

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

Bárbara Brzezinski Azevedo

**AGRUPAMENTO DE TRABALHADORES COM
PERFIS SEMELHANTES DE APRENDIZADO
UTILIZANDO TÉCNICAS MULTIVARIADAS**

Porto Alegre

2013

Bárbara Brzezinski Azevedo

**Agrupamento de Trabalhadores com Perfis Semelhantes de Aprendizado utilizando
Técnicas Multivariadas**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Acadêmica, na área de concentração em Sistemas de Produção.

Orientador: Michel José Anzanello, *Ph.D.*

Porto Alegre

2013

Bárbara Brzezinski Azevedo

**Agrupamento de Trabalhadores com Perfis Semelhantes de Aprendizado utilizando
Técnicas Multivariadas**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Acadêmica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof. Michel José Anzanello, *Ph.D.*

Orientador PPGEP/UFRGS

Prof. José Luis Duarte Ribeiro, Dr.

Coordenador PPGEP/UFRGS

Banca Examinadora:

Professor Danilo Marcondes Filho, Dr. (DEST/UFRGS)

Professora Liane Werner, Dr. (PPGEP/UFRGS)

Professor Ricardo Augusto Cassel, Dr. (PPGEP/UFRGS)

AZEVEDO, Bárbara Brzezinski. *Agrupamento de Trabalhadores com Perfis Semelhantes de Aprendizado utilizando Técnicas Multivariadas*, 2013. Dissertação (Mestrado em Engenharia) – Universidade Federal do Rio Grande do Sul, Brasil.

RESUMO

A manufatura de produtos customizados resulta em variedade de modelos, redução no tamanho de lotes e alternância frequente de tarefas executadas por trabalhadores. Neste contexto, tarefas manuais são especialmente afetadas por conta do processo de adaptação do trabalhador a novos modelos de produtos. Este processo de aprendizado pode ocorrer de maneira distinta dentro de um grupo de trabalhadores. Assim, busca-se o agrupamento dos trabalhadores com perfis similares de aprendizado, monitorando a formação de gargalos em linhas de produção constituídas por dissimilaridades de aprendizado em processos manuais. A presente dissertação apresenta abordagens para clusterização de trabalhadores baseadas nos parâmetros oriundos da modelagem de Curvas de Aprendizado. Tais parâmetros, os quais caracterizam o processo de adaptação de trabalhadores a tarefas, são transformados através da Análise de Componentes Principais e então utilizados como variáveis de clusterização. Na sequência, testam-se outras transformações nos parâmetros utilizando funções *Kernel*. Os trabalhadores são clusterizados através do método *K-Means* e *Fuzzy C-Means* e a qualidade dos agrupamentos formados é medida através do *Silhouette Index*. Por fim, sugere-se um índice de importância de variável baseado em parâmetros obtidos na Análise Componentes Principais com o objetivo de selecionar as variáveis mais relevantes para clusterização. As abordagens propostas são aplicadas em um processo da indústria calçadista, gerando resultados satisfatórios quando comparados a clusterizações realizadas sem a transformação prévia dos dados ou sem seleção das variáveis.

Palavras-chave: Curvas de Aprendizado, Clusterização, Análise de Componentes Principais, Transformação de Dados, Seleção de Variáveis

ABSTRACT

Manufacturing of customized products relies on a large menu choice, reduced batch sizes and frequent alternation of tasks performed by workers. In this context, manual tasks are especially affected by workers' adaptation to new product models. This learning process takes place in different paces within a group of workers. This thesis aims at grouping workers with similar learning process tailored to avoid bottlenecks in production lines due to learning dissimilarities among workers. For that matter, we present a method for clustering workers based on parameters derived from Learning Curve (LC) modeling. Such parameters are processed through Principal Component Analysis (PCA), and the PCA scores are used as clustering variables. Next, Kernel transformations are also used to improve clustering quality. The data is clustered using K-Means and Fuzzy C-Means techniques, and the quality of resulting clusters is measured by the Silhouette Index. Finally, we suggest a variable importance index based on parameters derived from PCA to select the most relevant variables for clustering. The proposed approaches are applied in a footwear process, yielding satisfactory results when compared to clustering on original data or without variable selection.

Key words: Learning Curves, Clustering, Principal Components Analysis, Data Transformation, Variable Selection

LISTA DE FIGURAS

Figura 2.1- Passos do Método	24
Figura 2.2- Dendograma AAV	27
Figura 2.3- Dendograma AACP	28
Figura 2.4- Variação dos Índices Silhouette Médios para distintos K s e CPs para o método AACP	29
Figura 2.5- Agrupamentos gerados pelas variáveis latentes oriundas de 3 CPs retidos	31
Figura 3.1- Fluxograma do Método	45
Figura 4.1- Fluxograma do Método proposto	66

LISTA DE TABELAS

Tabela 2.1- Modelos de CAs (adaptado de Wong et al., 2007).	18
Tabela 2.2- Índices Silhouette Médios para os métodos AAV e AACP (de acordo com o número de componentes retidos).....	27
Tabela 2.3 - Autovalores do conjunto de dados e Variância Explicada por cada CP	29
Tabela 2.4- <i>Cluster</i> de destino dos trabalhadores e SI para cada trabalhador através da AACP	30
Tabela 2.5 - Escores do Primeiro CP	31
Tabela 2.6- Fatores e níveis do experimento de simulação.....	32
Tabela 2.7- Resultados da simulação para a escolha do melhor método de agrupamento.....	32
Tabela 3.1- Funções Kernel.....	44
Tabela 3.2- Modelos de CAs (adaptado de Wong et al., 2007)	46
Tabela 3.3- Resultados dos agrupamentos gerados pelos métodos propostos	50
Tabela 4.1- Modelos de CAs (adaptado de Wong et al., 2007)	62
Tabela 4.2- Hierarquia das variáveis de acordo com o IIV	67
Tabela 4.3- SIs médio das clusterizações.....	68

SUMÁRIO

1. INTRODUÇÃO	10
1.1. Considerações Iniciais	10
1.2. Objetivos.....	11
1.3. Justificativa do Tema e dos Objetivos.....	11
1.4. Procedimentos Metodológicos	12
1.5. Estrutura da Dissertação	13
1.6. Delimitações do Estudo	13
1.7. Referências Bibliográficas.....	14
2. PRIMEIRO ARTIGO: AGRUPAMENTO DE TRABALHADORES COM PERFIS SEMELHANTES DE APRENDIZADO APOIADA EM ANÁLISE DE COMPONENTES PRINCIPAIS	15
2.1. Introdução.....	15
2.2. Referencial Teórico	17
2.2.1. Curvas de Aprendizado (CAs)	18
2.2.2. Técnicas de Análise Multivariada.....	21
2.2.2.1. Análise de Componentes Principais (ACP).....	21
2.2.2.2. Análise de Agrupamentos	22
2.3. Metodologia Proposta.....	23
2.4. Aplicação em um processo da indústria calçadista	26
2.5. Experimento de simulação.....	31
2.6. Conclusões.....	33
2.7. Referências Bibliográficas.....	34
Apêndice - Parâmetros CAs	37
3. SEGUNDO ARTIGO: FORMAÇÃO DE GRUPOS HOMOGÊNEOS DE TRABALHADORES BASEADA EM FUZZY C-MEANS E TRANSFORMAÇÃO DE DADOS	38
3.1. Introdução.....	38
3.2. Fundamentação Teórica.....	40
3.2.1. Agrupamentos envolvendo Curvas de Aprendizado	40
3.2.2. Fuzzy C-Means	42
3.2.3. Métodos <i>Kernel</i>	43
3.3. Metodologia Proposta.....	45

3.4. Resultados e Discussão.....	49
3.5. Considerações Finais	51
3.6. Referências	52
Apêndice - Parâmetros CAs	55
4. TERCEIRO ARTIGO: SELEÇÃO DE VARIÁVEIS PARA FORMAÇÃO DE GRUPOS HOMOGÊNEOS DE TRABALHADORES BASEADA EM CURVAS DE APRENDIZADO	56
4.1. Introdução	56
4.2. Fundamentação Teórica.....	58
4.2.1. Métodos de Seleção de Variáveis para Clusterização	58
4.2.2. Análise de Componentes Principais.....	59
4.2.3. A Análise envoltória de dados	60
4.3. Método proposto.....	62
4.4. Resultados e Discussão.....	66
4.5. Conclusões.....	69
4.6. Referências	70
Apêndice - Parâmetros CAs	74
5. Considerações Finais	75

1. INTRODUÇÃO

1.1. Considerações Iniciais

O perfil do consumidor atual demanda cada vez mais produtos personalizados, com elevado padrão de qualidade, entregues de maneira rápida. No entanto, estas exigências devem ser atendidas a baixos custos de produção. Este fenômeno é conhecido como Customização em Massa, o qual se baseia no fornecimento de produtos personalizados em altos volumes a custos baixos (DA SILVEIRA *et al.*, 2001), próximos aos custos de produtos padronizados (MACCARTHY *et al.*, 2003).

Neste cenário, as opções de catálogo tendem a ser numerosas e o tamanho dos lotes reduzido, o que implica em uma alternância frequente de modelos em linha de produção (FOGLIATTO; DA SILVEIRA, 2011). Em processos apoiados majoritariamente em procedimentos manuais, tal alternância submete o trabalhador a um processo de adaptação aos requisitos do novo procedimento. Este processo de adaptação pode ser modelado através de Curvas de Aprendizado (CA). O modelo de Wright (1936), precursor na literatura de CAs, é uma progressão geométrica que expressa a diminuição de tempo necessário para realizar tarefas repetitivas. Modelos posteriores em Anzanello e Fogliatto (2011); Nembhard e Uzumeri (2000); Mazur e Hastie (1978); Towill (1990) corroboram o fato de o desempenho aumentar com a repetição das tarefas.

Cada modelo de CA, descrito por uma equação própria, utiliza parâmetros distintos para quantificar o processo de aprendizado do trabalhador. Tais parâmetros são capazes de informar as particularidades de aprendizado de cada indivíduo, incluindo a velocidade de aprendizado, tipo de tarefa a que o trabalhador melhor se adequa, experiência prévia e estado estacionário de desempenho quando o processo de aprendizado é concluído (UZUMERI; NEMBHARD, 1998). Em uma linha de produção com alternância frequente de tarefas, percebe-se como mandatória a alocação de trabalhadores com características semelhantes de aprendizado ao mesmo grupo de trabalho. Tal prática visa monitorar e evitar a formação de gargalos produtivos e permite melhor monitoramento dos processos executados pelos integrantes do grupo.

A fim de agrupar trabalhadores com perfis similares de aprendizado, esta dissertação aplica métodos de clusterização baseados em transformações de dados através da Análise de Componentes Principais (ACP), funções *Kernel* e seleção de variáveis de aprendizado. A

dissertação é composta por artigos que apresentam as abordagens propostas. No primeiro artigo é apresentada uma sistemática de clusterização baseada nos componentes principais informados pela ACP e a ferramenta *K-Means*. Este artigo também avalia o método proposto frente a variações nos níveis de correlação, presença de ruído nos dados e proporção de observações em relação ao número de variáveis através de um experimento de simulação. O segundo artigo propõe uma sistemática de clusterização com a transformação prévia dos dados através de funções *Kernel* e ACP, e posterior clusterização através do método *Fuzzy C-Means*. O terceiro propõe uma sistemática de seleção de variáveis embasada em um índice de importância gerado através dos pesos oriundos da ACP; a eficiência desta é avaliada através da Análise Envoltória de dados. A qualidade dos agrupamentos formados nos três artigos é avaliada através do *Silhouette Index*. Todas as abordagens propostas foram aplicadas no setor de costura de uma empresa calçadista.

1.2. Objetivos

O objetivo principal da dissertação é propor sistemáticas para formação de agrupamentos homogêneos de trabalhadores em processos que envolvem aprendizado utilizando ferramentas multivariadas.

Como objetivos específicos surgem:

- Integrar a ACP às técnicas de clusterização com vistas ao aumento da qualidade dos agrupamentos formados;
- Avaliar o uso de funções *Kernel* para aprimorar o desempenho das ferramentas de clusterização;
- Verificar o desempenho das sistemáticas propostas através de um experimento de simulação que abranja fatores relevantes para formação de agrupamentos;
- Testar diferentes técnicas de clusterização e avaliar seu desempenho para agrupamento de trabalhadores;
- Selecionar as variáveis de aprendizado mais relevantes para a clusterização através de um índice de importância gerado por meio de parâmetros oriundos da ACP.

1.3. Justificativa do Tema e dos Objetivos

Conforme exposto anteriormente, exigências de mercado levam as empresas a alterar seus sistemas produtivos e modelos de produtos de forma a atender às demandas dos clientes,

exigindo do trabalhador frequente readaptação a novas tarefas. No entanto, o decurso do aprendizado pode não ocorrer de forma igual para todos os trabalhadores e a alocação inapropriada de trabalhadores com diferentes velocidades de aprendizado, experiência prévia e níveis de desempenho a um mesmo grupo (linha de montagem, por exemplo) podem resultar em gargalos produtivos, ociosidade e baixo rendimento em termos quantitativos e qualitativos. Desta forma, o agrupamento de trabalhadores com perfis semelhantes de aprendizado via CAs encontra justificativa prática, visto que permite reduzir as deficiências acima apontadas.

