

7

Modelo Híbrido: Conjugação de Redes Neurais Artificiais e Análise de Fronteira Estocástica para Cálculo do X^0

7.1.

Classificação de Padrões: Encontrando Grupos de Similaridade

Para a definição de fronteiras de eficiência, um importante aspecto que precisa ser considerado na análise é o aspecto de similaridade entre as empresas distribuidoras de energia. Isto é, dadas as heterogeneidades deste mercado no Brasil, antes de se definir uma fronteira de eficiência que não leve em consideração diversas particularidades regionais e de operação das empresas, é plausível que se faça uma análise de reconhecimento ou classificação de padrões destas empresas. Propõe-se portanto que, posteriormente à identificação de grupos de similaridade, dentro de cada grupo, sejam estimadas as fronteiras de eficiência.

A metodologia de comparação escolhida foi a de montagem de famílias de empresas semelhantes, considerando variáveis de forte influência nos Custos Operacionais das distribuidoras, utilizando técnicas de análise estatística multivariada. Destas técnicas a escolhida foi a de análise de *cluster*.

A análise estatística multivariada²² que envolve técnicas como *cluster analysis*, *factor analysis*, *principal component analysis*, *discrimination analysis* entre outros, parte de uma matriz de dados com objetivos tais como:

- determinar o grau de associação entre conjuntos de variáveis;
- desenvolver métodos de interpretação visual (importante na fase inicial de exploração de dados);
- redução de dimensionalidade, isto é, diminuir o número de variáveis com perda mínima de informação;
- identificar um número reduzido de fatores que expliquem a maior parte da variação dos dados
- e determinar e caracterizar grupos de observações semelhantes.

²² Johnson & Wichern (1994).

Este último item caracteriza o que se denomina de análise de *cluster*. O objetivo é encontrar grupos “naturais”, ou seja, identificar grupos ou classes de objetos semelhantes entre si dentro de um conjunto de objetos. Obtém-se como resultado uma organização natural do banco de dados. As aplicações para este tipo de metodologia são as mais variadas, podendo-se citar: identificação de segmentos de mercado, compressão de informação (substituir dados originais por estatísticas descritivas de cada grupo, por exemplo); classificação de animais e plantas, classificação de tipos de personalidade, identificação de pontos discrepantes entre outras.

A análise de agrupamentos é aplicada então sobre uma tabela de dados multivariados X , como apresentada na figura 17, com o objetivo de redução do número de linhas desta matriz.

$$X = \begin{bmatrix} & VAR_1 & VAR_2 & . & . & . & VAR_k \\ OBJ_1 & x_{11} & x_{12} & . & . & . & x_{1k} \\ OBJ_2 & x_{21} & x_{22} & . & . & . & x_{2k} \\ . & & & . & & & . \\ . & & & & . & & . \\ . & & & & & . & . \\ OBJ_n & x_{n1} & x_{n2} & . & . & . & x_{nk} \end{bmatrix}$$

Figura 17. Representação da Tabela de Dados Multivariados

Onde:

OBJ_n = n-ésimo objeto

VAR_k = k-ésima variável (ou atributo) relativa aos objetos

x_{nk} = valor da variável k para o n -ésimo objeto

Ressalta-se que existem alguns **passos fundamentais no processo de obtenção de *clusters*** e que precisam ser definidos pelo analista:

- 1) **Escolha das variáveis** que serão utilizadas para caracterizar os objetos (denominado de vetor característico);
- 2) Similaridade em estatística é calculada através da distância matemática entre vetores numéricos. Portanto, é necessário que se **escolha o tipo de métrica²³ para distância**.

²³ Existem várias medidas de distância matemática como, por exemplo: Canberra, Czekanowski, Euclideana, Minkowski, Mahalanobis entre outras. Ver Johnson & Wichern (1994).

3) **Escolha do algoritmo**²⁴ que será utilizado para comparar os vetores característicos com fins de formar os grupos homogêneos.

4) **Definição do número de grupos (padrões)** que se quer encontrar.

Dado que existem vários métodos para obtenção dos grupos homogêneos, nesta tese, usou-se as Redes Neurais, mais especificamente os Mapas Auto-Organizáveis de Kohonen²⁵ devido às vantagens advindas desta metodologia e exploradas no capítulo 5.

Ressalta-se, entretanto, que no trabalho de investigação da tese, implementou-se uma metodologia para auxiliar na determinação do número de neurônios na topologia da rede neural. A metodologia utilizada inicialmente na tese era a de, definida uma malha bidimensional (ver figura 15), adotar uma análise *top-down*, isto é, partia-se de um número maior de neurônios e ia-se reduzindo o número de grupos através da análise de protótipos (ver Zanini, Souza & Mattos, 2003a; Zanini, Souza & Mattos, 2003b; Zanini *et al*, 1999; Parente *et al*, 2003).

