PROPOSTA DE NOVAS CLASSES DE RISCO DE ACIDENTE DO TRABALHO NO BRASIL COM A UTILIZAÇÃO DE MÉTRICAS DE VALIDAÇÃO DE AGRUPAMENTOS

MOACYR MACHADO CARDOSO JUNIOR
Instituto Tecnológico de Aeronáutica – ITA
Praça Mal. Eduardo Gomes, 50 – São José dos Campos
moacyr@ita.br

RODRIGO ARNALDO SCARPEL
Instituto Tecnológico de Aeronáutica – ITA
Praça Mal. Eduardo Gomes, 50 – São José dos Campos
rodrigo@ita.br

RESUMO

O presente trabalho tem por objetivo propor novas classes de risco de acidente do trabalho no Brasil por meio da utilização de métodos de agrupamentos hierárquicos e não hierárquicos. O número ideal de classes foi obtido com o auxílio de métricas de validação interna e de estabilidade. Um método de classificação conhecido como Vizinhos mais próximos – KNN, foi aplicado na base de dados de indicadores de acidentes do ano de 2007 e com as instâncias, denominadas por código nacional de atividades econômicas – CNAE, classificadas segundo a Portaria n° 76 do Ministério do Trabalho e Emprego. Como resultado verificou-se um desempenho ruim do classificador devido à incoerência dos indicadores de acidentes com as atuais classes de risco. Novas classes foram então propostas com a utilização das métricas de validação de agrupamentos.

PALAVRAS CHAVE. Acidente do Trabalho. Classificação de Risco. Validação de Agrupamentos.

Área de classificação principal (Outras aplicações em PO (OA))

ABSTRACT

The main issue of this study is to propose new occupational accident risk classes in Brazil by hierarchical and non-hierarchical clustering methods. The ideal number of classes was obtained with internal and stability validation techniques. The K-nearest neighbor classifier was applied to 2007 annual occupational accident indicators data and to instances designed by economic activity national code – CNAE, classified by employment and labour ministry in an official regulation n° 76. The results showed a high misclassification though to the contradiction imposed by accident indicators and the nowadays risk classification. New classes were then proposed using clustering validation metrics.

KEYWORDS. Cluster validation. Risk classification. Occupational accident.

Main area (Other applications in OR (OA)).

1. Introdução

O Ministério do Trabalho e Emprego adota as classes de riscos dos diversos setores econômicos, representados pelo Código Nacional de Atividade Econômica (CNAE), a partir dos dados constantes da Portaria Nº 76, atualizada em 25 de novembro de 2008. A referida Portaria apresenta os diversos setores econômicos – CNAE distribuídos em quatro classes de Risco: 1, 2,3 e 4, e apesar dos critérios para esta classificação não estarem especificados nesta Portaria e nem nas suas antecessoras, Portaria Nº 04 publicada em Outubro de 1991 e Portaria Nº 01 de Maio de 1995, o mesmo deve estar baseado nos indicadores de acidentes do trabalho no Brasil, conforme MTE (2008).

O Ministério da Previdência Social compila anualmente os indicadores de acidentes do trabalho com a finalidade de mensurar a exposição dos trabalhadores aos níveis de risco inerentes à atividade econômica, permitindo o acompanhamento das flutuações e tendências históricas dos acidentes e seus impactos nas empresas e na vida dos trabalhadores. Além disso, fornecem subsídios para o aprofundamento de estudos sobre o tema e permitem o planejamento de ações nas áreas de segurança e saúde do trabalhador, (MPS, 2009).

Os indicadores de acidentes de trabalho disponibilizados pela Previdência Social para o ano de 2007 e suas respectivas descrições estão na Figura 1.

FIGURA 1 - Indicadores de Acidentes do Trabalho divulgados pela Previdência Social.

Indicador	Definição
Taxa de Incidência-TxAc	Expressa a relação entre o número de novos acidentes do trabalho registrados no ano e a população exposta ao risco.
Taxa de incidência de Doenças do Trabalho-TxIncDoe	Expressa a relação entre o número de novos casos de doenças do trabalho registrados no ano e a população exposta ao risco.
Taxa de incidência de Acidentes Típicos–TxIncAcTip	Expressa a relação entre o número de novos casos de Acidentes Típicos do trabalho registrados no ano e a população exposta ao risco.
Taxa de incidência de acidentes que causam incapacidade temporária–TxIncAcIncTemp	Expressa a relação entre o número de novos casos de Acidentes do trabalho registrados que causam a incapacidade temporária no ano e a população exposta ao risco.
Taxa Mortalidade-TxMort	Expressa a relação entre o número total de óbitos decorrentes de acidente do trabalho no ano e a população exposta.
Taxa Letalidade-TxLetal	Expressa a gravidade dos acidentes ocorridos, relacionando o número de óbitos com a quantidade de acidentes registrados no ano.
Taxa de incidência específico para a faixa etária de 16-34 anos–Tx16-34	Expressa a relação entre o número de novos acidentes do trabalho registrados no ano e a população exposta ao risco, para um grupo extremamente crítico em termos de acidentes do trabalho, pois afeta uma parcela jovem da população trabalhadora.