Em termos acadêmicos, a presente abordagem justifica-se por contribuir com um número restrito de abordagens que utilizam curvas de aprendizado com vistas à formação de agrupamentos de trabalhadores com perfis semelhantes de aprendizado. Em uma pesquisa realizada nas bases *Science Direct* e *Web of Knowledge*, utilizando a palavra-chave *Learning Curve*, retornaram 3.522 e 2.097 publicações, respectivamente, desde o ano de 1998 a 2013. Porém, apenas 3 publicações utilizaram CAs com o objetivo de agrupar trabalhadores: UZUMERI; NEMBHARD, 1998; STROIEKE *et al.*, 2011; WONG *et al.*, 2010. Embora tenham endereçado o problema em questão, tais abordagens apoiam-se na análise dos perfis gerados pelas curvas ou na avaliação direta dos parâmetros. Nas proposições desta dissertação, os parâmetros são transformados através de ferramentas multivariadas, o que possibilita aumentar a precisão na formação dos grupos de trabalhadores. Desta forma, justificam-se academicamente as sistemáticas apresentadas nesta dissertação, as quais buscam propor abordagens originais voltadas ao tema agrupamento de trabalhadores.

1.4. Procedimentos Metodológicos

Quanto aos objetivos, essa dissertação é classificada como pesquisa exploratória, dado que permite conhecer o problema e possibilita construir hipóteses para solucioná-lo. Quanto à natureza, é considerada como pesquisa aplicada, tendo em vista que seu conteúdo teórico é explorado e direcionado à solução de problemas genéricos (GIL, 2002). A dissertação é enquadrada como pesquisa quantitativa, pois faz uso de análises numéricas.

1.5. Estrutura da Dissertação

A dissertação está organizada em 5 capítulos. O primeiro capítulo apresenta uma visão geral do tema proposto, destacando a importância do estudo. Ele também justifica o tema abordado, os objetivos do trabalho e o método utilizado.

O segundo capítulo apresenta o primeiro artigo, que propõe uma sistemática de clusterização baseada nos componentes principais oriundos da ACP e a clusterização dos dados através do algoritmo *K-Means*. O método proposto é avaliado frente a variações nos níveis de correlação, presença de ruído nos dados e proporção observações em relação ao número de variáveis através de um experimento de simulação.

O terceiro capítulo traz o segundo artigo, que propõe uma sistemática de clusterização apoiada na transformação prévia dos dados através de funções *Kernel* e agrupamento através do algoritmo *Fuzzy C-Means*. A transformação dos dados através da ACP também é feita, sendo que os componentes principais provenientes desta transformação são clusterizados através do mesmo algoritmo.

O quarto capítulo apresenta o terceiro artigo, o qual propõe uma sistemática de seleção de variáveis embasada em um índice de importância gerado através dos pesos informados pela ACP. A eficiência desta sistemática é avaliada através da Análise Envoltória de dados. Por fim, o quinto capítulo apresenta as conclusões da dissertação, retomando os principais resultados obtidos.

1.6. Delimitações do Estudo

Este trabalho considera o aprendizado proveniente da repetição de uma tarefa manual, desconsiderando a qualidade dos produtos fabricados e o processo de perda de rendimento devido à fadiga. Fatores como dia da semana, dia do mês, turno de trabalho e sexo do trabalhador não foram monitorados na coleta de dados, o que pode influenciar nos parâmetros gerados pela modelagem. Também não é considerada a relação entre o modelo de produto que acaba de ser produzido e o novo produto que entrará em produção sobre a adaptação do trabalhador.

Além disso, não é objeto desta dissertação criar um novo algoritmo de clusterização, mas sim integrar ferramentas multivariadas e sistemáticas de seleção das variáveis a

algoritmos já existentes. O objetivo do trabalho limita-se a métodos não supervisionados de clusterização, não incluindo técnicas de classificação (métodos supervisionados).

1.7. Referências Bibliográficas

- Anzanello, M. J., Fogliatto, F.S., 2011. Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics* 41, 573-583.
- Da Silveira, G., Borestein, D., Fogliatto, F.S., 2001. Mass customization: literature review and research directions. *International Journal of Production Economics* 72 , 1–13.
- Fogliatto, F.S., Da Silveira, G., Borestein, D., 2011. The mass customization decade: An updated review of the literature. *International Journal of Production Economics* 138, 14–25.
- Gil, A.C., 2002. Como elaborar projetos de pesquisa. 4 ed. São Paulo- Atlas.
- MacCarthy, B., Brabazon, Philip G., Bramham, J., 2003. Fundamental modes of operation for mass customization. *International Journal of Production Economics* 85, 289-304.
- Mazur, J.E., Hastie, R., 1978. Learning as accumulation: a reexamination of the learning curve. *Psychology Bulletin* 85 (6),1256–1274.
- Nembhard, D.A., Uzumeri. M.V., 2000. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* 25, 315-326.
- Stroieke, R. E.; Fogliatto, F.S, Anzanello, M.J., 2011. Formação de Agrupamentos Homogêneos de Trabalhadores Através de Curvas de Aprendizado. Pré-anais XLIII Simpósio Brasileiro de Pesquisa Operacional.
- Uzumeri, M.; Nembhard, D.,1998. A Population of Learners: A New Way to Measure Organizational Learning. *Journal of Operations Management, Netherlands*, v. 16, n. 5,515-528.
- Towill, D.R., 1990. Forecasting learning curves. *International Journal of Forecasting* 6 (1), 25 - 38.
- Wong, P.S.P., Cheung, S.O., Wu, R.T.H., 2010. Learning from project monitoring feedback: A case of optimizing behavior of contractors. *International Journal of Project Management* 28, 469 – 481.

2. PRIMEIRO ARTIGO: AGRUPAMENTO DE TRABALHADORES COM PERFIS SEMELHANTES DE APRENDIZADO APOIADA EM ANÁLISE DE COMPONENTES PRINCIPAIS

Bárbara Brzezinski Azevedo, Michel José Anzanello

Universidade Federal do Rio Grande do Sul (UFRGS), Av. Osvaldo Aranha, 99, Porto Alegre, CEP 90.035-190, Rio Grande do Sul, Brasil.

Resumo: A manufatura de produtos personalizados em larga escala, chamada customização em massa, implica no aumento da variedade de modelos e redução no tamanho dos lotes de produção. Tarefas que dependem da habilidade humana são especialmente afetadas nesse contexto, visto que os trabalhadores precisam se adaptar às características do novo modelo. Esse processo de adaptação pode ocorrer de forma distinta dentro de um grupo de trabalhadores, justificando o desenvolvimento de sistemáticas com vistas ao agrupamento de indivíduos com características de aprendizado semelhantes. Este artigo propõe um método para formação de grupos homogêneos de trabalhadores de acordo com seus perfis de aprendizado através da integração de curvas de aprendizado e clusterização. Para tanto, dados de desempenho são coletados e modelados através de curvas de aprendizado; os parâmetros oriundos da modelagem quantificam o processo de adaptação dos trabalhadores às tarefas, servindo de base para o agrupamento dos trabalhadores. Na primeira proposição deste artigo, os parâmetros são clusterizados através do método *K-Means*, e a qualidade dos agrupamentos formados é avaliada por intermédio do Silhouette Index (SI). Em uma proposição alternativa, a Análise de Componentes Principais é aplicada sobre os parâmetros de aprendizado e as variáveis latentes geradas (escores) são clusterizados através do *K-Means*. A clusterização utilizando os escores apresentou um SI 0,968, valor satisfatório comparado ao valor de 0,392 obtido com a clusterização nas variáveis originais. Um estudo de simulação foi realizado para corroborar a eficácia do método proposto, o qual se mostrou robusto quando submetido a diferentes níveis de ruído, correlação e proporções de variáveis e observações.

Palavras-chave: Curvas de Aprendizado, Clusterização, Análise de Componentes Principais

2.1. Introdução

A redução do ciclo de vida dos produtos e o aumento da competição industrial fizeram com que muitas indústrias de massa repensassem suas estratégias de produção, focando-se nas necessidades individuais dos clientes. Assim, no final dos anos 1980, surge o conceito de Customização em Massa (CM), que é baseado no fornecimento de produtos personalizados através de processos flexíveis em altos volumes a custos razoavelmente baixos (DA SILVEIRA *et al.*, 2001), semelhantes aos custos de produtos padronizados (MACCARTHY *et al.*, 2003) e com integração entre os membros da cadeia de suprimentos (FOGLIATTO *et al.*, 2011).

Devido a características específicas, diversos produtos customizados apoiam-se em processos de produção manuais (FOGLIATTO; DA SILVEIRA, 2011), sendo confeccionados

em pequenos lotes (ANZANELLO; FOGLIATTO, 2011a). Esta realidade, relacionada a exigências dos consumidores por produtos diferenciados e maiores opções de catálogo, faz com que os processos produtivos alterem-se frequentemente, expondo constantemente o trabalhador à produção de novos modelos, tecnologias e processos. Tal exposição tende a reduzir os patamares de produção e níveis de qualidade dos produtos nos ciclos iniciais, devido ao processo de adaptação (aprendizado) a que o trabalhador é submetido quando da solicitação de um novo produto (NEMBHARD; UZUMERI, 2000a). Como ferramenta de auxílio à solução de alguns problemas oriundos da CM, a modelagem por Curvas de Aprendizado (CA) vem ganhando importância no contexto produtivo.

CAs são representações matemáticas do desempenho do trabalhador em tarefas repetitivas (FIORETTI, 2007; RIED; MIRKA, 2007; JABER; KAHN, 2010; ANZANELLO; FOGLIATTO, 2011a). CAs têm sido vastamente utilizadas para estudar as alterações de desempenho de trabalhadores relacionados aos efeitos do aprendizado (PLAZA; ROLF, 2008). Assim, a modelagem de CAs pode auxiliar na gestão da produção para a estimação dos tempos de produção, prazos de entrega, desenvolvimento de estratégias de produção mais precisas (ANZANELLO; FOGLIATTO, 2011a) e classificação de grupos de trabalhadores de acordo com seus perfis de adequação a tarefas de complexidades distintas (UZUMERI; NEMBHARD, 1998; NEMBHARD; UZUMERI, 2000a; STROIEKE *et al.*, 2011).

O agrupamento de trabalhadores em equipes tipicamente apoia-se nos tempos médios de execução de tarefas, visando minimizar a variância entre estes. Porém, em cenários de CM caracterizados por produção de lotes em tamanhos reduzidos, a formação das equipes pode ser mais eficiente considerando os perfis de aprendizado, os quais capturam o desempenho do trabalhador desde o início do processo analisado - e não somente quando o processo de aprendizado está concluído (ANZANELLO; FOGLIATTO, 2007a). Tal agrupamento permite a alocação de trabalhadores com padrões similares de produtividade à mesma equipe, reduzindo níveis de ociosidade (ANZANELLO; FOGLIATTO, 2007a).

Apesar da grande importância e aplicabilidade do tema, encontram-se na literatura poucos estudos que se valem das CAs endereçando este problema de pesquisa. Uzumeri e Nembhard (1998) identificaram perfis distintos de trabalhadores em uma empresa de manufatura americana através da análise dos parâmetros gerados pela modelagem das CAs, evidenciando as diferenças entre os trabalhadores para adequação a tarefas. Wong *et al.* (2010) avaliaram o padrão de desempenho de empreiteiros submetidos a *feedbacks* regulares e identificaram perfis diferentes dentre os trabalhadores, os quais foram agrupados de acordo com os parâmetros da CA. Com propósitos semelhantes, Stroieke *et al.* (2011) propuseram

um método para caracterizar diferentes grupos de trabalhadores utilizando curvas de aprendizado e técnicas de clusterização nos dados de desempenho dos trabalhadores. Porém, os estudos apresentados não supriram todas as lacunas existentes na literatura referentes ao agrupamento de trabalhadores segundo seus perfis de aprendizado, principalmente com vistas à utilização de técnicas multivariadas como ferramentas de suporte.

Assim, este artigo propõe uma sistemática para a formação de grupos homogêneos de trabalhadores de acordo com seus padrões de aprendizado em linhas de produção customizadas. O método inicia com a coleta de dados de desempenho de trabalhadores, os quais são ajustados a modelos distintos de CAs. Os parâmetros gerados, os quais descrevem o processo de aprendizado dos trabalhadores, são remapeados através da Análise de Componentes Principais (ACP) visando aprimorar a qualidade do procedimento de clusterização. A qualidade de agrupamentos formados é avaliada através do Índice Silhouette (SI), conforme proposto por Rousseeuw (1987). Quando aplicado em um processo da indústria calçadista, o método proposto apresenta um incremento na qualidade dos agrupamentos medida através do SI de 0,392 para 0,968 quando comparado à clusterização aplicada às variáveis originais. Para corroborar a eficácia do método, realizou-se um experimento com simulação avaliando-se distintos níveis de ruído, correlação e proporções de variáveis por observações.

