Tecido, Vestuário e Calçado (1.509); S31 - Móveis e Eletrodomésticos (1.522); S32 - Automóveis, Motocicletas, Peças (1.548).
Consumo de Energia Elétrica: S34 - Comercial (1.402); S35 - Residencial (1.403); S36 - Industrial (1.404); S37 - Outros (1.405).
Taxa de Desemprego (10.777).
Tempo Médio de Procura de Emprego: S6 - Até 30 dias (10.796); S7 - De 31 dias até 6 meses (10.797); S8 - De 7 meses a 11 meses (10.798); S9 - Mais de 1 ano (10.799).
Exportações: S22 - Total (kg) (4.193); S23 - Produtos Básicos (kg) (4.194); S24 - Produtos Semifaturados (kg) (4.221); S25 - Produtos Manufaturados (kg) (4.248); S18 - Total (US$) (2.946); S19 - Produtos Básicos (US$) (2.947); S20 - Produtos Semifaturados (US$) (2.974); S21 - Produtos Manufaturados (US$) (3.001).
Importações: S12 - Total (US$) (3.034); S13 - Matérias-Primas e Produtos Intermediários (US$) (3.050); S14 - Bens de Capital (US$) (3.062); S15 - Total (kg) (4.281); S16 - Matérias-Primas e Produtos Intermediários (kg) (4.297); S17 - Bens de Capital (kg) (4.309).
Investimentos Diretos: S26 - Total Líquido (2.851); S28 - Estrangeiro Direto (2.860); S27 - Brasileiro Direto (2.852).
O processo de geração de agrupamentos consiste em 4 etapas:

1. Adquirir a base de dados.
2. Escolher a métrica de dissimilaridade.
3. Calcular as coordenadas das series em um espaço bidimensional baseado na dissimilaridade.
4. Escolher o método hierárquico de geração de agrupamentos mais conveniente.

Como o propósito do trabalho é conseguir classificar a situação econômica do país, sempre no dia de hoje, não faz muito sentido trabalhar com séries de tamanhos desiguais apesar do fato da DTW permitir que as séries tenham tamanhos distintos entre si. Assim, as séries estudadas possuem 91 medidas mensais (*time points*) de out/01 até abr/09, não possuindo valores faltantes (*missing values*), como mostra a tabela abaixo.

Quantidade de séries estudadas	37
Comprimento	Equal
Time points de cada série	91 (meses)
Primeiro *time point*	Outubro de 2001
Último *time point*	Abril de 2009
Valores faltantes	Nenhum
Intervalo entre *time points* adjacentes	1 mês

Tabela 4.1: descrição do banco de dados.

Nota: É importante ressaltar que todas as séries foram padronizadas para que tivessem média zero e desvio padrão 1. Isso é importante no caso de geração de agrupamentos de séries temporais, pois algumas séries são índices, outras tem como unidade R$, outras são em Gigawatts (GW).

A etapa 2 é escolher uma métrica de dissimiliridade capaz de gerar clusters interpretáveis, não existe uma regra bem definida sobre qual métrica é mais indicada para esse ou aquele grupo de séries, geralmente a escolha faz-se após a experimentação.
No caso do presente trabalho foram testadas três métricas:

1. **STS- *short time series*** - ela se baseia parte do pressuposto que entre um *time point* e outro as séries podem ser escritas como segmentos de retas. Assim, essa métrica compara a inclinação desses segmentos para medir a dissimilaridade.
2. **Distância baseado em um índice de correlação cruzada** - através de uma correlação cruzada de duas séries é medido a dissimilaridade entre elas.
3. **DTW**- é um método construído para bioinformática muito eficiente para o calculo da dissimilaridade entre duas séries.

Será feito a descrição apenas do método do DTW, isso, pois foi o que gerou resultados mais relevantes. As métricas 1 e 2 não formaram agrupamentos muito nítidos.
A STS apresentou deficiências quando aplicada a séries temporais econômicas, pois ela considera que séries que possuem movimentos opostos (quando uma cresce a outra diminui) possuem a maior dissimilaridade, contudo isso não é verossímil na maior parte das series econômicas, pois séries com comportamento absolutamente opostos são fenômenos regidos pelas mesmas leis, mas que reagem de maneira inversa as variações, devendo ter baixa dissimilaridade e não alta como calcula o STS.