Fonte: MPS, 2009.

A adoção das atuais classes de risco influi diretamente no dimensionamento do quadro do Serviço de Segurança e Medicina do Trabalho das Empresas, e indiretamente no estabelecimento de políticas de fiscalização, pesquisa, dentre outros. Desta forma é de vital importância que as classes reflitam corretamente os indicadores de acidentes, evitando onerar algum segmento econômico, ou mesmo a não adoção de ações corretas em outros setores.

Desta forma este trabalho tem por objetivo avaliar as atuais classes de risco de acidente, confrontando as mesmas com os indicadores de acidentes do trabalho divulgados pelo Ministério da Previdência Social para o ano de 2007, que é o conjunto de dados mais atualizados de que se dispõe no momento da realização deste trabalho e estabelecer novas classes de risco em função de novo agrupamento de classes utilizando técnicas de formação de agrupamentos hierárquicas e não hierárquicas, bem como a validação dos agrupamentos obtidos.

Neste trabalho considera-se a hipótese de que as classes de risco para as empresas podem ser representadas pelos indicadores de acidentes do trabalho divulgadas pela Previdência Social, e espera-se como contribuição o aperfeiçoamento da classificação de riscos trabalhistas com a utilização de técnicas adequadas.

Este trabalho encontra-se organizado da seguinte forma: na seção de introdução, foi abordada a motivação e os objetivos para desenvolvimento deste trabalho; a Seção 2 o teste de coerência das atuais Classes definidas pela Portaria n° 76, por intermédio da utilização do classificador vizinhos mais próximos – KNN. A Seção 3 apresenta as métricas para validação de agrupamentos hierárquicos e não hierárquicos; a Seção 4 apresenta as análises e resultados obtidos com a utilização de algoritmos de agrupamentos hierárquicos, com método Ward e não hierárquicos com o *K-médias* e *K-medóides*, assim como a validação dos agrupamentos obtidos com o auxílio de métricas de validação interna e de estabilidade; e finalmente a Seção 5 com as conclusões finais.

2. Testando as Classes atuais

Para verificar a coerência das atuais classes definidas pela Portaria n° 76, ou seja, demonstrar que as atuais classes de risco não estão condizentes com os atributos (indicadores de acidentes) será utilizado o classificador K-*nearest neighbour* – KNN.

O método KNN é um método não paramétrico para classificação de objetos baseado na densidade obtida quando ao fixarmos uma instância vamos crescendo a distância até alcançar K-vizinhos mais próximos. O KNN é o melhor classificador quando não se tem idéia da distribuição de probabilidades, é simples e intuitivo e funciona bem quando o número de amostras é grande. (DUDA, HART e STORK, 2001).

Para tanto os dados originais que estão em uma matriz de 655x8, ou seja, 655 grupos de CNAE, 7 atributos e 1 descritor da classe de risco (1, 2, 3 ou 4). Os dados serão divididos em 1 grupo de treinamento com 350 instâncias e 1 grupo de teste com as 305 restantes.

Utilizando o KNN implementado no pacote "*Class*" do software R, e o valor de K fixado em 1, obteve-se a matriz de confusão apresentada na Tabela 1.

TABELA 1. Matriz de confusão obtida a partir da classificação realizada pelo método KNN.

Classe atribuída pelo KNN	Classe original			
	1	2	3	4
1	27	11	11	1
2	16	32	37	4
3	12	16	97	14
4	2	1	20	4

A partir deste resultado verifica-se que os dados originais de classes de risco, não guardam relação com os indicadores de acidentes oficiais disponibilizados pela Previdência Social. O erro de classificação da classe 1, por exemplo, é de 53%, o da classe 2 é de 47%, o da classe 3, 41% e finalmente o da classe 4, 83%.