Este artigo apresenta duas contribuições principais: (i) propõe um método de agrupamento de trabalhadores balizado por perfis de aprendizado que apresenta qualidade superior à clusterização tradicional, e (ii) permite agrupamentos nos quais há poucas observações coletadas (poucos trabalhadores) e um número elevado de variáveis (parâmetros oriundos das CAs) a serem consideradas na clusterização.

O restante do artigo é organizado como segue. A seção 2.2 apresenta uma breve introdução sobre CAs, ACP e técnicas para análise da qualidade de formação de agrupamentos. Na seção 2.3 é apresentado o método para formação de equipes balizado pelo aprendizado dos trabalhadores. A seção 2.4 apresenta os resultados obtidos com o método e posterior aplicação da sistemática recomendada em um estudo de caso na indústria calçadista. A seção 5 traz a conclusão do artigo e sugestões para pesquisas futuras.

2.2. Referencial Teórico

Nesta seção, revisam-se os fundamentos das ferramentas utilizadas na sistemática proposta: CAs e técnicas de análise multivariada para formação de agrupamentos.

2.2.1. Curvas de Aprendizado (CAs)

CAs são modelos de regressão não-lineares que associam o desempenho de trabalhadores (normalmente medido em número de unidades produzidas em determinado intervalo de operação) a características das tarefas (ANZANELLO; FOGLIATTO, 2007b). Sua utilização para a formação de agrupamentos foi proposta por Uzumeri e Nembhard (1998), Wong *et al.* (2010) e Stroeike *et al.* (2011). Os principais modelos de CAs são classificados em: (i) potenciais, (ii) exponenciais e (iii) hiperbólicos, e suas fundamentações teóricas são detalhadas na sequência. A Tabela 2.1 apresenta um resumo dos modelos, a descrição de seus parâmetros e suas aplicações. Detalhes dos modelos são apresentados a seguir.

Tabela 2.1- Modelos de CAs (adaptado de Wong *et al.*, 2007).

Modelo	Equação	Parâmetros	Aplicações	Referências
Modelos Potenciais				
Wright	$Y = Y_1 x^b$ (1)	Y = tempo demandado para a repetição x da tarefa; Y_1 = tempo de execução da primeira repetição; b = expoente de aprendizagem, onde $-1 < b < 0$, $b = \frac{\log \emptyset}{\log 2}$ e \emptyset indica a taxa de aprendizado expressa percentualmente.	Expressar o efeito da aprendizagem no decréscimo do custo médio (ou tempo) ao longo das repetições de tarefas.	Wright (1936).
Stanford-b	$Y = Y_1(x + B)^b$ (2)	B = unidades equivalentes à experiência prévia do trabalhador. *	Incorporar a experiência prévia do trabalhador no início de um ciclo de produção.	Anzanello e Fogliatto (2011); Yeh e Rubin (2012).
DeJong	$Y = Y_1[M + (1 - M)x^b]$ (3)	M = fator de incompressibilidade, onde $0 < M < 1$, representa a proporção do tempo total de operação constituído por procedimentos automatizados. Quando $M = 0$, o modelo reduz-se ao modelo de Wright.*	Incorporar a influência de maquinário no processo de aprendizado.	Anzanello e Fogliatto, (2011).
Curva S	$Y = Y_1[M + (1 - M)(x + B)^b]$ (4)	Os parâmetros assumem o mesmo significado dos modelos em que a equação foi baseada.	Descrever o processo de aprendizagem quando há a intervenção de maquinário.	Nembhard e Uzumeri(2000b); Anzanello e Fogliatto (2011).
Plateau	$Y = C + Y_1 x^b$ (5)	C = constante aditiva que descreve o estado estacionário do aprendizado.*	Descrever o estado estacionário do processo de aprendizado.	Anzanello e Fogliatto (2011).
Modelos Exponenciais				
Knecht	$Y = Y_1 x^b e^{c'x}$ (6)	C' = segunda constante. *	Modelar processos produtivos que apresentam elevado número de repetições.	Anzanello e Fogliatto (2011).
Exponencial de 3 Parâmetros	$Y = k(1 - e^{-(x+p)/r})$ (7)	Y = desempenho do trabalhador em itens produzidos; x = tempo de operação na tarefa; k = patamar máximo de desempenho a ser atingido (expresso em número de unidades produzidas por tempo de operação); p = experiência prévia do trabalhador na execução da tarefa; r = a taxa de aprendizado do trabalhador, onde: $Y, x, k \geq 0$ e x, p e r são dados em unidades de tempo.	Adequa-se bem a cenários onde os trabalhadores possuem experiência prévia na tarefa.	Mazur e Hastie (1978); Anzanello e Fogliatto (2011).
Tempo constante	$Y = y_c + y_f(1 - e^{-t/\tau})$ (8)	y_c = desempenho inicial; y_f = patamar máximo de aprendizado; t = tempo acumulado de operação, onde y_c e y_f são	Expressar os efeitos do aprendizado, após a adaptação do	Towill (1990); Dardan et al. (2006);

		dados em unidades de tempo, sendo que t possui o mesmo significado do número de unidades produzidas adotado pelos demais modelos de CAs.	trabalhador à tarefa.	Anzanello e Fogliatto (2011).
Modelos Hiperbólicos				
Hiperbólico de 2 Parâmetros	$Y = k\left(\frac{x}{x+r}\right)$ (9)	Y= o número de unidades produzidas em um dado intervalo de tempo; x = intervalo de tempo considerado; r = parâmetro de aprendizado, expresso em unidades de tempo; k = nível máximo de aprendizado, expresso em número de unidades produzidas por tempo de operação.	Modelar a CA de acordo o número de unidades produzidas em conformidade com o número total de unidades.	Mazur e Hastie (1978).
Hiperbólico de 3 Parâmetros	$Y = k\left(\frac{x+p}{x+r+p}\right)$ (10)	p = parâmetro que representa experiência prévia do trabalhador na execução da tarefa, expresso em unidades de tempo*.	Possui a mesma aplicação que o modelo anterior, porém considera a experiência prévia do trabalhador.	Mazur e Hastie (1978).

*Os demais parâmetros apresentam as mesmas definições anteriores.

O primeiro modelo potencial foi desenvolvido por Wright (1936), após observar que os custos de montagem de aviões reduziam-se em 20% a cada duplicação da quantidade de aeronaves produzidas, conforme a equação (1). Devido à simplicidade matemática e capacidade de ajuste a dados empíricos, o modelo de Wright encontrou vasta aplicação prática: Jaber *et al.* (2010) o utilizaram para analisar uma cadeia de suprimentos na qual a produção é submetida a processos de melhoria contínua, concluindo que a inclusão dos efeitos da aprendizagem no processo produtivo traz benefícios para toda a cadeia, reduzindo significativamente os custos associados ao abastecimento. Zorgios *et al.* (2009), utilizando o mesmo modelo de CA, verificaram que flutuações nas taxas de aprendizado provêm da variabilidade de capital humano e que os custos envolvidos em um ambiente de conhecimento podem ser racionalizados e monitorados através da análise da CA. Outras aplicações do modelo de Wright podem ser encontradas em Jaber e Glock (2013) e Yeh e Rubin (2012).

Devido a limitações em situações práticas, o modelo de Wright sofreu alterações ao longo tempo. O modelo Stanford-b acrescenta um parâmetro ao modelo da equação (1) para representar as unidades equivalentes à experiência do trabalhador no início do ciclo de produção; ver equação (2) (Yeh e Rubin, 2012). Para representar processos que apresentam tarefas automatizadas, desenvolveu-se o modelo de DeJong, no qual a velocidade das operações é influenciada pelo maquinário que limita a capacidade do trabalhador em reduzir o tempo por unidade produzida, conforme a equação (3) (NEMBARD; UZUMERI, 2000b). Contemplando os parâmetros envolvidos nos modelo de Stanford-b e de DeJong, foi proposta a Curva S, a qual é baseada na suposição de que o início do processo de aprendizagem é mais gradual do que representado pelo modelo de Wright, ver equação (4) (NEMBARD; UZUMERI, 2000b).

Com vistas a contornar o comportamento assintótico do modelo de Wright quando repetições tendem ao infinito, o modelo de Plateau insere uma constante aditiva C para descrever o estado estacionário de desempenho do trabalhador, fazendo com que o tempo/unidade tenda àquele valor quando o volume de produção for elevado, conforme a equação (5). O estado estacionário é atingido quando o processo de aprendizado é encerrado ou quando há limitações de maquinário restringindo a melhoria de desempenho (ANZANELLO; FOGLIATTO, 2011a).

Com o objetivo de aprimorar as previsões em corridas de produção de longa duração, surgiram os modelos exponenciais. O primeiro deles, na equação (6), foi proposto por Knecht (1974) objetivando aprimorar a modelagem de processos que possuem um elevado número de repetições (NEMBHARD; UZUMERI, 2000b). O modelo propõe a utilização combinada de funções exponenciais e potenciais, conforme apresentado na Tabela 2.1.

Mazur e Hastie (1978) propuseram o modelo Exponencial de 3 Parâmetros (equação (7)) para expressar a aprendizagem em três frentes: (i) patamar máximo de desempenho a ser atingido pelo trabalhador quando a aquisição de conhecimento for completa; (ii) experiência prévia do trabalhador na execução da tarefa observada e (iii) taxa de aprendizado, a qual quantifica a velocidade de aquisição de destreza (as unidades dos parâmetros são detalhadas na Tabela 2.1). Estudos realizados pelos autores demonstram que o modelo apresenta ajustes deficientes quando o trabalhador é submetido a tarefas complexas e que exigem grande quantidade de novos conhecimentos. Por outro lado, o modelo adequa-se bem a situações nas quais o trabalhador possui experiência prévia (ANZANELLO; FOGLIATTO, 2007a).

Towill (1990) propôs o modelo de Tempo Constante estruturado de forma análoga ao Exponencial de 3 Parâmetros, conforme a equação (8). O autor recomenda a utilização deste modelo quando a coleta de dados de desempenho inicia depois de uma curta adaptação do trabalhador à tarefa. Dardan *et al.* (2006) utilizaram este modelo para avaliar o tempo de vida de um investimento em tecnologia, considerando os impactos do aprendizado.

Mazur e Hastie (1978) propuseram um modelo de CA que relaciona o número de unidades produzidas em conformidade com o número total de unidades produzidas. Sua descrição matemática consistiu em uma curva hiperbólica de 2 parâmetros (resultando no modelo Hiperbólico de 2 Parâmetros), conforme a equação (9). Buscando considerar a experiência prévia no trabalhador na execução da tarefa, Mazur e Hastie (1978) propuseram a inclusão de um parâmetro p na equação do modelo Hiperbólico de 2 Parâmetros, dando

origem ao modelo Hiperbólico de 3 Parâmetros, conforme equação (10). Tal modelo tem sido amplamente aplicado em cenários produtivos: Wong *et al.* (2007) utilizaram o modelo para prever o desempenho de contratantes durante um projeto de construção civil, enquanto que Guimarães *et al.* (2012) obtiveram uma redução significativa no número de acidentes e no índice de absenteísmo em uma indústria de calçados utilizando uma metodologia de rotação de trabalhadores a partir da aplicação do modelo Hiperbólico de 3 Parâmetros.

2.2.2. Técnicas de Análise Multivariada

Esta seção revisa as duas técnicas de análise multivariada utilizadas neste artigo: (i) análise de componentes principais e (ii) análise de agrupamentos (*clustering*).

2.2.2.1. Análise de Componentes Principais (ACP)

A ACP é um método utilizado para reduzir a dimensionalidade de uma matriz de dados, preservando a maior parte da variabilidade existente no conjunto de dados (Rencher, 1995). Sua operacionalização ocorre através do cálculo dos autovalores e dos autovetores da matriz de covariância dos dados originais (RENCHE, 1995; JOLLIFFE, 2002). Os componentes principais (CPs) resultantes, Z_1, Z_2, \dots, Z_n , são combinações lineares das variáveis originais, X_1, X_2, \dots, X_p , $z_{ni} = w_{1n} x_{1i} + w_{2n} x_{2i} + \dots + w_{pn} x_{pi}$, e são mutuamente ortogonais (Rencher, 1995). Devido à ortogonalidade, é possível examinar as relações entre o conjunto de dados e as variáveis através dos gráficos dos escores (z) e dos pesos (w), cujos eixos são CPs nos quais os dados são projetados. Os escores fornecem a composição dos CPs em relação às amostras de dados, enquanto que os pesos fornecem a mesma composição em relação às variáveis, pois os pesos decorrem da maximização da variância entre as variáveis latentes Z (JOLLIFFE, 2002). O número de CPs a ser retido pode ser definido com base no percentual de variância explicado por cada componente. Assim, somam-se os autovalores – arranjados do maior ao menor – até atingir o valor esperado de variância (RENCHE, 1995); pode-se também utilizar o *Scree Graph* nesta análise (RENCHE, 1995).