4.1 Descrição do DTW (Dynamic Time Warping)

O DTW (*dynamic time warping*) ou alinhamento temporal dinâmico é a generalização dos algoritmos clássicos para comparar sequências discretas com sequências de valores contínuos, muito usado também para comparar sequências de DNA/RNA . Dado duas séries temporais, $Q=q1, q2, \ldots, qi, \ldots, qn$ e $R=r1, r2, \ldots, rj, \ldots, rm$, DTW alinha as duas séries minimizando suas diferenças. Para esse fim, é construída uma matriz $n \times m$ onde o elemento (i , j) contem a distância $d\left(q_i, r_j\right)$ entre dois pontos q_i e r_j . A distância euclidiana é normalmente usada.
Depois de construída a matriz de distâncias, é necessário obter-se o caminho *warping (warping path*), afinal a dissimilaridade depende disso.
O caminho *warping*, $W=w1,w2, \ldots, wk, \ldots,$ wk onde $\max(m,n) \leq K \leq m+n-1$, é o conjunto dos elementos da matriz de distancia que satisfaz três condições: condição de contorno , continuidade e monotonicidade.
A condição de contorno requer que o caminho path inicie-se em $w_1 = (1,1)$ e termina em $w_k = (m,n)$. Já a condição de continuidade restringe que os passos dados na matriz de distancia para construir o caminho *warping* sejam dados para células adjacentes. Por fim, a condição de monotonicidade força que os pontos no caminho *warping* sejam monotonicamente espaçados no tempo.
Levado isso em conta, o caminho *warping* é aquele que minimiza a distância entre as duas séries temporais de interesse. Matematicamente temos:

$$d_{DTW} = \min \frac{\sum_{k=1}^{K} w_k}{K}$$

Programação dinâmica pode ser usada de maneira muito eficiente para encontrar o caminho *warping*, utilizando o algoritmo abaixo.

dcum(i, j) = d(qi, rj) + min{dcum(i − 1, j − 1), dcum(i − 1, j), dcum(i, j − 1)}.

O algoritmo computacional para calcular o DTW foi feito no software R .

4.2 Obtenção da matriz de dissimilaridade

O método DTW é uma maneira de obter a dissimilaridade entre duas séries temporais, contudo para que a geração de agrupamentos seja possível necessita-se da dissimilaridade de uma com todas e todas com uma. Logo, precisa-se de uma matriz de dissimilaridades, que no caso desse estudo contenha 37×37 elementos, uma vez que temos 37 series.
O algoritmo para o calculo da matriz de dissimilaridade foi feito no software R .
As medidas de dissimilaridade possuem algumas propriedades, elas são:
1. $d(x,y) \geq 0$
2. Simetria: $d(x,y) = d(y,x)$
3. Se $d(x,y) \neq 0$ então $x \neq y$
4. Se $x = y$, então $d(x,y) = 0$

4.3 Geração do MDS (*multi dimensional scale*)

Tendo como entrada a matriz de dissimilaridade, construiu-se um código no R capaz de gerar coordenadas e *plotar* um gráfico bidimensional, onde pontos próximos são mais similares que pontos distantes. O MDS é uma boa ferramenta de redução de dimensão e visualização de cluster, pois é uma foto do melhor plano bidimensional para visualizar as diferenças entre as séries.
Existem duas formas de MDS: métrico (assume-se que as dissimilaridades são proporcionais às distâncias Euclidianas) e não-métrico (assume-se que as dissimilaridades são relacionadas às distâncias Euclidianas por alguma transformação monotônica).