Conclui-se então que este baixo nível de acerto do classificador KNN se deve a um desajuste na estrutura dos dados, indicando que as classes não seguem nenhum padrão que possa ser reconhecido por um classificador.

3. Estabelecendo novas classes

O número de técnicas formais para a determinação do número ideal de agrupamentos passa de trinta, segundo Milligan e Cooper (1985) e Everitt *et al*.(2001). Estes autores avaliaram a habilidade de tais métodos indicarem o número ideal de agrupamentos com dados simulados e suas conclusões são de que não se pode generalizar, uma vez que dependem da estrutura desconhecida dos dados, assim como do algoritmo utilizado.

Para determinação do número ideal de classes utilizaremos a comparação de vários métodos de geração agrupamento, hierárquicos e não hierárquicos juntamente com métricas de validação do número de classes encontradas.

Os agrupamentos hierárquicos são gerados com base em uma matriz de dissimilaridade, que no caso deste estudo utilizou-se a distância Euclidiana. O método utilizado é o de Ward, ou algoritmo de variância mínima, que busca aglomerar as observações de forma a minimizar a variância total, conforme Theodoridis e Koutroumbas (2009).

No caso dos métodos não hierárquicos parte-se de um número definido de agrupamentos que se deseja particionar os dados. Estes métodos buscam maximizar a distância entre agrupamentos diferentes e minimizar as distâncias intra-agrupamento. Neste trabalho são utilizados dois métodos *K-médias* e sua variante *K-medóides* (*PAM – partioning around medoids*), que difere do primeiro pelo fato de determinar um centróide de cada agrupamento, enquanto que o segundo parte de um centro coincidente com umas das instâncias, de acordo com Theodoridis e Koutroumbas (2009).

A formulação, por programação matemática, do algoritmo *K-médias*, conforme Webb (2002).

$$Min \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{c=1}^{k} z_{ic} \left[\sum_{\mathrm{j}=1}^{\mathrm{p}} (x_{ij} - m_{cj})^2 \right]^{\frac{1}{2}}$$

$$S.A. \quad \sum_{c=1}^{k} z_{ic} = 1 \qquad \mathrm{i} = 1,\ldots,\mathrm{n}$$

$$z_{ic} = \begin{cases} 1, \text{se o ponto i pertencer ao agrupamento c} \\ 0, \text{caso contrário} \end{cases} \qquad , \mathrm{c} = 1,\ldots,\mathrm{k} \;\; \mathrm{e} \;\; \mathrm{i} = 1,\ldots,\mathrm{n}$$

em que

$$m_{cj} = \frac{\sum_{i=1}^{n} z_{ic}\, x_{ij}}{\sum_{i=1}^{n} z_{ic}} \qquad \mathrm{c} = 1,\ldots,\mathrm{k} \quad \mathrm{e} \quad \mathrm{j} = 1,\ldots,\mathrm{p}$$

sendo m_{cj} o centróide do agrupamento c na dimensão j.

3.1. Métricas utilizadas

As métricas utilizadas para validação do número de agrupamentos ou classes nas quais os dados serão particionados são divididas em dois grandes grupos: Interna e Estabilidade, conforme proposto por Brock *et al*.(2008).

Para a validação interna foram selecionados medidas que reflitam a compactação, conectividade e separação dos agrupamentos gerados. A conectividade refere-se a extensão na qual uma instância está alocada a um mesmo agrupamento dos seus vizinhos mais próximos. A compactação avalia a homogeneidade do agrupamento, usualmente calculando a variância intra-agrupamento, enquanto a separação quantifica o grau de separação dos agrupamentos, usualmente pela medida da distância dos centróides. Uma vez que a compactação e a separação apresentam tendências opostas, ou seja, compactação aumenta com o numero de agrupamentos e a separação diminui, uma opção é unir as duas métricas. Duas medidas que representam uma combinação não linear de compactação e separação são o índice de Dunn e a largura de Silhouette. (EVERITT *et al*., 2001)

A conectividade é definida por:

$$con(C) = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{i,} nn_{i(j)} \tag{2}$$

onde N representa o número total de observações e M o número de dimensões.

$nn_{i(j)}$ é o j-ésimo vizinho mais próximo da instância i na dimensão j, e $x_{i,} nn_{i(j)} = 0$ se i e j estiverem no mesmo agrupamento e 1/j caso contrário.

A conectividade tem valores entre $0 \leq con(C) \leq \infty$, e é uma métrica que deve ser minimizada, ou seja, quanto menor o valor melhor será a estrutura proposta pelo algoritmo, conforme Everitt *et al*. (2001).