Entretanto, como mencionando na seção 5.3, a quantidade de unidades ou neurônios de saída pode ser arbitrada e mantida fixa, ou então definida automaticamente pelo algoritmo de treinamento. Com base nos trabalhos de Fritzke (1994), Cho (1997) e de Castro & Von Zuben (1999b) foi então implementado o denominado PSOM (*Pruning Self Organizing Maps*).

7.1.1.PSOM: *Pruning Self Organizing Maps*

Apesar de alguns resultados estarem disponíveis na literatura (ver Castro, Von Zuben & Martins, 1998; e Cho, 1997) no que se diz respeito a técnicas de *pruning* para o SOM, a técnica proposta por de Castro & Von Zuben (1999a) para topologias unidimensionais apresenta alguns aspectos inovadores. Como

²⁴ Ressalta-se que existem vários métodos para obtenção dos *clusters*. Tem-se desde os métodos estatísticos clássicos hierárquicos (vizinho mais próximo, vizinho mais afastado, *linkage*, *ward* dentre outros) e não hierárquicos (k-means), até os métodos de inteligência artificial como as redes neurais (SOM, *Learning Vector Quantization* (LVQ)), lógica *fuzzy* (*Fuzzy C-Means* (FCM)), dentre outros.

²⁵ Como a tese adotou o SOM, pode-se ainda acrescentar mais alguns passos ao processo de obtenção dos grupos: **escolha da topologia** (arquitetura: bidimensional ou unidimensional), **definição da distância entre neurônios** (existem também várias métricas como Manhattan, Grid, dentre outras), **definição da função de vizinhança entre os neurônios**, **definição da taxa de aprendizagem**, dentre outros (Ver Haykin, 1998).

mencionado anteriormente, este método objetiva reduzir a dimensão da topologia do mapa (número de neurônios) a partir do algoritmo de treinamento da rede.

O objetivo é achar e eliminar (“podar”) neurônios que não representam um número pré-especificado de padrões. Uma medida de *clustering* (MC) é definida para avaliar o grau de representatividade de cada neurônio. Aqueles neurônios cujo MC é inferior ao *threshold* (ξ) pré-especificado são candidatos a serem eliminados. Se um determinado neurônio tiver um MC alto, um termo de penalidade (τ) pode ser adicionado ao seu MC, forçando outras unidades de saída a aumentarem seu MC.

Usualmente o critério de parada para o SOM é um limitado número de iterações, ou um menor valor para a taxa de aprendizagem (α), assumindo que α decresce ao longo do processo adaptativo dos pesos (treinamento). A fim de obter o MC, o procedimento de poda precisa ser “retardado”, isto é, a rede é treinada da maneira usual para algumas iterações, e tão logo o mapa constitua uma certa definição de topologia, a poda dos neurônios com menor MC acontece. É óbvio que, depois de podar um neurônio numa fase avançada do processo adaptativo, o reduzido valor da taxa de aprendizagem implicará também em um reduzido ajuste do vetor de pesos. Então, um procedimento de reinicialização para o algoritmo é necessário, onde o valor da taxa de aprendizagem (α) e a região de vizinhança (N_c) são inicializados. Como consequência, a poda do neurônio deve ser associada com a restauração dos parâmetros de treinamento. O estado inicial (valores dos pesos) da rede podada corresponde ao estado final da rede anterior (antes do procedimento de poda).

Fazendo uso do estado final da rede anterior, aumentam as chances de que a nova rede apresente resultados melhores, uma vez que se garanta que a rede seja suficientemente capaz de representar o conjunto de dados. Nesta tese, seguindo recomendação de de Castro e Von Zuben (1999a), o critério de parada é o valor mínimo (α_{\min}) da taxa da aprendizagem (α), juntamente com um número fixo de iterações (épocas).

O valor de ξ é um valor percentual, isto é, se um neurônio não representa pelo menos $\xi\%$ dos dados, então ele é um candidato a ser podado. Se a distribuição do conjunto de dados é conhecida a priori, então sugere-se que não mais do que metade do valor percentual da classe menos representativa seja usado

para o parâmetro ξ . Quando o conjunto de dados é completamente desconhecido, valores muito baixos (tipo 0,5% ou 1%) são sugeridos para o parâmetro ξ de forma a garantir que a rede possa ainda representar apropriadamente o conjunto de dados.