4.3.1 Equações do MDS

É construída uma função chamada STRESS que deve ser minimizada, ela é responsável pela decomposição das dissimilaridades.

$$Min\ STRESS = \left[\sum_i \sum_j \left(d_{ij} - D_{ij} \right)^2 \Big/ \sum_i \sum_j d_{ij}^{\ 2} \right]$$

Em que Dij é a dissimilaridade entre as observações i e j e dij é a distância euclidiana entre as coordenadas das observações i e j no espaço k-dimensional, calculada por:

$$d_{ij} = \sqrt{\sum_{v=1}^{k<p} \left(x_{iv} - x_{jv} \right)^2}$$

4.4 Geração de agrupamentos através de métodos hierárquicos.

A geração de *clusters* não precisa ser feita por métodos hierárquicos, contudo eles têm uma vantagem que é a formação de um dendograma que permite visualizar agrupamentos naturais. Os dendogramas são construídos baseados na matriz de dissimilaridade, que no caso desse trabalho, foi obtida utilizando DTW.

5.0 Análises e resultados

5.1 MDS bidimensional

Como foi dito, a partir da matriz de dissimilaridade construída com DTW foi possível gerar coordenadas espaciais para cada uma das séries temporais, as coordenadas plotadas em um gráfico bidimensional é apresentado abaixo.

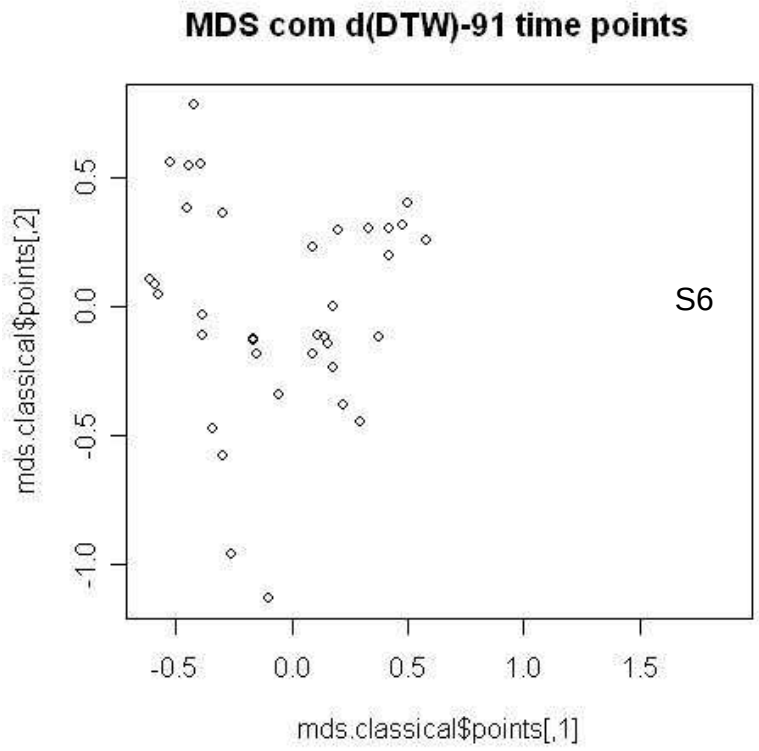

Figura 5.1: MDS construído com DTW das 37 séries temporais

Os gráficos acima têm a seguinte propriedade: séries temporais que estão próximas são mais semelhantes entre si, que séries temporais que estão distantes.
Sem a necessidade de nenhum método adicional notamos que a série S6 comporta-se como um *outlier* sendo distante de todas as outras séries.

S6	10796 - Distribuição de desocupados segundo o tempo de procura de emprego - Até 30 dias - %

Explicação do porquê a S6 é aparentemente um *outlier*: É de pensar ao olhar essa série que se o desemprego aumenta isso deveria refletir uma situação ruim do país, contudo ela aponta pessoas desempregadas até 30 dias, e se uma pessoa não passa mais do que trinta dias desempregada é de se imaginar que o país esteja em boa situação econômica , assim essas duas possíveis interpretações contraditórias se refletem na matriz de dissimilaridade e mostra que é impossível prever a situação do pais olhando para a porcentagem de desempregados até 30 dias.