O índice Dunn é a razão entre a menor distância entre instâncias que não estejam no mesmo agrupamento e a maior distância intra-agrupamento. O valor do índice Dunn varia de 0 a 1 e quanto mais próximo de 1 melhor é o resultado, de acordo com Brock *et al*. (2008).

A largura de Silhouette proposta por Kauffman e Rousseeuw (1990) e recomendada por Everitt *et al*., (2001). Para cada instância i é definido um índice $S(i) \in [-1,1]$ que mede a diferença entre b(i) e a(i), onde a(i) é a dissimilaridade média da instância i em relação ao seu agrupamento e b(i) é a dissimilaridade média da instância i em relação a todas instâncias no agrupamento mais próximo. Quando S(i) está próximo de 1a instância i está mais próxima do seu agrupamento do que do agrupamento vizinho, e, portanto representa uma boa alocação. Quando S(i) está próximo de -1, a instância está mal alocada. Os autores da proposta indicam ainda que valores acima de 0,5 representam um bom resultado e que valores abaixo de 0,2 podem indicar ausência de estruturação clara dos dados.

Finalizando as métricas de validação interna, Everitt *et al*. (2001), alertam não ser prudente confiar em apenas uma das métricas para selecionar o número ideal de agrupamentos.

A validação utilizando o conceito da estabilidade se baseia na retirada de uma coluna por vez da matriz inicial e compara os agrupamentos formados antes e depois da retirada. As métricas utilizadas foram proporção média de não sobreposição – (*Average Proportion of Non-overlap* - APN), Distância Média – (*Average Distance* - AD), Distância média entre centros dos agrupamentos - (*Average Distance between Means* - ADM), e Figura de Mérito – (*Figure of Merit* - FOM).

A APN mede a proporção média de instâncias não alocadas ao mesmo agrupamento quando a matriz completa é utilizada e quando uma coluna é retirada. Seja $C^{i,0}$ o agrupamento que contém a instância i usando a matriz completa e $C^{i,l}$ quando uma coluna l foi retirada. Considerando o número total de agrupamentos igual a K, define-se:

$$APN(K) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}\left(1 - \frac{n(C^{i,l} \cap C^{i,0})}{n.C^{i,0}}\right) \tag{3}$$

Com $APN(K) \in [0,1]$, sendo que valores próximos a zero correspondem a um agrupamento consistente. (BROCK *et al*., 2008)

AD mede a distância média entre instâncias de um mesmo agrupamento baseada na matriz completa e com a retirada de uma coluna, é definida por:

$$AD(K) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}\frac{1}{n(C^{i,0})n(C^{i,l})}\left[\sum_{i \in C^{i,0}, j \in C^{i,l}} dist(i,j)\right] \tag{4}$$

com $AD(K) \in [0,\infty]$, sendo que valores próximos a zero os melhores resultados.

ADM mede a distância média entre centros dos agrupamentos, e é definida por:

$$ADM(K) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M} dist(\overline{X}_{C^{i,l}}, \overline{X}_{C^{i,0}}) \tag{5}$$

com $ADM(K) \in [0,\infty]$, e da mesma forma que AD, quanto menor melhor.

E finalmente FOM mede a variância média intra-agrupamento das instancias na coluna suprimida, onde o agrupamento está baseado nas colunas restantes. É estimado o erro médio usando previsões baseadas nas médias dos agrupamentos. Para uma coluna l deixada de fora, temos:

$$FOM(K) = \sqrt{\frac{1}{N}\sum_{k=1}^{K}\sum_{i \in CK(l)} dist(x_{i,l}, \overline{x}_{CK(l)})} \tag{6}$$

onde $x_{i,l}$ é o valor da i-esima instancia da l-esima coluna no agrupamento $CK(l)$ e $\overline{x}_{CK(l)}$ e a média do agrupamento. Finalmente FOM é multiplicado por um fator $\sqrt{\frac{N}{N-K}}$, para minimizar a tendência de FOM decrescer à medida que K cresce. O escore final e calculado pela média de todas as colunas removidas, e os valores obtidos estão entre zero e ∞, com os menores valores representando a melhor configuração, conforme Brock *et al*. (2008).

4. Análises e Resultados

Todas as métricas descritas anteriormente foram aplicadas ao conjunto de dados de indicadores de acidentes na busca pela melhor estruturação dos dados em termos de classes. Para tanto foram testadas de 2 a 4 classes, com os métodos de formação de agrupamentos hierárquicos Ward e não-hierárquicos *K-médias* e *K-medóides*. Foi utilizado o pacote do software R Versão 2.9 com o pacote clValid, implementado por Brock et al. (2008).