Para um pequeno conjunto de dados, com não mais que alguns exemplos, o número inicial de neurônios pode ser assumido como o número de padrões disponíveis, isto é, $m = N$. Este fato pode ser explicado pelo fato de que em um conjunto de dados com N amostras, não pode haver mais do que N diferentes grupos de dados (cada amostra representando um diferente grupo). Se o conjunto de dados é muito grande, esta estratégia é não recomendada devido ao alto esforço computacional aplicado a toda inicialização do processo de treinamento.

O algoritmo de treinamento PSOM pode ser definido como se segue:

1) Definição e inicialização paramétrica

- Inicialize os pesos w_{ij}
- Defina-se: Nc_0 , τ , ξ , I (número de iterações), m_{\min} , $delay$ e α_0 .

2) Quando o critério de parada é falso, fazer:

2.1) Para cada I determinar:

$$2.1.1) J = \arg \min_j \{ \|w_j - x_j\| \}$$

2.1.2) Aumentar MC (J)

2.1.3) $\forall j \in Nc(J)$, e $\forall k$:

$$w_{jk}(\text{novo}) = w_{jk}(\text{velho}) + \alpha[x_{jk} - w_{jk}(\text{velho})]$$

2.2) Se $(m > m_{\min})$ e $(\text{épocas} > delay)$ e $(\text{épocas} \bmod I \neq 0)$

2.2.1) Pode o neurônio que agrupa menos que $\xi\%$ das amostras, restaure $\alpha = \alpha_0$ e $Nc(j) = \text{floor}(m/2)$

2.2.2) Caso contrário, mantenha todos os neurônios

2.3) Se $MC(j)$, $j = 1, \dots, N$ é grande, penalize-o por τ

2.4) Diminua Nc e α

Este algoritmo foi implementado nas pesquisas realizadas nesta tese, sendo que os resultados serão comentados no capítulo 8.

7.2. Análise de Fronteira Estocástica

A implementação dos modelos de fronteira estocástica, apresentados no capítulo 4, exigiu que antes fosse feito um tratamento dos dados de forma a adaptar a estimação dos parâmetros dos modelos. Foram abordados basicamente:

- Hipótese sobre retornos de escala
- Tratamento de *outliers* (pontos discrepantes)

Ressalta-se que para as formas de função de produção e custo e distribuição dos resíduos foram utilizadas abordagens comumente utilizadas na literatura e especificadas no capítulo 4 e pesquisadas no capítulo 3, isto é, função Cobb-Douglas e Translog para formas de função e distribuições Normal/Half-Normal e Normal Truncada para distribuições dos resíduos.

7.2.1. Hipótese sobre Retornos de Escala

Este é um dos pontos principais de discussão nas metodologias de definição de *frontier benchmarking* (seja utilizando o DEA, seja utilizando AFS). A hipótese de retornos de escala constantes (*constant return to scale - CRS*) merece uma análise cuidadosa.

Na figura 02 foi apresentada uma função com o conjunto de possibilidades de produção (CPP). Considerando que para produzir o nível de produto y sejam necessários dois fatores de produção: x_1 e x_2 . Então $y = f(x_1, x_2)$. Com base neste CPP, pode-se abordar os chamados rendimentos de escala (Varian, 1999). Imagine que se aumente todos os insumos da função de produção, isto é, que se multiplique a quantidade de todos os insumos por algum fator constante, por exemplo, que se use o dobro do fator 1 e o dobro do fator 2. A pergunta é: caso se utilize o dobro de cada insumo, qual a quantidade de produto será produzida? O resultado mais provável é que se obtenha o dobro de produto. Isso é chamado de **rendimentos constantes de escala**. Em termos de função de produção significa que o dobro de cada insumo dá o dobro do produto. No caso de dois insumo pode-se indicar matematicamente pela expressão:

$$2f(x_1, x_2) = f(2x_1, 2x_2) \quad (7.1)$$

Em geral, se a escala de todos os insumos aumenta em uma certa quantidade t , os rendimentos constantes de escala implicam que se obtenha então t vezes mais produto:

$$tf(x_1, x_2) = f(tx_1, tx_2) \quad (7.2)$$

Se diz que este é o produto provável pelas seguintes razões: normalmente a firma poderia replicar suas atividades anteriores. Se a firma tem o dobro de cada insumo, ela pode simplesmente instalar duas fábricas idênticas e, portanto, obter o dobro do produto. Tendo o triplo de cada insumo, a firma poderia instalar três fábricas idênticas e assim por diante.