5.2 Comparação entre a métrica DTW e a métrica STS

Como foi dito nos itens anteriores a DTW mostrou-se uma métrica mais indicada para gerar agrupamentos, contudo anteriormente não foi dito o porquê ela é mais indicada. Uma boa métrica

permite que a dissimilaridade ressalte as diferenças, já uma métrica inadequada não consegue deixar claras as diferenças entre as diversas séries.
Far-se-á agora uma comparação entre a DTW e a STS, onde será nítido que a DTW é uma métrica mais adequada para o estudo em questão.
Não foi feito uma grande explanação sobre a métrica STS, mas isso não é importante, pois o foco desse item não está sobre as propriedades da STS e sim sobre a comparação entre duas métricas. As conclusões das comparações devem advir dos resultados e não de suas propriedades especificas.

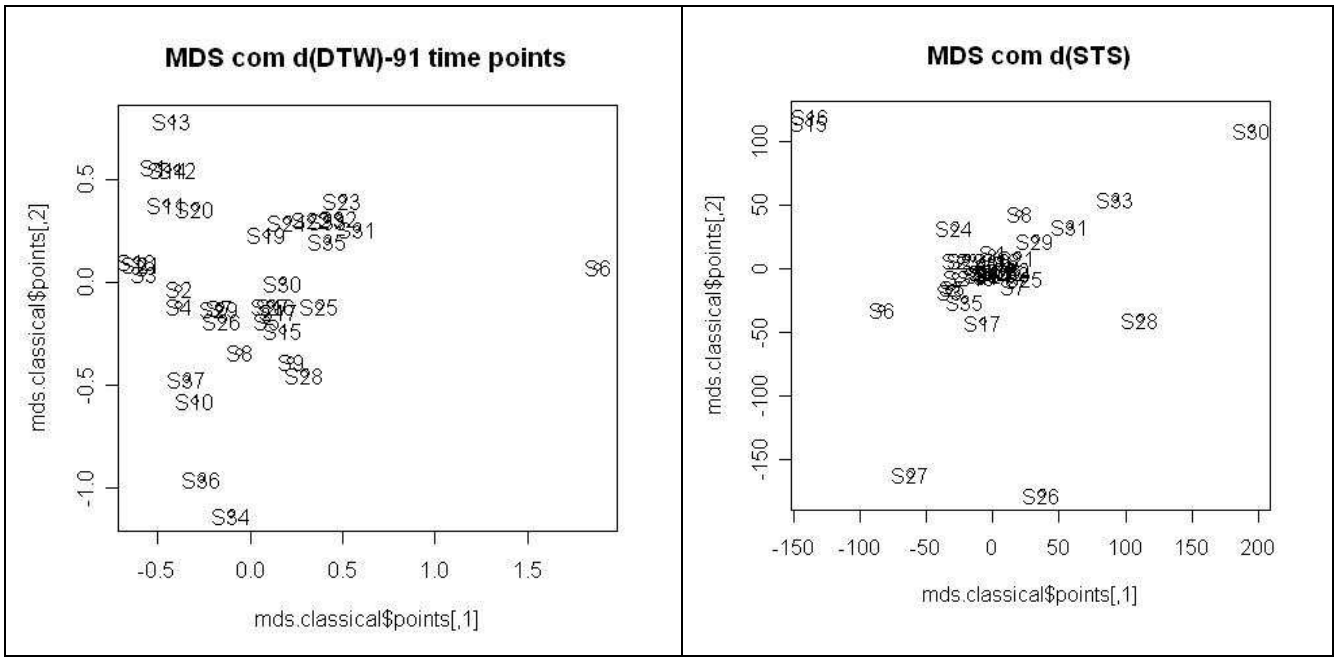

Figura 5.3: Comparação entre os MDS formados por DTW e STS

Note que o MDS formado com a métrica STS não permite visualizar claramente as diferenças entre varias séries, tão pouco é possível enxergar facilmente a existência de clusters naturais. Essa visualização lado a lado permite justificar que a DTW é uma métrica mais aconselhável que a STS para a construção de agrupamentos das séries estudadas.