Dentre suas diversas aplicações, Mao *et al.* (2009) utilizaram a ACP para estruturar um questionário em componentes, buscando investigar os fatores em cada componente que influenciam na indústria automotiva chinesa, enquanto que Engblom *et al.* (2012) fizeram uso da ACP para agrupar custos logísticos em componentes e verificar a influência destes nos custos totais de companhias finlandesas. Lu *et al.* (2011) utilizaram CPs e análise de conglomerados para agrupar estações de monitoramento da qualidade do ar com base nos

níveis de poluição em Hong Kong, verificando que a utilização combinada dos dois métodos possui elevada aplicação na gestão de sistemas com muitas variáveis. Gong *et al.* (2012) também utilizaram os dois métodos a fim de avaliar as relações existentes entre a concentração de elementos químicos presentes na substância própolis. Os métodos permitiram a formação de agrupamentos homogêneos e maior conhecimento sobre as variações das características observadas nas substâncias entre as áreas estudadas.

2.2.2.2. Análise de Agrupamentos

A análise de agrupamentos ou análise de conglomerados (*cluster analysis*) é um método estatístico que permite agrupar observações em grupos homogêneos em função do grau de similaridade entre as observações (HAIR *et al.*, 1998; FÁVERO *et al.*, 2009). Os agrupamentos resultantes devem possuir alta homogeneidade interna e alta heterogeneidade externa; quando representados graficamente, as observações dentro dos agrupamentos estarão próximas, enquanto que as observações de diferentes grupos estarão distantes (HAIR *et al.*, 1998). Na análise de agrupamentos, a variável de agrupamento (conjunto de variáveis que representam as características utilizadas para comparar observações) é especificada pelo pesquisador, não sendo estimada empiricamente (HAIR *et al.*, 1998; FÁVERO *et al.*, 2009).

A análise de clusterização demanda que se determine o procedimento para agrupamento (hierárquico ou não hierárquico). Os procedimentos hierárquicos envolvem a construção de uma estrutura do tipo árvore, denominada dendograma (HAIR *et al.*, 1998); neste procedimento não é necessário determinar previamente o número de *clusters*. Os procedimentos não hierárquicos, por sua vez, demandam a definição do número inicial de *clusters* para inserção das observações em grupos (FÁVERO *et al.*, 2009). O método *K-Means* é um dos procedimentos não hierárquicos mais conhecidos (NALDI *et al.*, 2011) e utiliza a distância euclidiana como critério para formação de grupos (FÁVERO *et al.*, 2009). É considerado um método simples e pode ser aplicado a um grande conjunto de dados (NALDI *et al.*, 2011). O *K-Means* é operacionalizado com base nos seguintes procedimentos: (i) partição inicial dos indivíduos em *K clusters* definidos pelo pesquisador; (ii) cálculo dos centroides para cada *cluster* e cálculo da distância euclidiana dos centroides para cada observação da base de dados; e (iii) agrupamento das observações aos *clusters* cujos centroides sejam mais próximos; (iv) voltar ao passo (ii) até que não ocorra variação significativa na distância mínima de cada observação a cada um dos centroides dos *K clusters* (NALDI *et al.*, 2011).

Para medir a qualidade dos agrupamentos formados, Rousseeuw (1987) apresenta o índice Silhouette (SI), que mede o grau de similaridade de uma observação em relação às demais em seu próprio grupo, comparada às observações alocadas ao grupo mais próximo a ela. O SI é calculado através da equação (11), onde $a(i)$ é média da distância da i -ésima observação a todas as demais pertencentes ao grupo no qual a observação foi alocada e $b(i)$ é a média da distância entre a i -ésima observação a todas as observações do grupo vizinho mais próximo.

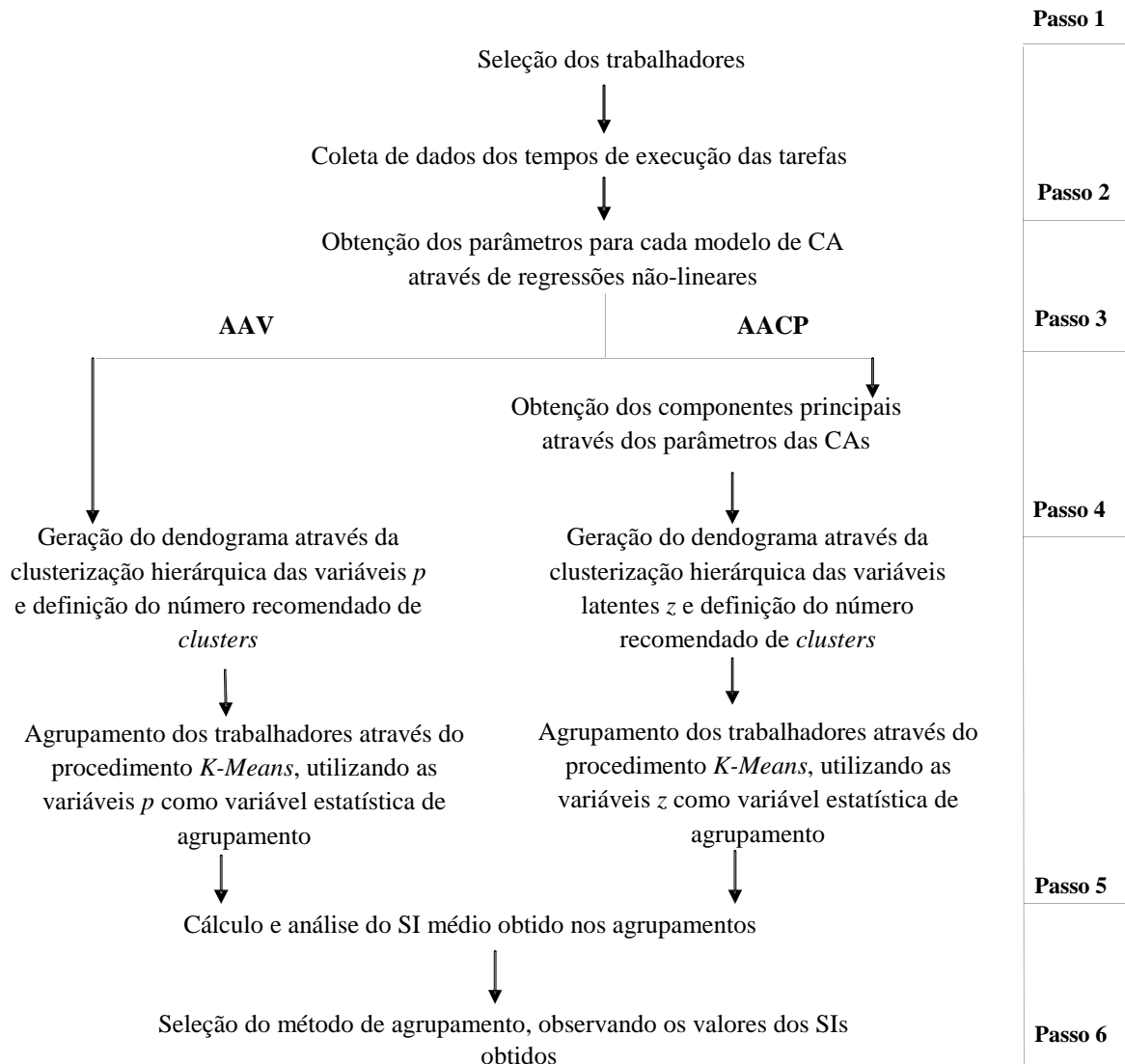
$$SI = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}} \quad (11)$$

Os valores de SI estão no intervalo de -1 a 1, sendo que valores próximos a -1 indicam que a observação foi erroneamente inserida no grupo de destino; valores de próximos a zero denotam que a observação poderia estar tanto em seu grupo de destino quanto em algum outro grupo; e valores próximos a 1 indicam que a observação está corretamente alocada em seu grupo (ROUSSEEUW, 1987). Dentre as suas aplicações, o SI foi utilizado por Brida *et al.* (2012) para selecionar o melhor método de agrupamento de turistas no mercado italiano, enquanto que Jung e Koh (2009) o utilizaram para medir a qualidade de agrupamentos, propondo uma metodologia para a classificação e localização de danos estruturais em engenharia. Saha *et al.* (2011) aplicaram o SI para a avaliação do desempenho de algoritmos de clusterização criados para encontrar grupos de genes. Por fim, Stroiecke *et al.* (2011) valeram-se do SI para avaliação a qualidade dos agrupamentos de trabalhadores formados a partir dos dados de desempenho obtidos por CAs.

2.3. Metodologia Proposta

O método proposto visa agrupar equipes de trabalhadores de acordo com seus perfis de aprendizado em ambientes de CM. O tempo de processamento das tarefas depende da velocidade de aprendizado dos trabalhadores e, por consequência, impacta na formação de equipes. A Figura 2.1 representa as etapas do método proposto, as quais são detalhadas na sequência. A sistemática proposta apoia-se em dois processos distintos de clusterização: Análise de Agrupamentos sobre Variáveis de Clusterização p (AAV) e Análise de Agrupamentos sobre Componentes Principais (AACP).

Figura 2.1- Passos do Método



Passo 1: Seleção dos trabalhadores

Este passo inicia com a seleção de I ($i=1, \dots, I$) trabalhadores que terão seus processos de aprendizado monitorados. É desejado que cada trabalhador conheça os princípios das tarefas. No momento em que o trabalhador é apresentado a uma nova tarefa, deve-se cronometrar o tempo demandado para execução de cada repetição. Os dados de desempenho devem ser coletados até que não sejam observadas maiores alterações no padrão de desempenho. A coleta de dados é realizada através da contagem do número de itens produzidos em determinado intervalo de operação.

Passo 2: Obtenção dos parâmetros para cada equação de CA

Os dados de desempenho dos trabalhadores são então modelados através das CAs apresentadas na Tabela 2.1. Os parâmetros das CAs podem ser estimados através de rotinas disponíveis em aplicativos de análise estatística. Assim, cada trabalhador i é caracterizado por um conjunto de parâmetros estimados a partir dos dados de desempenho obtidos através da modelagem de cada CA. Tais parâmetros são normalizados para evitar efeitos de escala no processo de agrupamento. Os parâmetros passam a ser considerados variáveis de clusterização p .

Passo 3: Obtenção dos componentes principais através dos parâmetros das CAs

Neste passo, aplica-se a ACP sobre os parâmetros, que resultará em variáveis latentes z . O número de CPs a ser retido pode ser definido com base no *Scree Graph* (Rencher, 1995). Após a conclusão deste passo, haverá duas opções de variáveis de clusterização: p (parâmetros originais) e z (parâmetros no formato de variáveis latentes).

Passo 4: Geração do dendograma através da clusterização hierárquica das variáveis p e z

Iniciam-se duas sistemáticas paralelas de agrupamento dos trabalhadores: AAV e AACP. No processo AAV, estima-se o número recomendado de agrupamentos de acordo com o dendograma resultante e executa-se a clusterização hierárquica utilizando as variáveis p através do método *K-Means*. A combinação dos métodos hierárquicos e não hierárquicos é feita para se obter os benefícios de ambos os métodos: o método hierárquico permite estimar um número adequado de agrupamentos de acordo com as características das observações avaliadas, sem que o pesquisador tenha de fazer esta escolha previamente, como ocorre no procedimento não hierárquico. Porém, a precisão de agrupamento é maior em métodos não hierárquicos (FÁVERO *et al.*, 2009). Na segunda sistemática de clusterização, AACP, extraem-se variáveis latentes z dos parâmetros de aprendizado para então realizar o agrupamento sobre tais variáveis, o que pode reduzir o efeito de correlação das variáveis sobre a clusterização. Na sequência, executa-se a clusterização hierárquica sobre as variáveis z e verifica-se o número recomendado de agrupamentos de acordo com o dendograma resultante. O número estimado de agrupamentos obtido é gerado através do método *K-Means* utilizando as variáveis latentes z .

Passo 5: Cálculo e análise do SI médio obtido nos agrupamentos

A qualidade dos agrupamentos gerados é avaliada através da média de SIs dos I trabalhadores considerados. Devem ser gerados dois SIs médios, um para a sistemática AAV e outro para AACP. Os resultados obtidos são armazenados para comparações posteriores.