O caso de rendimentos constantes de escala é o caso mais “natural” tendo em vista o argumento da replicagem, mas isso não quer dizer que outros resultados não possam ocorrer. Por exemplo, poderá acontecer que, ao multiplicar ambos os insumos por algum fator t , obtenha-se mais de t vezes o produto anterior. Isso é conhecido como o caso de **rendimentos crescentes de escala**²⁶. Matematicamente, este tipo de rendimentos de escala significa que:

$$f(tx_1, tx_2) > tf(x_1, x_2) \quad (7.3)$$

para todo $t > 1$.

O outro caso a considerar é aquele de **retornos decrescentes de escala**²⁷, onde:

$$f(tx_1, tx_2) < tf(x_1, x_2) \quad (7.4)$$

para todo $t > 1$.

Assumir CRS na análise comparativa do desempenho operacional das concessionárias favorece os consumidores, assumir retornos de escala variáveis (VRS, do inglês *variable returns to scale*) favorece as empresas. Desta forma a escolha de qual o tipo de retornos de escala considerar tem uma influência

²⁶ Um exemplo de uma tecnologia com rendimentos crescentes de escala é o caso das canherias utilizadas na extração de petróleo. Caso se duplique o diâmetro da canheria, estará se utilizando o dobro de materiais, mas a circunferência da canheria será quatro vezes maior. Portanto, poderá ser bombeada mais do que o dobro de petróleo. Importante ressaltar que os rendimentos crescentes de escala só ocorrem dentro de certos níveis de produção. No exemplo da canheria, caso se continue duplicando o diâmetro, ela acabaria cedendo por causa do peso.

²⁷ Esse caso é um pouco peculiar. Caso se obtenha menos que o dobro de produto ao duplicar cada um dos insumos, deve haver alguma coisa errada. Ao final das contas, pode-se simplesmente duplicar o que se estava fazendo anteriormente. A forma usual na qual rendimentos decrescentes de escala aparecem é quando se “esquece” de levar em consideração algum insumo. Caso se tenha o dobro de todos os insumos, exceto um deles, não se poderá duplicar o que se estava fazendo antes, de modo que não é obrigatório obter o dobro de produto. Os rendimentos decrescentes de escala são, na verdade, um fenômeno de curto prazo, quando alguma coisa está fixa.

importante sobre a redução de OPEX a exigir de cada empresa no âmbito da fixação do seu *price-cap*.

Um argumento comum em favor da adoção de CRS é baseado na noção que o regulador tem o dever de salvaguardar o interesse público. Claramente, para isso o regulador não deve permitir que as distribuidoras transfiram para o público as suas eventuais ineficiências de escala. Além disso, assumindo CRS incentiva as empresas a continuamente melhorarem sua escala de operação.

O argumento contrário é baseado na noção de que a escala em que cada empresa opera não está completamente sob o seu controle, uma vez que a escala foi herdada em boa parte da situação pré-privatização e o gerenciamento não pode alterar a dimensão da sua área de concessão. Isto significa que é difícil promover ações conducentes à mudança para uma escala de operação mais eficiente, ainda mais porque não são permitidas fusões ou aquisições. Se o *benchmark* de uma empresa DEA-ineficiente, por exemplo, for uma empresa de escala significativamente diferente, a redução de OPEX que lhe for exigida pode ser, naquelas condições, impraticável. O mesmo não aconteceria na hipótese de VRS.

7.2.2.

Tratamento de *Outliers* nos Modelos de Fronteira Estocástica

A partir da análise de pontos discrepantes utilizando gráficos de dispersão das variáveis²⁸ das concessionárias, foram testadas diversas possibilidades de inclusão de variáveis *dummies*²⁹ nos modelos com o objetivo de capturar as especificidades das concessionárias atípicas e assim não contaminar os indicadores de eficiência.

7.3.

Metodologia Proposta

A metodologia proposta pela tese para cálculo da eficiência operacional (X^0), componente da fórmula de cálculo do fator X (capítulo 6, equação 6.7) também proposta por este trabalho, consiste então em conjugar uma técnica de

²⁸ A seleção de variáveis utilizadas nos modelos será discutida no capítulo 8.

²⁹ Este tipo de tratamento de discrepâncias nos modelos pode ser verificado em Kumbhakar & Lovell, 2000.

classificação de padrões através do algoritmo SOM adaptado, isto é, o PSOM (apresentado na seção 7.1) com um modelo econométrico de análise de fronteira estocástica modificado para englobar a presença de pontos discrepantes no universo das empresas de distribuição de energia elétrica no Brasil. Os resultados da aplicação desta metodologia serão apresentados no capítulo 8.