5.3 Geração dos dendogramas por métodos hierárquicos.

Foi escolhido o método do centróide para a construção dos agrupamentos devido a sua robustez e capacidade de gerar clusters significativamente interpretativos. O código fonte do programa que gerou o dendograma foi executado no software R .
Esse dendograma reafirma o que havia sido dito sobre a série S6 ser um *outlier*, note que a altura de ligação é proporcional a diferença/dissimilaridade.
Pode-se observar a formação de sete agrupamentos naturais sendo que um deles contém apenas a série S6 que tem características de *outlier*. No gráfico abaixo podemos ver o dendograma formado pelo método do centróide e o realce dos *clusters* naturais é evidenciado.

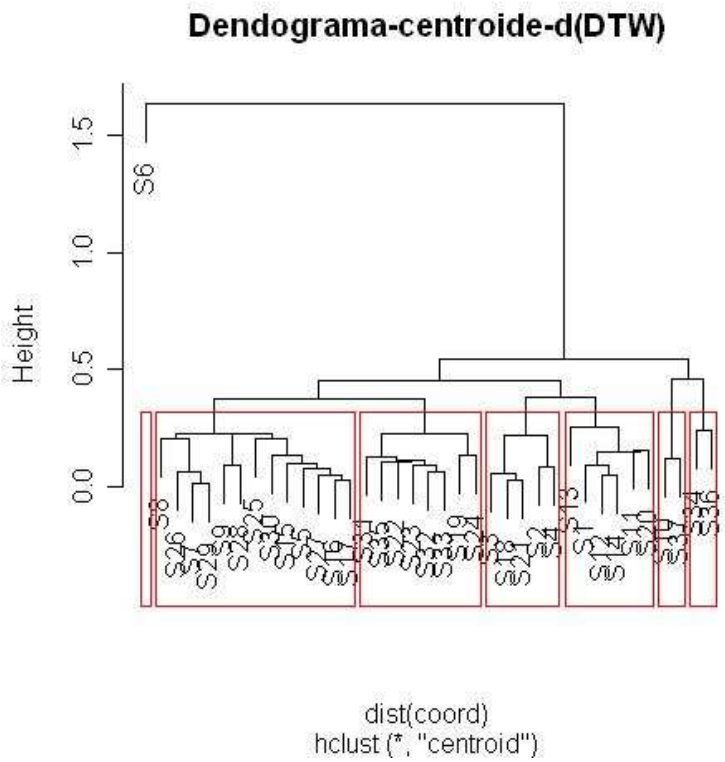

Figura 5.2 : realce dos sete agrupamentos naturais formados pelo método do centróide utilizando a matriz de dissimilaridade construída por DTW

É possível unir as informações do dendograma com as informações do MDS gerando o gráfico abaixo:

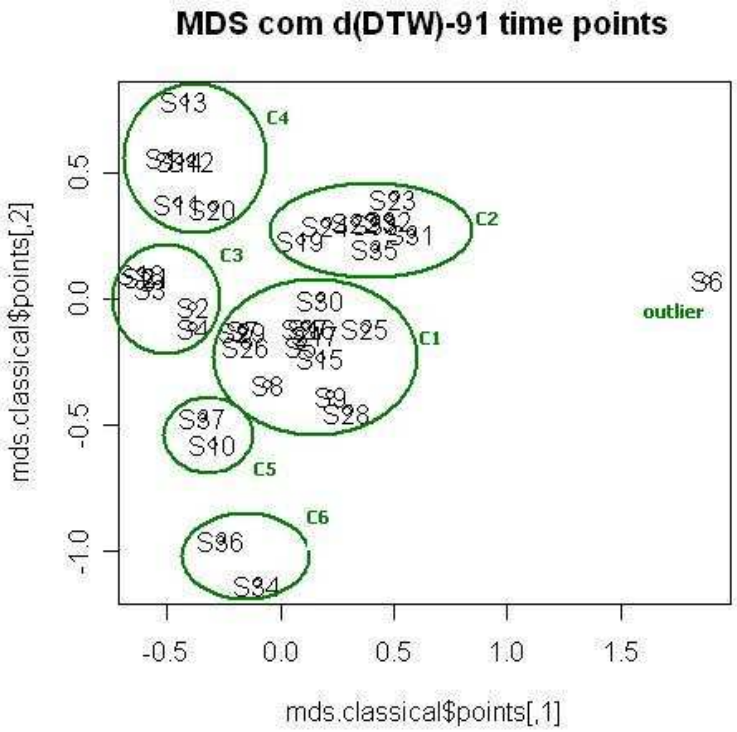

Figura 5.3: realce dos clusters formados pelo método do centróide no gráfico do MDS