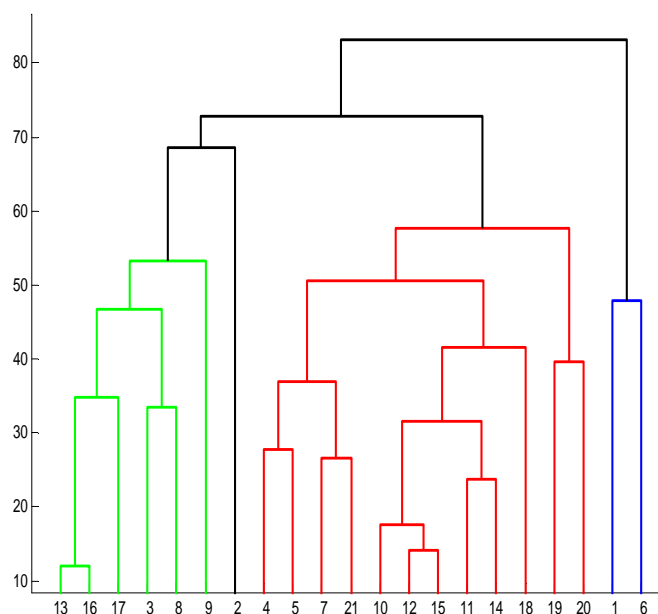
Passo 6: Seleção do método de agrupamento

Por fim, comparam-se os valores de SI obtidos no passo anterior. O objetivo é identificar a sistemática de clusterização responsável pelo maior SI, a qual é recomendada para futuros procedimentos de clusterização.

2.4. Aplicação em um processo da indústria calçadista

Um processo da indústria calçadista foi utilizado como exemplo numérico para avaliação do método proposto. O setor calçadista é caracterizado por uma crescente tendência de CM evidenciada pela redução no tamanho dos lotes de produção e alternância frequente de modelos, fazendo com que a produção seja dependente da habilidade manual dos trabalhadores (Anzanello e Fogliatto, 2011b).

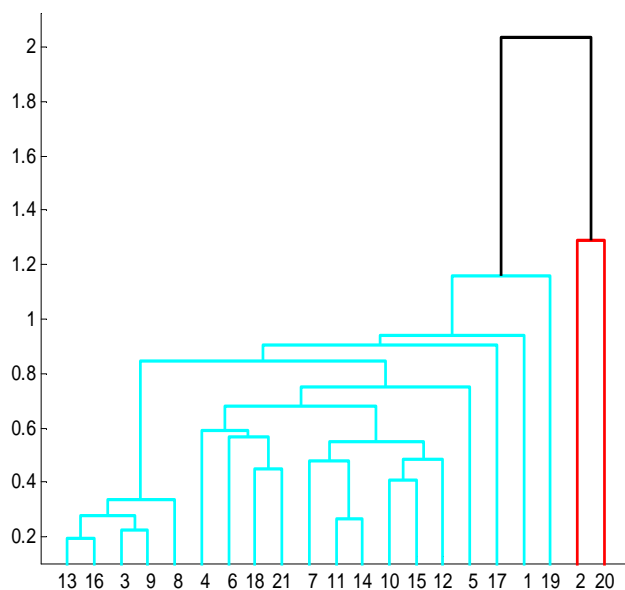
Dados de desempenho de 21 trabalhadores foram coletados e modelados de acordo com as equações da Tabela 2.1. Foram gerados 29 parâmetros, apresentados no Apêndice, que representam os perfis de aprendizado dos trabalhadores. Tais parâmetros foram padronizados e utilizados como variáveis de clusterização p , de acordo com os passos 1 e 2 descritos no método. Na sequência, aplicou-se a ACP sobre os parâmetros originais, gerando as variáveis latentes z . Para executar a sistemática AAV, aplicou-se uma rotina de clusterização hierárquica sobre as variáveis p utilizando a distância euclidiana e o algoritmo *Average Linkage* para agrupar as observações (trabalhadores). Este algoritmo utiliza como critério de agrupamento as distâncias médias entre todos os pares constituídos pelas observações pertencentes a um e outro grupo; não depende de valores extremos e a decisão de agrupamento considera todos os elementos de um aglomerado. Assim, este procedimento tende a formar aglomerados com aproximadamente a mesma variância interna (HAIR *et al.*, 1998). O dendograma resultante sugeriu a formação de 4 *clusters*, conforme a Figura 2.2. Na sequência, as mesmas variáveis foram clusterizadas através do método *K-Means*, utilizando $K=4$. O valor de SI médio resultante na AAV foi de 0,392, conforme a Tabela 2.2.

Figura 2.2- Dendograma AAV

A seguir, as variáveis latentes z foram utilizadas para dar início à AACP. A clusterização hierárquica foi realizada para estimar o número ideal de *clusters*; utilizando o mesmo algoritmo (*Average Linkage*) e métrica de distância (euclidiana) utilizados para gerar o dendograma na AAV. O dendograma resultante sugere a formação de 2 *clusters*, conforme a Figura 2.3. A partir dos CPs gerados, utilizaram-se as variáveis latentes z para proceder ao método *K-Means*, utilizando $K=2$. O valor de SI médio resultante na AACP foi de 0,968, retendo-se um componente principal, explicando 75% da variância na análise, conforme Tabela 2.2.

Tabela 2.2- Índices Silhouette Médios para os métodos AAV e AACP (de acordo com o número de componentes retidos)

Número de Componentes Principais Retidos	Variância Acumulada Explicada	AACP	AAV
1	75,00%	0,968	0,392
2	89,98%	0,921	
3	95,74%	0,906	
4	97,80%	0,902	
5	98,82%	0,899	
6	99,36%	0,898	

Figura 2.3- Dendograma AACP

A Tabela 2.2 apresenta a média dos valores de SIs obtidos pelos 21 trabalhadores agrupados nos dois métodos analisados. Os valores médios de SIs da AACP foram obtidos fixando-se $K=2$ e alterando-se o número de componentes principais retidos na análise. Reteu-se 1 CP na análise, pois apresentou o maior valor de SI. Com base em tais resultados, o método de agrupamento recomendado é a AACP com retenção de 1 CP, o qual gera o maior SI médio.

A Figura 2.4 apresenta a variação nos valores médios de SIs para diferentes valores de K e componentes principais retidos. Percebe-se que, para os dados em análise, a retenção de 1 CP e a utilização de 2 *clusters* na análise gera o maior valor de SI. Tal resultado corrobora a sugestão de 2 *clusters* indicada pelo dendograma da Figura 2.3. Percebe-se ainda uma tendência de redução na qualidade de clusterização com a inclusão de maior número de componentes principais, o que pode ser justificado pela inserção de variáveis latentes de menor importância no procedimento de clusterização (o que acaba por comprometer a estrutura dos *clusters* formados).

0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%
0	0,00%	100,00%

A AACP permitiu a formação de agrupamentos homogêneos, conforme valores individuais de SI na Tabela 2.4. Os trabalhadores 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,19 e 21 foram inseridos no *cluster* 1, ao passo que os trabalhadores 2 e 20 compõem o *cluster* 2. Tendo em vista que todos os trabalhadores geraram SI próximo a 1, pode-se afirmar que a clusterização foi adequadamente realizada.

Tabela 2.4– *Cluster* de destino dos trabalhadores e SI para cada trabalhador através da AACP

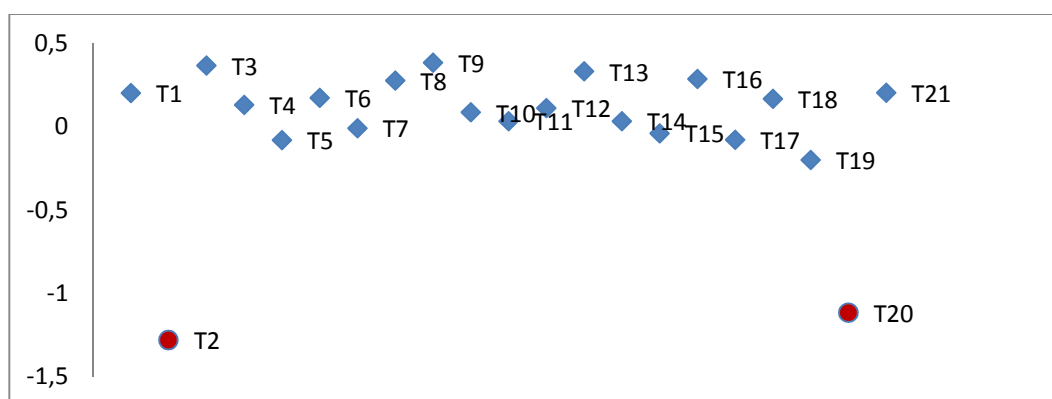
Trabalhador	Cluster	SI
1	1	0,983
2	2	0,986
3	1	0,964
4	1	0,985
5	1	0,943
6	1	0,984
7	1	0,968
8	1	0,976
9	1	0,961
10	1	0,983
11	1	0,976
12	1	0,984
13	1	0,969
14	1	0,976
15	1	0,959
16	1	0,975
17	1	0,944
18	1	0,985
19	1	0,862
20	2	0,983
21	1	0,983
SI médio		0,968

A Tabela 2.5 apresenta os escores, z , do primeiro CP obtido com a AACP. A primeira coluna representa o trabalhador a ser agrupado e na segunda, o respectivo escore do primeiro componente principal. Pode-se verificar que os trabalhadores 2 e 20 apresentam valores de escores bastante distintos dos demais, o que é melhor visualizado graficamente na Figura 2.5.

Tabela 2.5 - Escores do Primeiro CP

Trabalhador	Score do 1º CP
T1	0,2025
T2	-1,2782
T3	0,3682
T4	0,1314
T5	-0,0801
T6	0,1743
T7	-0,0082
T8	0,2781
T9	0,3844
T10	0,0867
T11	0,0336
T12	0,1114
T13	0,3324
T14	0,0333
T15	-0,0389
T16	0,2868
T17	-0,0777
T18	0,1686
T19	-0,1989
T20	-1,1133
T21	0,2035

Os *clusters* apresentados na Tabela 2.4 são ilustrados na Figura 2.5, onde se utiliza o primeiro CP para geração de um gráfico. Como o CP resultante é uma combinação linear dos parâmetros de aprendizado, os aglomerados formados no gráfico permitem identificar dois grupos de trabalhadores com perfis distintos de aprendizado.

Figura 2.5- Agrupamentos gerados pelas variáveis latentes oriundas de 1º CP

2.5. Experimento de simulação

Devido ao satisfatório desempenho gerado pelo método AACCP frente ao AAV no exemplo numérico aplicado em dados industriais, realizaram-se experimentos de simulação com o objetivo de avaliar o comportamento do método diante da inclusão de fatores como ruído, correlação nos dados e diferentes proporções entre o número de variáveis e de

observações. No exemplo numérico da seção 2.4, a AACP apresentou maior desempenho com a retenção de 1 CP e a utilização de $K=2$ clusters. Sendo assim, estes fatores foram mantidos fixos na simulação para avaliação do método. Todas as simulações são baseadas no conjunto de dados utilizado no exemplo numérico. Os experimentos foram realizados em Matlab 7.8.

Os dados foram gerados através de uma distribuição multinormal, combinando os níveis de três fatores analisados no experimento, apresentados na Tabela 2.6. O nível nominal de variância do ruído gera vetores de médias obtidos através de uma distribuição uniforme no intervalo $[0;0,8]$. O nível baixo corresponde a $[0;0,1]$, enquanto o nível alto a $[0,5;0,8]$. Tais níveis foram definidos de forma empírica. Para o fator de correlação, foi gerada a matriz de correlação (Γ) a partir dos dados originais; Γ^3 corresponde ao nível baixo e Γ ao nível alto. A potência ímpar foi utilizada para preservar o sinal das correlações entre as variáveis. Há três níveis de razão entre observações e variáveis: (i) razão de 0,72 (21 observações /29 variáveis), denotando o nível baixo, (ii) razão de 1,0 (29 observações/29 variáveis), denotando o nível médio, e (iii) razão de 7,2 (210/29), correspondendo ao nível alto. Foram gerados $3 \times 2 \times 3 = 18$ cenários diferentes para cada método de agrupamento (AACP e AAV) e simuladas 1.000 repetições para cada cenário.

Tabela 2.6- Fatores e níveis do experimento de simulação

Fator	Nível
Variância do ruído	$[0;0,1]$, $[0,2;0,4]$, $[0,5;0,8]$
Correlação dos parâmetros	$\Gamma^3; \Gamma$
Proporção de observações por variáveis	0,72; 1; 7,2

A Tabela 2.7 apresenta a média e desvio-padrão das 1000 repetições para cada cenário avaliado. Percebe-se que a AACP apresenta desempenho superior frente à AAV quando da inserção de ruído, correlação e diferentes razões entre observações e variáveis em todos os níveis dos fatores inseridos na simulação.