Os resultados obtidos para a validação interna estão na Tabela 2.

TABELA 2 - Resultado da validação interna do número ideal de agrupamentos a serem gerados com a estrutura dos dados de indicadores de acidentes, utilizando 3 algoritmos de agrupamento

Método de agrupamento	Métrica	2 Classes	3 Classes	4 Classes
Hierárquico – Ward	Conectividade	36,2806	100,2135	100,2135
	Dunn	0,0069	0,0036	0,0036
	Silhouette	0,4827	0,1601	0,1739
K-médias	Conectividade	**2,9290** *	47,1214	111,0123
	Dunn	**0,9881** *	0,0142	0,0057
	Silhouette	**0,9429** *	0,4589	0,3339
K-medóides (pam)	Conectividade	115,8095	151,8845	221,3377
	Dunn	0,0033	0,0036	0,0014
	Silhouette	0,2483	0,2180	0,1353

Os resultados da validação interna mostram claramente que duas classes geram o melhor resultado em todas as avaliações, e que ainda o melhor método foi o *K-médias*.

Na Figura 2 podem-se visualizar os resultados das métricas conectividade, Dunn e Silhoutte na definição do melhor número de classes.

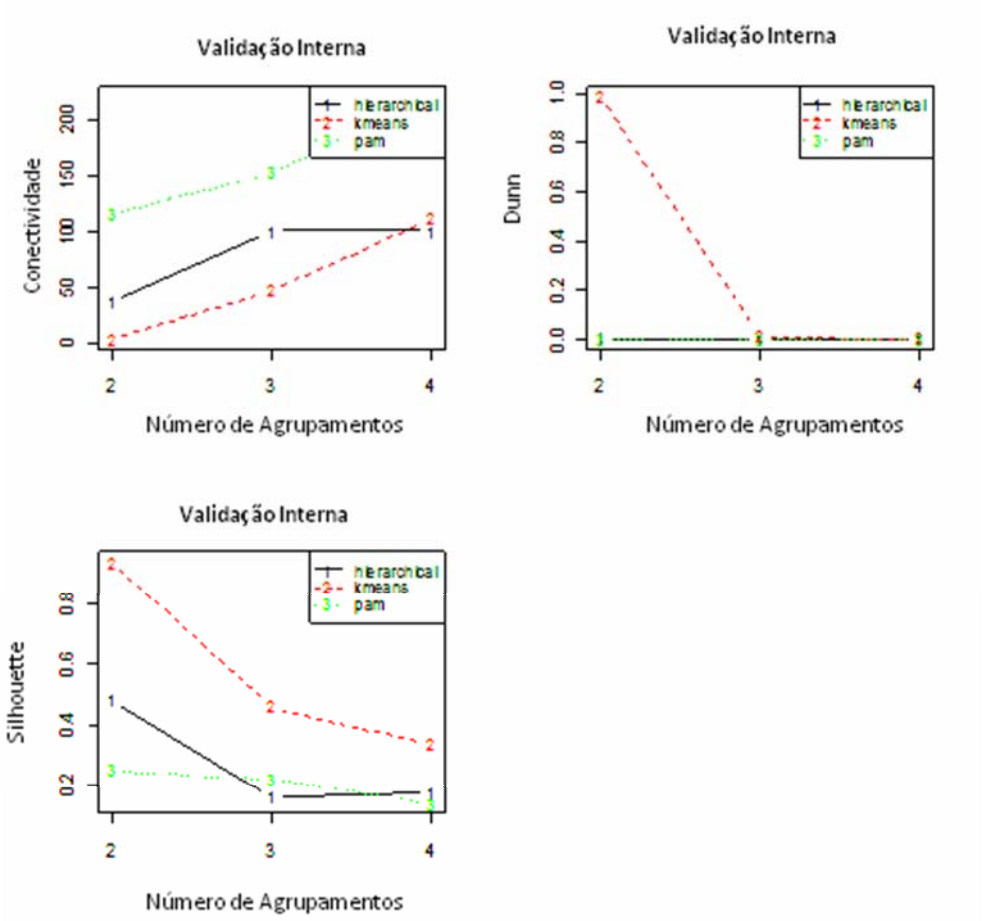

FIGURA 2 - Resultados das Métricas de Validação Interna para 2, 3 e 4 classes.