5.4 Resultados da geração de agrupamentos

	Série e código da serie no SGS
Cluster 1	S8(10798); S26(2851); S7(10797); S29(1483); S9(10799); S28(2860); S25(4248); S30(1509); S15(4281); S5(10777); S27(2852); S16(4297); S17(4309)
Cluster 2	S31(1522); S35(1403); S22(4193); S23(4194); S32(1538); S33(1561); S19(2947); S24(4221)
Cluster 3	S3(11070); S18(2946); S21(3001); S2(11068); S4(11071)
Cluster 4	S13(3050); S1(11067); S12(3034); S14(3062); S11(4382); S20(2974)
Cluster 5	S10(4380); S37(1405)
Cluster 6	S34(1402); S36(1404)
Cluster 7	S6(10796)

Tabela 5.1: Séries que compõem cada um dos agrupamentos/ entre parênteses o código da serie no SGS

5.5 Exemplo de como extrair conclusões dos clusters

É possível extrair muitas conclusões tendo em mãos a relação das séries que formam os clusters, tantas que seria inviável listar todas. As conclusões dependem do tipo de enfoque, do objetivo da análise e da visão particular do analista. Extrair conclusões não é uma ciência exata, nem tão pouco há como saber, a priori, quantas conclusões são possíveis de se extrair de um dado conjunto de clusters.
Aqui será apresentada apenas uma conclusão para reforçar a robustez do modelo adquirido.
Conclusão 1: Ao se olhar o **cluster C1** tem-se a relação de S26,S27,S28 que referem-se a Investimentos diretos, com as séries S7, S8, S9, S5 que remetem a taxas de desemprego de períodos mais longos, isso revela que a falta de investimento direto impacta sobre a qualidade de vida das cidades e sobre os indivíduos que são mão de obra mais qualificada gerando taxa de desemprego mais elevadas no longo prazo.

5.6 Índices obtidos capazes de classificar a situação econômica atual do país

Desde o inicio do trabalho o objetivo da geração de agrupamentos das séries temporais foi identificar alguns poucos índices capazes de classificar a situação econômica do país no momento atual (dia de hoje), conseguir esses índices permite deixar de usar 37 séries e passar a usar poucos indicadores, uma vez que, séries que estão no mesmo cluster são similares e reagem de maneira similares aos fenômenos econômicos. Levando o que foi dito em conta, extraímos **quatro indicadores relevantes**, eles são citados e descritos abaixo:

Cluster	Nome do Indicador
C1	Produção Intensiva em Capital
C2	Consumo
C3	Produção para Consumo