Tabela 2.7– Resultados da simulação para a escolha do melhor método de agrupamento

Razão de Observações por Variáveis	Correlação	Ruído	AACP		AAV	
			SI médio	Desvio-Padrão	SI médio	Desvio-Padrão
0,72	Γ^3	$[0;0,1]$	0.7417	0.0684	0.2826	0.0590
		$[0,2;0,4]$	0.7428	0.0693	0.2854	0.0655
		$[0,5;0,8]$	0.7386	0.0729	0.2821	0.0604
	Γ	$[0;0,1]$	0.7391	0.0713	0.3528	0.0671
		$[0,2;0,4]$	0.7386	0.0678	0.3519	0.0702
		$[0,5;0,8]$	0.7355	0.0728	0.3534	0.0679

1	Γ^3	[0;0,1]	0.7304	0.0571	0.3030	0.0522
		[0,2;0,4]	0.7400	0.0560	0.2612	0.0458
		[0,5;0,8]	0.7383	0.0559	0.2990	0.0510
	Γ	[0;0,1]	0.7317	0.0591	0.3601	0.0596
		[0,2;0,4]	0.7380	0.0583	0.3364	0.0578
		[0,5;0,8]	0.7419	0.0570	0.3765	0.0598
7,2	Γ^3	[0;0,1]	0.7213	0.0203	0.2559	0.0168
		[0,2;0,4]	0.7199	0.0201	0.2678	0.0171
		[0,5;0,8]	0.7197	0.0201	0.2463	0.0159
	Γ	[0;0,1]	0.7112	0.0211	0.3013	0.0200
		[0,2;0,4]	0.7201	0.0213	0.3327	0.0209
		[0,5;0,8]	0.7115	0.0207	0.3196	0.0190

A simulação comprova a robustez do AACCP, que utiliza as variáveis latentes para agrupar as observações frente à utilização das variáveis originais utilizadas pelo AAV. Tal resultado decorre da redução do efeito de correlação das variáveis sobre a clusterização obtida pelo uso dos componentes como variáveis de clusterização.

2.6. Conclusões

Este trabalho apresentou um método para a formação de grupos homogêneos de trabalhadores de acordo com seus perfis de aprendizado, o qual integra curvas de aprendizado e análise de agrupamentos. As curvas são utilizadas para a geração de parâmetros que descrevem características dos perfis de aprendizado dos trabalhadores; tais parâmetros são avaliados através da Análise de Componentes Principais (ACP) com vistas à melhoria da precisão da clusterização. A qualidade dos agrupamentos formados foi avaliada através do Índice Silhouette (SI).

Ao ser aplicado em um processo da indústria calçadista, o método na qual os parâmetros são transformados em componentes principais apresentou um índice Silhouette 0,968, valor superior a 0,392 obtido com a clusterização tradicional. Além disso, o método proposto permitiu gerar agrupamentos nos quais há poucas observações disponíveis (21 trabalhadores observados, e um número superior de variáveis a serem consideradas na clusterização, 29 parâmetros de aprendizado), cenário este que pode comprometer a eficiência de algumas ferramentas multivariadas. Para avaliar a eficácia do método, foi realizado um experimento de simulação, o qual comprovou o desempenho superior do AACCP frente ao AAV, mesmo com a inserção de fatores de ruído, correlação e diferentes razões entre variáveis e observações nos agrupamentos formados.

Neste trabalho os agrupamentos foram formados incluindo todas as variáveis (parâmetros de aprendizado), tanto na AAV, quanto na AACCP que utiliza todas as avariáveis

para gerar 1 componente principal. Porém, trabalhos futuros visam à construção de uma sistemática para seleção das variáveis que serão utilizadas na clusterização. Pretende-se avaliar a contribuição efetiva das variáveis de clusterização nos agrupamentos formados.

2.7. Referências Bibliográficas

- Anzanello, M. J., Fogliatto, F.S., 2007a. Curvas de aprendizado: estado da arte e perspectivas de pesquisa. *Revista Gestão e Produção*, São Carlos, v. 14, n. 1,109-123.
- Anzanello, M. J., Fogliatto, F.S., 2007b. Learning curve modelling of work assignment in mass customized assembly lines. *International Journal of Production Research*, Vol. 45, No. 13, 2919 –2938.
- Anzanello, M. J., Fogliatto, F.S., 2011a. Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics* 41, 573-583.
- Anzanello, M. J., Fogliatto, F.S., 2011b. Programação de Tarefas Baseada em Curvas de Aprendizado para Linhas de Produção Customizadas. *Revista Produção Online*. Florianópolis, SC, v.11, n. 3, 851-870.
- Brida, J.,G.,Disegna, M.,Osti, L., 2012. Segmenting visitors of cultural events by motivation: A sequential non-linear clustering analysis of Italian Christmas Market visitors. *Expert Systems with Applications* 39 ,11349–11356.
- Dardan, S., Busch, D., Sward, D., 2006. An application of the learning curve and the nonconstant-growth dividend model: IT investment valuation at Intel Corporation. *Decision Support Systems* 41 (4), 688 - 697.
- Da Silveira, G., Borestein, D., Fogliatto, F.S., 2001. Mass customization: literature review and research directions. *International Journal of Production Economics* 72 , 1–13.
- Engblom, J., Solakivi, T.,Töyli, J.,Ojala,L., 2012. Multiple-method analysis of logistics costs. *International Journal of Production Economics* 137, 29–35.
- Fávero, L. P.; Belfiore, P.; da Silva, F.L.; Chan, B. L., 2009. *Análise de Dados*. Rio de Janeiro. Editora Campus.
- Fioretti, G., 2007. The organizational learning curve. *European Journal of Operational Research* 177, 1375 - 1384.
- Fogliatto, F.S.,Da Silveira, G., Borestein, D., 2011. The mass customization decade: An updated review of the literature. *International Journal of Production Economics* 138, 14–25.
- Gong, S., Luo, L.,Gong, W.,Gao, Y., Xie, M., 2012. Multivariate analyses of element concentrations revealed the groupings of propolis from different regions in China. *Food Chemistry* 134 , 583 –588.

- Guimarães, L. B., Anzanello, M. J., Renner, J. S., 2012. A learning curve-based method to implement multifunctional work teams in the Brazilian footwear sector. *Applied Ergonomics*, 43, 541-547.
- Hair, J; Anderson, R. E; Tatham, R; Black, W.C., 1998. *Multivariate Data Analysis*. New Jersey: Prentice Hall, 1998.
- Hamade, R. F., Jaber, M. Y., Sikström, S., 2009. Analyzing CAD competence with univariate and multivariate learning curve models. *Computers & Industrial Engineering* 56, 1510 – 1518.
- Jaber, M. Y., Khan, M., 2010. Managing yield by lot splitting in a serial production line with learning, rework and scrap. *International Journal of Production Economics*. 124 (1), 32 e 39.
- Jaber, M. Y., Bonney, M., Guiffrida, A.L, 2010. Coordinating a three-level supply chain with learning-based continuous improvement. *International Journal of Production Economics*. 127, 27–38.
- Jaber, M. Y., Glock, C.H., 2013. A learning curve for tasks with cognitive and motor elements. *Computers & Industrial Engineering* 64, 866–871.
- Jolliffe, I.T., 2002. *Principal Component Analysis*. Springer Series in Statistics, New York, 2nd-ed.
- Jung, U., Koh, B.H, 2009. Structural damage localization using wavelet-based silhouette statistics. *Journal of Sound and Vibration* 321, 590–604.
- Lu, W.Z., He, H.D., Dong, L.Y, 2011. Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Building and Environment* 46, 577 – 583.
- MacCarthy, B., Brabazon, Philip G., Bramham, J., 2003. Fundamental modes of operation for mass customization. *International Journal of Production Economics* 85, 289-304.
- Mao, G., Hu, B., Song, H., 2009. Exploring talent flow in Wuhan automotive industry cluster at China. *International Journal Production Economics* 122, 395–402.
- Mazur, J.E., Hastie, R., 1978. Learning as accumulation: a reexamination of the learning curve. *Psychology Bulletin* 85 (6), 1256–1274.
- Nadeau, M.C., Jar, A., Roth, R., Kirchain, R., 2010. A dynamic process-based cost modeling approach to understand learning effects in manufacturing. *International Journal of Production Economics* 128, 223–234.
- Naldi, M.C., Campello, R.J.G.B., Hruschka, E.R. , Carvalho, A.C.P.L.F., 2011. Efficiency issues of evolutionary k-means. *Applied Soft Computing* 11, 1938–1952.
- Nembhard, D.A., Uzumeri. M.V., 2000a. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* 25, 315-326.
- Nembhard, D.A., Uzumeri, M.V., 2000b. An individual-based description of learning within an organization. *IEEE Transactions on Engineering Management* 47 (3), 370-378.

- Plaza, M. Rohl, K., 2008. Learning and performance in ERP implementation projects: A learning-curve model for analyzing and managing consulting costs. *International Journal of Production Economics* 115, 72-85.
- Ried, S., Mirka, G., 2007. Learning curve analysis of a patient lift-assist device. *Applied Ergonomics* 38, 765 –771.
- Rencher, A. C.,1995. *Methods of Multivariate Analysis*. John Wiley & Sons, Inc., New York.
- Rousseeuw, P.J.,1987.Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* 20, 53 – 65.
- Saha, I., Maulik, U., Bandyopadhyay, S., Plewczynski, D. Improvement of new automatic differential fuzzy clustering using SVM classifier for microarray analysis. *Expert Systems with Applications* 38, 15122–15133.
- Stroieke, R. E.; Fogliatto, F.S, Anzanello, M.J., 2011. Formação de Agrupamentos Homogêneos de Trabalhadores Através de Curvas de Aprendizado. Pré-anais XLIII Simpósio Brasileiro de Pesquisa Operacional.
- Towill, D.R., 1990. Forecasting learning curves. *International Journal of Forecasting* 6 (1), 25 - 38.
- Uzumeri, M.; Nembhard, D.,1998. A Population of Learners: A New Way to Measure Organizational Learning. *Journal of Operations Management, Netherlands*, v. 16, n. 5,515-528.
- Wong, P.S.P., Cheung, S.O., HardCastle, C., 2007. Embodying learning effect in performance prediction. *Journal of Construction Engineering and Management, ASCE* 133 (6), 474 – 482.
- Wong, P.S.P., Cheung, S.O., Wu, R.T.H., 2010. Learning from project monitoring feedback: A case of optimizing behavior of contractors. *International Journal of Project Management* 28, 469 – 481.
- Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3 (4), 122-128.
- Yeh, S., Rubin, E., 2012. A review of uncertainties in technology experience curves. *Energy Economics* 34 ,762–771.
- Zorgios, Y., Vlismas O., Venieris, G. 2009. A learning curve explanatory theory for team learning valuation, *VINE*, Vol. 39, 1 , 20 – 39.