Com a adoção deste critério, poderíamos demonstrar a divisão das novas classes com o dendograma ilustrativo da Figura 3.

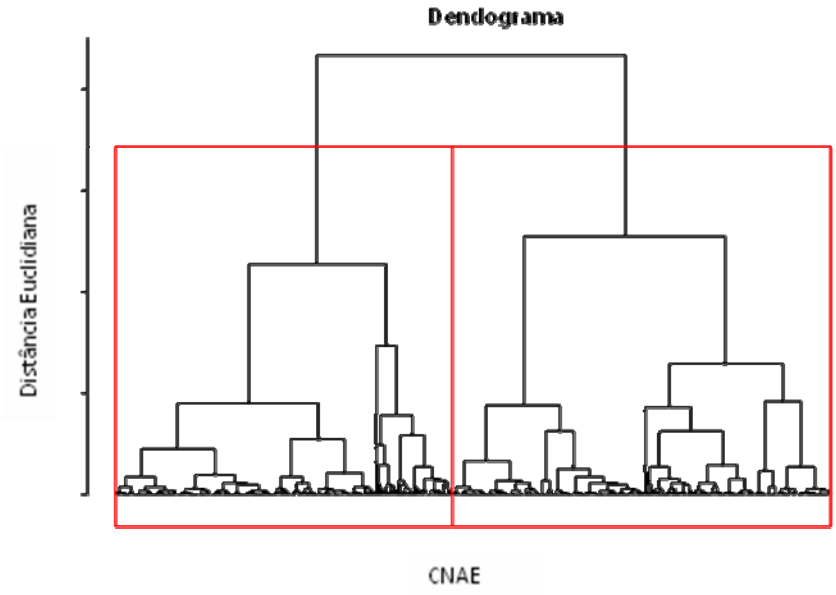

FIGURA 3 - Dendograma obtido com algoritmo hierárquico Ward, demonstrando as duas classes geradas.

Para as métricas relacionadas à estabilidade do agrupamento, os resultados foram diferentes das de validação interna, e nos quatro critério houve um empate entre 2 e 4 classes, com o melhores valores indicando o método pam, conforme Tabela 3.

TABELA 3 - Resultado da validação por estabilidade do número ideal de classes a serem geradas com a estrutura dos dados de indicadores de acidentes.

Método de agrupamento	Métrica	2 Classes	3 Classes	4 Classes
Hierárquico – Ward	APN	0,2945	0,3038	0,3620
	AD	2,2543	2,0414	1,9492
	ADM	0,7245	0,6241	0,6701
	FOM	0,9749	0,9457	0,9398
K-médias	APN	0,2250	0,2968	0,1956
	AD	2,3279	2,1653	1,9052
	ADM	0,5081	0,7712	0,5131
	FOM	0,9686	0,8504	**0,8369 ***
K-medóides (pam)	APN	**0,0973 ***	0,1275	0,2179
	AD	2,1221	1,9131	**1,8602 ***
	ADM	**0,2220 ***	0,2407	0,4361
	FOM	0,9531	0,9469	0,9426

A conclusão a que se chega é que os dois métodos utilizados apontam para lados diferentes, ou seja, métodos de validação interna apontam para 2 classes, enquanto que os métodos de estabilidade apontam um empate para 2 ou 4 classes. Considerando ainda que os dados originais estão agrupados em quatro classes, optou-se nesse trabalho em utilizar 4 classes com o agrupamento realizado pelo K-*medoides* (pam). Descartou-se o estudo com duas classes, pois os agrupamentos generalizam demais os grupos de CNAE, o que pode prejudicar a adoção de políticas e programas voltados a fiscalização das empresas de cada segmento. Na figura 4 visualiza o gráfico *biplot* obtido pelo método das componentes principais (duas primeiras componentes) para facilitar a visualização das classes obtidas. A análise de Componentes Principais para as duas primeiras dimensões explica aproximadamente 72% da variância total.