C4	Insumos para Produção

Tabela 5.2: relaciona o cluster a sua interpretação

Dar-se-á uma breve explicação sobre os indicadores:
Produção Intensiva em Capital - cluster C1: nota-se claramente que esse cluster relaciona os investimentos diretos aos tipos de mão de obra especializada através das taxas de desemprego, em especial a taxa de desemprego em regiões metropolitanas. O cluster reúne também importações de bens de capital, esse tipo de importação é típico de empresas intensivas em capital. Como o objetivo é reduzir o número de séries à serem analisadas, uma boa série capaz de substituir as treze séries contidas no cluster C1, para o efeito de verificar a atividade da produção intensiva em capital é a S5 - (Taxa de desemprego – Região metropolitana – Brasil (na semana) %).
A S5 é uma boa candidata a ser indicador de atividade produtiva intensivas em capital, pois para as empresas produzirem mais, necessitam de mais mão de obra especializada encontrada nas cidades.
Indicador de Consumo - Cluster C2: Esse *cluster* relaciona índice de volume de vendas no varejo a exportações. Obviamente temos a seguinte relação diretamente proporcional: um aumento nas exportações é devido ao aumento da produção dos produtos destinados a exportação, que traz consigo um aumento dos postos de trabalho, que traz consigo um aumento do poder de compra da população que se reflete em um maior consumo no varejo. Para esse cluster o consumo de energia elétrica residencial (S35) seria uma boa escolha para representar o consumo, nota-se que o consumo de eletricidade pelas casas é fortemente afetado por variações na renda familiar.
Indicador de Produção para Consumo – cluster C3: Esse *cluster* contém essencialmente indicadores de produção de bens de consumo, possui também exportações do mesmo tipo de bens.
Insumos para Produção – cluster C4: Esse *cluster* associa, principalmente, as importações e o indicador de produção de bens de capital, isso pode ser entendido como insumos para a produção, sem tais coisas a indústria não é capaz de produzir.
O cluster C5 essencialmente possui o PIB que é a métrica de avaliar o desempenho de um país, ou seja, um país cresce o que o seu PIB (produto interno bruto) cresce.
O cluster C6 mostra que tanto o consumo de energia elétrica comercial quanto industrial não são bons indicadores para classificar a situação do país, uma explicação possível para esse fato é de que o gasto de energia para esses setores é bastante inelástico com a demanda, por exemplo, um comércio não gasta menos luz se, em um dia, tem o tráfego de clientes reduzidos pela metade, pois, é claro, os clientes que foram até o comércio em questão necessitam de um nível adequado de serviço, não podendo ser atendidos no escuro.

6.0 Conclusão

Concluiu-se que o objetivo do trabalho de gerar agrupamentos hierárquicos com a finalidade de obter um número restrito de indicadores capazes de classificar a situação econômica do país foi alcançado. Obtiveram-se quatro indicadores: Produção Intensiva em Capital, Consumo, Produção para Consumo, Insumos para Produção; esses indicadores substituem a análise de 37 séries temporais para classificar a situação atual do país.
Justificou-se durante o trabalho que existem métricas de dissimilaridade mais eficientes que outras e que a escolhida foi a DTW (*Dynamic Time Warping*), por gerar clusters mais bem definidos e mais interpretáveis. O MDS por sua vez, mostrou seu poder de reduzir as dimensões do espaço de atributos e gerar coordenadas para as séries temporais onde séries mais próximas eram mais similares que as mais distantes, a relação de proximidade entre as séries pôde ser confirmada pelo dendograma.

O trabalho permitiu também constatar que os métodos de geração de agrupamentos de séries temporais são úteis para analistas, pois possibilitam revelar conclusões, antes difíceis de visualizar ou provar.

Bibliografia

T.W. Liao, Clustering of time series data – a survey, The Journal of the pattern recoginition society (2005).

Caiado, J. , Crato, N. e Peña, D. (2006). A periodogram – based metric for time series classification. Computational Statistics & Data Analysis, 50, 2668 – 2684.

Sistema Gerenciador de Séries Temporais – SGS https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries –- origem da base de dados - acessado dia 1 de setembro de 2009.

X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, P.Boesiger, A new correlation-based fuzzy logic clustering algorithm for fMRI, Mag. Resonance Med. 40 (1998) 249–260.

C.S. Möller-Levet, F. Klawonn, K.-H. Cho, O. Wolkenhauer, Fuzzy clustering of short time series and unevenly distributed sampling points, Proceedings of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany, August 28–30, 2003.

M. Kumar, N.R. Patel, J. Woo, Clustering seasonality patterns in the presence of errors, Proceedings of KDD ’02, Edmonton, Alberta, Canada.

K. Košmelj, V. Batagelj, Cross-sectional approach for clustering time varying data, J. Classification 7 (1990) 99–109.

T.W. Liao, B. Bolt, J. Forester, E. Hailman, C. Hansen, R.C. Kaste, J. O’May, Understanding and projecting the battle state, 23rd Army Science Conference, Orlando, FL, December 2–5, 2002.

T.W. Liao, Mining of vector time series by clustering,Working paper, 2005.