Apêndice - Parâmetros CAs

Trabalhador	Wright	Stanford-b	DeJong	Curva S	Plateau	Knecht	Exponencial de 3 parâmetros	Tempo constante	Hiperbólico de 2 Parâmetros	Hiperbólico de 3 Parâmetros
1	Y_1 119.66 $b=-$ -0.05	Y_1 175.10 $B=$ 335.67 $b=-$ -0.09	Y_1 127.93 M 0.75 $b=-$ -0.56	Y_1 102.62 M 1.00 $B=$ 0.00 $b=$ 0.00	$C=$ 95.51 Y_1 31.42 $b=-$ -0.56	Y_1 130.14 $b=-$ -0.13 $c=$ 0.01	$k=$ 9.26 $\rho=$ 100.98 $r=$ 143.63	y_c 4.67 y_l 4.58 $t=$ 143.62	$k=$ 7.40 $r=$ 7.43	$k=$ 12.33 $\rho=$ 88.53 $r=$ 145.18
2	Y_1 62.06 $b=-$ -0.06	Y_1 1013.87 $B=$ 219.45 $b=-$ -0.54	Y_1 62.06 M 0.00 $b=-$ -0.06	Y_1 1370.71 M 1.00 $B=$ 241.54 $b=-$ -0.58	$C=$ 0.00 Y_1 61.98 $b=-$ -0.06 $c=$ 0.00	Y_1 51.20 $b=-$ -0.01 $c=$ 0.00	$k=$ 17.00 $\rho=$ 58.84 $r=$ 88.10	y_c 8.28 y_l 8.72 $t=$ 88.10	$k=$ 15.06 $r=$ 9.19	$k=$ 18.91 $\rho=$ 29.27 $r=$ 43.68
3	Y_1 59.40 $b=-$ -0.11	Y_1 59.43 $B=$ 0.01 $b=-$ -0.11	Y_1 191.92 M 0.42 $b=-$ -0.28	Y_1 40.67 M 1.00 $B=$ 1.00 $b=$ 0.00	$C=$ 26.31 Y_1 35.95 $b=-$ -0.28 $c=$ 0.00	Y_1 63.28 $b=-$ -0.16 $c=$ 0.00	$k=$ 16.74 $\rho=$ 17.78 $r=$ 22.71	y_c 9.09 y_l 7.65 $t=$ 22.71	$k=$ 17.43 $r=$ 5.08	$k=$ 18.71 $\rho=$ 10.53 $r=$ 7.17
4	Y_1 191.92 $b=-$ -0.17	Y_1 265.47 $B=$ 5.24 $b=-$ -0.24	Y_1 191.92 M 0.00 $b=-$ -0.17	Y_1 108.74 M 1.00 $B=$ 1.00 $b=$ 0.00	$C=$ 0.00 Y_1 191.26 $b=-$ -0.16 $c=$ 0.00	Y_1 162.00 $b=-$ -0.12 $c=$ 0.00	$k=$ 10.06 $\rho=$ 77.80 $r=$ 195.36	y_c 3.30 y_l 6.76 $t=$ 195.46	$k=$ 7.29 $r=$ 18.68	$k=$ 12.14 $\rho=$ 54.52 $r=$ 156.25
5	Y_1 202.08 $b=-$ -0.12	Y_1 324.31 $B=$ 15.03 $b=-$ -0.22	Y_1 202.08 M 0.00 $b=-$ -0.12	Y_1 324.28 M 1.00 $B=$ 15.03 $b=-$ -0.22	$C=$ 0.00 Y_1 201.58 $b=-$ -0.12 $c=$ 0.00	Y_1 185.00 $b=-$ -0.09 $c=$ 0.00	$k=$ 7.15 $\rho=$ 168.55 $r=$ 280.79	y_c 3.23 y_l 3.93 $t=$ 280.79	$k=$ 5.25 $r=$ 12.06	$k=$ 8.06 $\rho=$ 105.20 $r=$ 167.02
6	Y_1 155.83 $b=-$ -0.11	Y_1 155.83 $B=$ 0.00 $b=-$ -0.11	Y_1 170.87 M 0.55 $b=-$ -0.50	Y_1 155.83 M 0.00 $B=$ 0.00 $b=-$ -0.11	$C=$ 94.30 Y_1 76.57 $b=-$ -0.50 $c=$ 0.00	Y_1 183.00 $b=-$ -0.19 $c=$ 0.00	$k=$ 11.43 $\rho=$ 117.75 $r=$ 281.09	y_c 3.91 y_l 7.52 $t=$ 281.16	$k=$ 6.48 $r=$ 8.82	$k=$ 11.38 $\rho=$ 72.52 $r=$ 146.82
7	Y_1 205.84 $b=-$ -0.08	Y_1 278.82 $B=$ 15.09 $b=-$ -0.14	Y_1 205.84 M 0.00 $b=-$ -0.08	Y_1 279.97 M 0.27 $B=$ 8.85 $b=-$ -0.16	$C=$ 0.00 Y_1 205.75 $b=-$ -0.08 $c=$ 0.00	Y_1 196.00 $b=-$ -0.06 $c=$ 0.00	$k=$ 5.08 $\rho=$ 115.37 $r=$ 142.29	y_c 2.82 y_l 2.26 $t=$ 142.29	$k=$ 4.19 $r=$ 6.31	$k=$ 6.45 $\rho=$ 94.85 $r=$ 122.73
8	Y_1 101.43 $b=-$ -0.18	Y_1 101.43 $B=$ 0.00 $b=-$ -0.18	Y_1 108.81 M 0.27 $b=-$ -0.35	Y_1 108.81 M 0.27 $B=$ 1.00 $b=-$ -0.35	$C=$ 29.69 Y_1 79.12 $b=-$ -0.35 $c=$ 0.00	Y_1 109.57 $b=-$ -0.24 $c=$ 0.00	$k=$ 12.60 $\rho=$ 0.00 $r=$ 12.30	y_c -2.19 y_l 14.84 $t=$ 12.32	$k=$ 14.87 $r=$ 11.54	$k=$ 13.63 $\rho=$ -6.00 $r=$ 5.10
9	Y_1 35.97 $b=-$ -0.13	Y_1 35.97 $B=$ 1.00 $b=-$ -0.13	Y_1 43.13 M 0.43 $b=-$ -0.53	Y_1 43.13 M 0.43 $B=$ 1.00 $b=-$ -0.53	$C=$ 18.37 Y_1 24.76 $b=-$ -0.53 $c=$ 0.00	Y_1 44.00 $b=-$ -0.21 $c=$ 0.00	$k=$ 37.27 $\rho=$ 40.60 $r=$ 5.53	y_c 19.93 y_l 17.34 $t=$ 53.05	$k=$ 33.32 $r=$ 5.22	$k=$ 43.09 $\rho=$ 24.55 $r=$ 30.83
10	Y_1 184.77 $b=-$ -0.28	Y_1 220.81 $B=$ 1.13 $b=-$ -0.33	Y_1 184.77 M 0.00 $b=-$ -0.28	Y_1 220.81 M 0.00 $B=$ 1.13 $b=-$ -0.33	$C=$ 0.00 Y_1 183.13 $b=-$ -0.28 $c=$ 0.00	Y_1 163.00 $b=-$ -0.25 $c=$ 0.00	$k=$ 15.34 $\rho=$ 13.85 $r=$ 82.91	y_c 2.36 y_l 12.98 $t=$ 82.91	$k=$ 15.03 $r=$ 38.08	$k=$ 19.70 $\rho=$ 8.35 $r=$ 79.91
11	Y_1 246.67 $b=-$ -0.27	Y_1 250.20 $B=$ 0.08 $b=-$ -0.28	Y_1 257.57 M 0.12 $b=-$ -0.36	Y_1 250.19 M 0.00 $B=$ 0.08 $b=-$ -0.28	$C=$ 30.45 Y_1 227.11 $b=$ 0.36 $c=$ 0.00	Y_1 221.00 $b=-$ -0.24 $c=$ 0.00	$k=$ 8.46 $\rho=$ 17.23 $r=$ 61.18	y_c 2.08 y_l 6.38 $t=$ 61.18	$k=$ 9.28 $r=$ 28.98	$k=$ 10.19 $\rho=$ 4.50 $r=$ 45.09
12	Y_1 158.75 $b=-$ -0.24	Y_1 312.88 $B=$ 6.84 $b=-$ -0.41	Y_1 158.75 M 0.00 $b=-$ -0.24	Y_1 126.76 M 0.00 $B=$ 18.02 $b=-$ -0.12	$C=$ 0.00 Y_1 156.86 $b=-$ -0.23 $c=$ 0.00	Y_1 135.00 $b=-$ -0.19 $c=$ 0.00	$k=$ 11.45 $\rho=$ 7.61 $r=$ 41.62	y_c 1.91 y_l 9.54 $t=$ 41.62	$k=$ 13.48 $r=$ 23.93	$k=$ 15.12 $\rho=$ 5.10 $r=$ 42.58
13	Y_1 60.70 $b=-$ -0.09	Y_1 70.78 $B=$ 5.58 $b=-$ -0.12	Y_1 60.70 M 0.00 $b=-$ -0.09	Y_1 70.78 M 0.00 $B=$ 5.58 $b=-$ -0.12	$C=$ 0.00 Y_1 60.67 $b=-$ -0.09 $c=$ 0.00	Y_1 56.00 $b=-$ -0.06 $c=$ 0.00	$k=$ 16.37 $\rho=$ 49.31 $r=$ 50.96	y_c 10.15 y_l 6.22 $t=$ 50.96	$k=$ 15.75 $r=$ 4.97	$k=$ 18.66 $\rho=$ 29.55 $r=$ 26.28
14	Y_1 237.80 $b=-$ -0.15	Y_1 246.11 $B=$ 10.70 $b=-$ -0.13	Y_1 239.17 M 0.07 $b=-$ -0.17	Y_1 246.11 M 0.00 $B=$ 10.70 $b=-$ -0.13	$C=$ 16.91 Y_1 222.26 $b=-$ -0.17 $c=$ 0.00	Y_1 242.09 $b=-$ -0.17 $c=$ 0.00	$k=$ 6.72 $\rho=$ 58.17 $r=$ 135.31	y_c 2.35 y_l 4.37 $t=$ 135.29	$k=$ 5.85 $r=$ 22.46	$k=$ 7.04 $\rho=$ 20.43 $r=$ 59.63
15	Y_1 190.75 $b=-$ -0.25	Y_1 314.89 $B=$ 4.28 $b=-$ -0.38	Y_1 190.75 M 0.00 $b=-$ -0.25	Y_1 314.89 M 0.00 $B=$ 4.28 $b=-$ -0.38	$C=$ 0.00 Y_1 187.53 $b=$ 4.28 $c=$ 0.00	Y_1 85.49 $b=-$ -0.02 $c=$ 0.00	$k=$ 10.49 $\rho=$ 14.91 $r=$ 60.57	y_c 2.29 y_l 8.20 $t=$ 60.57	$k=$ 11.30 $r=$ 29.57	$k=$ 14.70 $\rho=$ 12.98 $r=$ 71.85
16	Y_1 90.49 $b=-$ -0.15	Y_1 97.90 $B=$ 1.20 $b=-$ -0.17	Y_1 90.49 M 0.00 $b=-$ -0.15	Y_1 97.90 M 0.00 $B=$ 1.20 $b=-$ -0.17	$C=$ 0.00 Y_1 90.39 $b=-$ -0.15 $c=$ 0.00	Y_1 90.49 $b=-$ -0.15 $c=$ 0.00	$k=$ 14.78 $\rho=$ 36.10 $r=$ 35.09	y_c 7.11 y_l 7.68 $t=$ 55.09	$k=$ 13.96 $r=$ 8.40	$k=$ 17.60 $\rho=$ 22.86 $r=$ 36.76
17	Y_1 39.06 $b=-$ -0.08	Y_1 358.82 $B=$ 132.60 $b=-$ -0.48	Y_1 39.06 M 0.00 $b=-$ -0.08	Y_1 358.82 M 0.00 $B=$ 132.60 $b=-$ -0.48	$C=$ 0.00 Y_1 39.05 $b=-$ -0.08 $c=$ 0.00	Y_1 37.94 $b=-$ -0.07 $c=$ 0.00	$k=$ 24.44 $\rho=$ 53.18 $r=$ 46.25	y_c 16.70 y_l 7.74 $t=$ 46.25	$k=$ 22.49 $r=$ 2.58	$k=$ 27.08 $\rho=$ 29.31 $r=$ 19.32
18	Y_1 162.04 $b=-$ -0.13	Y_1 162.04 $B=$ 0.00 $b=-$ -0.13	Y_1 173.97 M 0.33 $b=-$ -0.29	Y_1 173.09 M 0.33 $B=$ 1.00 $b=-$ -0.29	$C=$ 58.09 Y_1 115.88 $b=-$ -0.29 $c=$ 0.00	Y_1 175.00 $b=-$ -0.01 $c=$ 0.00	$k=$ 7.84 $\rho=$ 75.88 $r=$ 105.09	y_c 4.03 y_l 3.81 $t=$ 105.09	$k=$ 7.42 $r=$ 12.67	$k=$ 8.83 $\rho=$ 37.58 $r=$ 52.22
19	Y_1 375.31 $b=-$ -0.17	Y_1 385.59 $B=$ 3.29 $b=-$ -0.15	Y_1 375.31 M 0.00 $b=-$ -0.17	Y_1 385.59 M 0.00 $B=$ 3.29 $b=-$ -0.15	$C=$ 0.00 Y_1 375.31 $b=$ 3.29 $c=$ -0.17	Y_1 340.00 $b=-$ -0.14 $c=$ 0.00	$k=$ 4.00 $\rho=$ 50.00 $r=$ 122.00	y_c 1.30 y_l 1.60 $t=$ 121.90	$k=$ 3.57 $r=$ 24.35	$k=$ 4.98 $\rho=$ 45.39 $r=$ 120.08
20	Y_1 334.95 $b=-$ -0.26	Y_1 1050.23 $B=$ 11.68 $b=-$ -0.53	Y_1 334.95 M 0.00 $b=-$ -0.26	Y_1 1050.23 M 0.00 $B=$ 11.68 $b=-$ -0.53	$C=$ 0.00 Y_1 328.05 $b=-$ -0.25 $c=$ 0.00	Y_1 298.00 $b=-$ -0.22 $c=$ 0.00	$k=$ 7.35 $\rho=$ 38.54 $r=$ 157.89	y_c 1.59 y_l 5.58 $t=$ 157.89	$k=$ 6.99 $r=$ 56.61	$k=$ 11.15 $\rho=$ 37.12 $r=$ 220.62
21	Y_1 118.18 $b=-$ -0.08	Y_1 144.26 $B=$ 10.40 $b=-$ -0.12	Y_1 118.18 M 0.00 $b=-$ -0.08	Y_1 144.26 M 0.00 $B=$ 10.40 $b=-$ -0.12	$C=$ 0.00 Y_1 118.13 $b=-$ -0.08 $c=$ 0.00	Y_1 117.01 $b=-$ -0.07 $c=$ 0.00	$k=$ 9.14 $\rho=$ 161.14 $r=$ 180.05	y_c 5.40 y_l 3.73 $t=$ 180.03	$k=$ 7.62 $r=$ 6.35	$k=$ 9.76 $\rho=$ 77.90 $r=$ 68.67

3. SEGUNDO ARTIGO: FORMAÇÃO DE GRUPOS HOMOGÊNEOS DE TRABALHADORES BASEADA EM FUZZY C-MEANS E TRANSFORMAÇÃO DE DADOS

Bárbara Brzezinski Azevedo, Michel José Anzanello

Universidade Federal do Rio Grande do Sul (UFRGS), Av. Osvaldo Aranha, 99, Porto Alegre, CEP 90.035-190, Rio Grande do Sul, Brasil.