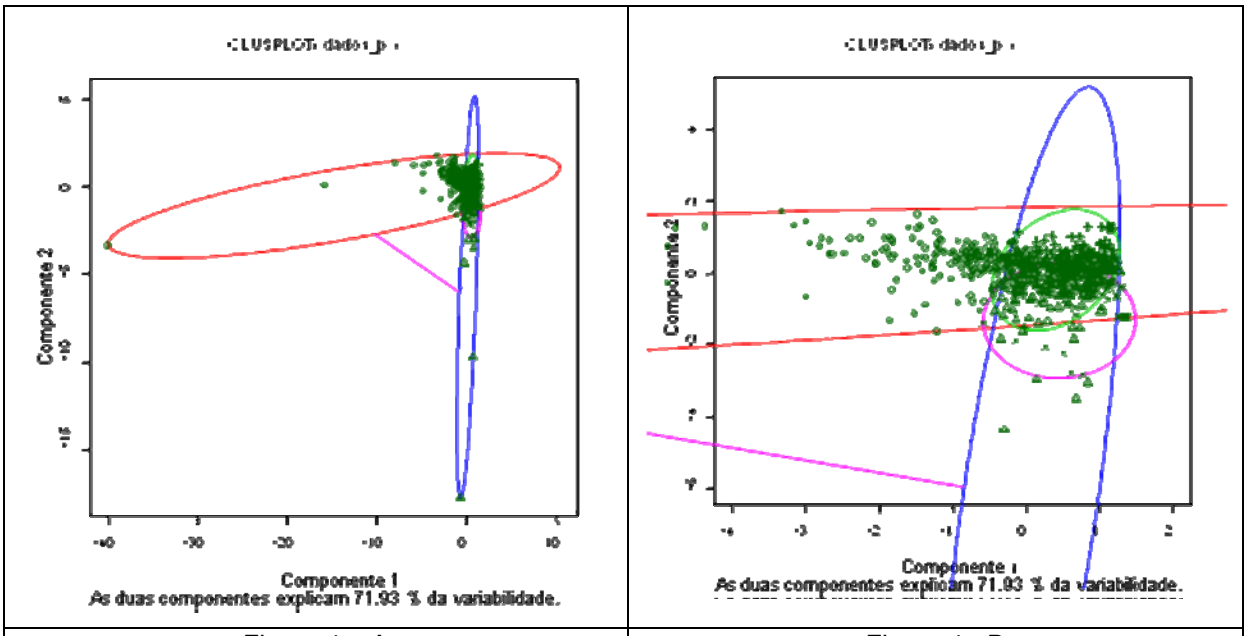

Figura 4 – A	Figura 4 - B

FIGURA 4 – Gráfico de visualização bivariada das 4 classes obtidas com o *K-medoides*, obtido por componentes principais. (4-A) e zoom (4-B).

Na Figura 5, apresenta-se o gráfico de Silhouette obtido após o agrupamento das novas classes pelo método *K-medoides*.

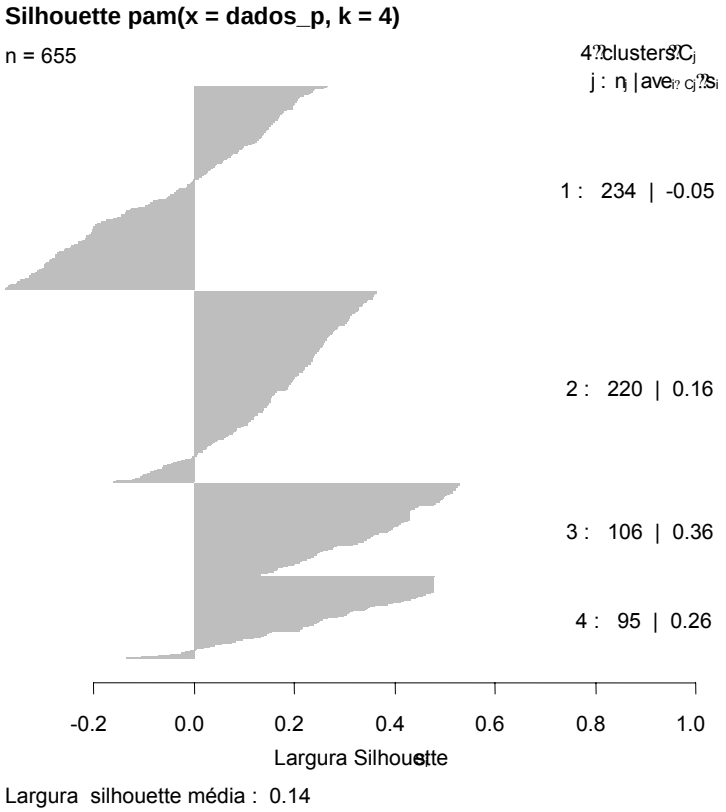

FIGURA 5 – Gráfico da distância de Silhouette para as quatro classes obtidas pelo método *K-medoides*.

Como resultado final do processo de redefinição das Classes de Risco previstas na Portaria n° 76, obtiveram-se novos agrupamentos, que refletem de forma mais adequada os indicadores de acidentes do trabalho dos segmentos econômicos.

Na tabela 4, apresenta-se um extrato comparativo das classes originais e das novas classes.

TABELA 4 - Extrato comparativo das classes originais da Portaria n° 76 e as classes obtidas após o processo de validação interna.

Descrição	Classe original	Nova Classificação obtida
Administração Pública em Geral	1	2
Regulação das Atividades de Saúde, Educa	1	4
Regulação das Atividades Econômicas	1	2
Atividades de Gravação de Som e de Edição	2	2
Atividades de Rádio	2	3
Atividades de Televisão Aberta	2	1
Cultivo de Cereais	3	1
Cultivo de Algodão Herbáceo e de Outras	3	2
Cultivo de Cana-De-Açúcar	3	3
Cultivo de Fumo	3	2
Cultivo de Soja	3	2
Cultivo de Oleaginosas de Lavoura Tempor	3	3
Cultivo de Plantas de Lavoura Temporária	3	3
Horticultura	3	3
Extração de Minério de Ferro	4	3
Extração de Minério de Alumínio	4	3
Extração de Minério de Estanho	4	2
Extração de Minério de Manganês	4	4

5. Conclusão

Neste trabalho foi apresentada uma nova forma de classificação das instâncias representadas pelos segmentos econômicos brasileiros. Esta proposta busca refletir melhor a estrutura dos dados representados pelos indicadores de acidentes e suas respectivas classes.

A utilização das métricas de validação interna e de estabilidade dos agrupamentos forneceu um critério mais sólido para a reclassificação dos segmentos econômicos com base nos indicadores reais de acidentes do Brasil para o ano de 2007, permitindo que a classificação apresente coerência com as estatísticas oficiais, apesar de não terem convergido para um valor único, o que não elimina a necessidade da ação do decisor na fixação do número ideal de agrupamentos.

Espera-se que este trabalho possa contribuir no sentido de mostrar a importância de se estudar a estrutura dos dados de forma a tomar a decisão mais acertada.

Como sugestão para trabalhos futuros a inclusão de novas instâncias, e /ou novos períodos a serem testados.

6. Referências.

Brock G.; Pihur, V.; Datta, S.; Datta, S. (2008), clValid: An R Package for Cluster Validation. Journal of Statistical Software, 25, 4.

Duda R. O.; Hart P. E.;Stork, D. G. (2001), Pattern Classification. - New York : John Wiley & Sons,V. 2.

Everitt B. S, Landau S. e Morven L. (2001), Cluster Analysis. London : Arnold. - Vol. 4 Ed.

Kaufman, L.; Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons.

Milligan, G. W.; Cooper, M. (1985), An examination of procedures for determining the number of clusters in a data set, *Psychometrika*, 50, 159-179.

[MPS] Ministério da Previdência Social Indicadores de Acidentes do Trabalho. Ministério da Previdência Social. - Acesso em: 07 de Novembro de 2009. Disponível em www.mps.gov.br.

[MPS] Ministério da Previdência Social Ministério da Previdência Social. - Acesso em: 12 de Fevereiro de 2009. Disponivel em:http://www.mps.gov.br/conteudoDinamico.php?id=423 .

[MTE] Ministério do Trabalho e Emprego Portaria Nº 04 Ministério do Trabalho e Emprego. - 10 de Outubro de 2001. - Acesso em: 07 de Novembro de 2009. Disponível em: http://www.mte.gov.br/legislacao/potarias.

[MTE] Ministério do Trabalho e Emprego Portaria Nº 1. - 12 de Maio de 1995. - Acesso em: 07 de Novembro de 2009. Disponível em: http://www.mte.gov.br/legislacao/portarias.

[MTE] Ministério do Trabalho e Emprego Portaria nº 76, de 21 de novembro de 2008 Minstério do Trabalho e Emprego. - 21 de Novembro de 2008. - Acesso em: 07 de Novembro de 2009. Disponível em: http://www.mte.gov.br/legislacao/portarias.

R Development Team (2006), R: A language and environment for statistical computing. Vienna : R. Foundation for Statistical Computing.

Theodoridis S. e Koutroumbas K. (2009), Pattern Recognition. London : Elsevier.

Venables A.C. e Ripley B.D. (2002), Modern Applied Statistics with S. - New York : Springer.

Webb, A.(2002) *Statistical Pattern Recognition, 2ª Ed*, Weat Sussex: John Wiley & Sons, Inc.