Resumo: A manufatura de produtos personalizados implica no aumento da variedade de modelos e redução no tamanho dos lotes de produção. Tarefas que dependem da habilidade humana são especialmente afetadas nesse contexto, visto que os trabalhadores precisam se adaptar às características de um novo modelo de produto. Esse processo de adaptação pode ocorrer de forma distinta dentro do grupo de trabalhadores. Este artigo propõe um método para a formação de grupos homogêneos de trabalhadores de acordo com seus perfis de aprendizado, o qual integra curvas de aprendizado e a ferramenta de clusterização *Fuzzy C-Means*. Parâmetros oriundos da modelagem das curvas, os quais caracterizam o processo de aprendizado de cada trabalhador, são remapeados através de métodos *Kernel* e Análise de Componentes Principais com vistas ao aprimoramento da qualidade da clusterização. A qualidade dos agrupamentos formados é avaliada através do Índice Silhouette. A utilização do *Kernel* Polinomial e da Análise de Componentes Principais apresentou ganhos no Índice Silhouette passando de um SI=0,32 utilizando a clusterização sem transformação de dados, para um SI=0,93 com o *Kernel* Polinomial e SI=0,97 com a Análise de Componentes Principais.

Palavras-chave: Clusterização, Curvas de Aprendizado, Métodos *Kernel*, *Fuzzy C-Means*, Análise de Componentes Principais

3.1. Introdução

A realidade competitiva impõe às empresas a adoção de estratégias de personalização de produtos e serviços a fim de atender às novas exigências dos clientes. Assim, a Customização em Massa (CM) tornou-se uma realidade que não pode ser negligenciada pelas corporações, demandando políticas específicas acerca de estratégia de produção com vistas à garantia de satisfação do cliente (DAABOUL *et al.*, 2011).

A CM é baseada no fornecimento de produtos personalizados através de processos flexíveis, a custos semelhantes aos de produtos padronizados (MACCARTHY *et al.*, 2003). Produtos customizados de determinados segmentos tipicamente requerem processos de manufatura manuais (FOGLIATTO; DA SILVEIRA, 2012) fabricados em lotes reduzidos (ANZANELLO; FOGLIATTO, 2011a). Associada a maiores opções de catálogo, os processos em CM alteram-se frequentemente, expondo o trabalhador à manufatura de novos modelos de produtos ou serviços constantemente. Tal exposição tende a reduzir os níveis de produção (quantitativamente e qualitativamente) nos ciclos iniciais, devido ao processo de

adaptação (aprendizado) a que cada trabalhador é submetido quando da solicitação de um novo produto (NEMBHARD; UZUMERI, 2000). Assim, a formação de equipes de trabalhadores com perfis semelhantes de aprendizado constitui-se em uma forma eficiente de equilibrar tais processos distintos de adaptação a novas tarefas. Em cenários de CM, agrupamentos permitem a formação de equipes de trabalhadores com velocidades semelhantes, monitorando gargalos nas linhas de produção, reduzindo níveis de ociosidade e proporcionando conhecimento acerca do perfil de aprendizado para rotação de trabalhadores dentro de cada equipe (ANZANELLO; FOGLIATTO, 2007).

A modelagem do perfil de aprendizado de um trabalhador pode ser operacionalizada através da Curva de Aprendizado (CA); CAs descrevem melhorias de desempenho dos trabalhadores em decorrência do acúmulo crescente de conhecimento em tarefas repetitivas (ARROW, 1962). CAs têm sido vastamente utilizadas para estudar as alterações de desempenho de trabalhadores relacionados aos efeitos do aprendizado (Plaza e Rolf, 2008), podendo auxiliar na gestão da produção para agrupamentos de trabalhadores de acordo com seus perfis de adequação a tarefas de complexidades distintas (UZUMERI; NEMBHARD, 1998; NEMBHARD; UZUMERI, 2000; STROIEKE *et al.*, 2011; WONG *et al.*, 2010, AZEVEDO; ANZANELLO, 2013).

Dentre tais abordagens, Uzumeri e Nembhard (1998) identificaram perfis distintos de trabalhadores através da análise dos parâmetros gerados pela modelagem das CAs, evidenciando as diferenças entre os trabalhadores para adequação às tarefas. Wong *et al.* (2010) avaliaram o padrão de desempenho de empreiteiros submetidos a *feedbacks* regulares e identificaram perfis distintos, os quais foram agrupados de acordo com os parâmetros da CA. Com propósitos semelhantes, Stroeike *et al.* (2011) caracterizaram diferentes grupos de trabalhadores, utilizando CAs e o método *K-Means* para clusterizar os dados de desempenho de trabalhadores. Já Azevedo e Anzanello (2013) propuseram a formação de grupos homogêneos de trabalhadores integrando parâmetros das CAs, Análise de Componentes Principais (ACP) e a técnica de clusterização *K-Means*; a qualidade dos agrupamentos realizados foi avaliada através do Índice Silhouette (SI). Naquele estudo, o mapeamento dos parâmetros oriundos das CAs através da ACP elevou o SI para 0,97, frente a 0,39 obtido quando a clusterização é operacionalizada diretamente nos parâmetros oriundos da CA. Apesar das diversas abordagens, os estudos apresentados acima não supriram todas as lacunas existentes na literatura referentes ao agrupamento de trabalhadores de acordo com perfis de aprendizado. A escassa literatura sobre o tema; apenas 4 artigos que utilizam CAs para analisar agrupamentos de trabalhadores foram encontrados. Destes, apenas em Stroeike *et al.*

(2011) e Azevedo e Anzanello (2013) os agrupamentos foram formados utilizando técnicas de clusterização. O que justifica a aplicação de outros métodos de clusterização em ambientes de manufatura customizada nos quais o trabalhador está exposto constantemente a novos processos produtivos.

Este artigo propõe uma sistemática para a formação de grupos homogêneos de trabalhadores de acordo com seus perfis de aprendizado em linhas de produção customizadas. O método inicia com a coleta de dados de desempenho de trabalhadores, os quais são ajustados a modelos distintos de CAs. Os parâmetros gerados, os quais descrevem o processo de aprendizado dos trabalhadores, são remapeados através de funções *Kernel* e da ACP visando aprimorar a qualidade do procedimento de clusterização. O método empregado para clusterização é o *Fuzzy C-Means*, o qual é testado em diversas transformações de dados. A qualidade de agrupamentos formados é avaliada através do SI, conforme proposto por Rousseeuw (1987).

O restante do artigo está organizado como segue. A seção 3.2 apresenta os fundamentos de CAs, *Fuzzy C-Means* e métodos *Kernel*. Na seção 3.3 é apresentado o método para formação de equipes caracterizadas pelos perfis de aprendizado dos trabalhadores. A seção 3.4 apresenta a aplicação da sistemática recomendada em um estudo de caso na indústria calçadista. Por fim, a seção 3.5 traz a conclusão do artigo e sugestões para pesquisas futuras.

3.2. Fundamentação Teórica

Esta seção está dividida em três partes. Inicialmente, revisam-se abordagens recentes sobre agrupamentos de trabalhadores através da utilização das CAs. Na sequência são apresentados os fundamentos analíticos dos métodos *Kernel*, seguido pela clusterização *Fuzzy C-Means*.

3.2.1. Agrupamentos envolvendo Curvas de Aprendizado

O agrupamento de trabalhadores por perfil de aprendizado ainda é pouco explorado na literatura, apesar de sua necessidade devido à crescente customização em massa de produtos e serviços. Segundo Cohen *et al.* (2008), a alocação de trabalhadores em atividades em conjunto é mais eficaz quando estes são agrupados de acordo com seus perfis de aprendizado. Uzumeri e Nembhard (1998) foram os pioneiros nesta linha de pesquisa ao aplicar CAs com o

objetivo de identificar agrupamentos naturais de trabalhadores; tais autores analisaram 3874 casos de aprendizado, modelando uma CA para cada trabalhador e identificando padrões de aprendizado distintos. Os autores utilizaram o modelo Hiperbólico de 3 Parâmetros (ver Mazur e Hastie, 1978) com vistas à descrição do comportamento do grupo de trabalhadores através de um gráfico tridimensional onde cada eixo representa um parâmetro do modelo; identificaram-se três *clusters* de trabalhadores. O primeiro era formado pelos trabalhadores que saíram antecipadamente do treinamento, indicando baixo desempenho; o segundo apresentava o comportamento oposto, sendo composto por trabalhadores que possuíam experiência prévia, desempenhando as atividades de forma mais eficiente; o terceiro *cluster* inclui os parâmetros da maioria dos trabalhadores, representando um perfil médio de aprendizado.

Similarmente a Uzumeri e Nembhard (1998), Wong *et al.* (2010) modelaram dados de desempenho de 53 projetos de construção civil através do modelo Hiperbólico de 3 Parâmetros, identificando quatro *clusters*: (i) *Competent learners*, (ii) *Incidental Learners*, (iii) *Indifferent Learners* e (iv) *Optimizers*. O grupo (i) apresentou melhoria contínua de desempenho durante os projetos, porém sua taxa de aprendizado era inferior à média dos componentes dos outros grupos de empreiteiros. Já o grupo (ii) possuía experiência em projetos de escalas maiores; seu desempenho melhorou inicialmente, mas piorou com a evolução do projeto. O grupo (iii) apresentou baixo desempenho durante o curso dos projetos, gerando os piores resultados entre os quatro grupos de empreiteiros. Por fim, o grupo (iv) apresentou pequenas flutuações de desempenho durante os projetos.

Diferentemente de Uzumeri e Nembhard (1998) e de Wong *et al.* (2010), Stroeke *et al.* (2011) e Azevedo e Anzanello (2013) utilizaram técnicas de clusterização para a formação de grupos de trabalhadores através da modelagem de CAs (e não apenas a sua identificação visual através de gráficos dos parâmetros oriundos da modelagem). Stroeke *et al.* (2011) caracterizaram dois grupos de trabalhadores por intermédio das CAs e dados de desempenho de trabalhadores da indústria calçadista foram utilizados como entrada na ferramenta de clusterização *K-Means*. Os autores reproduziram os dados de desempenho utilizando seis modelos de CAs: Wright (WRIGHT, 1936), Plateau (ANZANELLO; FOGLIATTO, 2011a), Stanford-b (ANZANELLO; FOGLIATTO, 2011a; YEH; RUBIN, 2012), Hiperbólico de 2 Parâmetros (MAZUR; HASTIE, 1978), Hiperbólico de 3 Parâmetros (MAZUR; HASTIE, 1978) e Exponencial de 3 Parâmetros (MAZUR; HASTIE, 1978). Já Azevedo e Anzanello (2013) agruparam trabalhadores utilizando os parâmetros de dez modelos de CAs, totalizando

29 parâmetros de aprendizado. Segundo os autores, a clusterização empregando todos os parâmetros permite a descrição de peculiaridades do aprendizado capturadas de forma distinta pelos diversos modelos de CAs. Alternativamente, os autores propuseram a transformação dos parâmetros através da ACP; a utilização dos componentes principais obtidos na análise permitiu o aumento no índice de qualidade de agrupamentos de 0,39 (valendo-se dos parâmetros originais) para 0,97 (parâmetros transformados em componentes principais). No presente estudo, serão testadas outras transformações nos dados - métodos *Kernel* - e a clusterização será feita através do procedimento de clusterização *Fuzzy C-Means*, onde se comparam os resultados obtidos com os demais autores. Estes métodos são detalhados nas seções 3.2.2. e 3.2.3.

3.2.2. Fuzzy C-Means

Fuzzy C-Means (FCM) é um algoritmo de agrupamento não supervisionado frequentemente aplicado no reconhecimento de padrões (Kril e Yuan, 1995). É baseado na distância entre pontos de dados de diferentes entradas. Os grupos são formados de acordo com a distância entre os pontos e os centros de fragmentação formados por cada *cluster*, assim cada observação possui uma probabilidade de pertencer a cada *cluster* (VELMURUGAN; SANTHANAM, 2010).

O método é baseado na minimização da função objetivo da equação (1) (CHANG *et al.*, 2011, VELMURUGAN; SANTHANAM, 2010; ROSS, 2004):

$$J_m = \sum_{j=1}^k \cdot \sum_{i=1}^N u_{i,j}^m d_{ij} \quad (1)$$

onde c_j é o centro de dimensão d do *cluster* C_j , u_{ij} é o grau de adesão de x_{ij} no *cluster* C_j , N é o número de pontos de dados, m é o parâmetro *fuzzifier* (qualquer número real maior do que 1); k é o número de agrupamentos, e d_{ij} é a distância euclidiana entre o ponto de dados x_{ij} e o centro do *cluster* C_j . É importante observar que u_{ij} deve satisfazer a seguinte restrição:

$$\sum u_{ij} = 1, \text{ para } i=1 \text{ a } N. \quad (2)$$

O particionamento *fuzzy* é realizado através de um processo iterativo de otimização da função objetivo (1), com a atualização de u_{ij} e dos centros dos *clusters* c_j dada por (3) e (4).

$$u_{ij} = \frac{1}{\sum_{k=1}^c \frac{\|x_i - c_j\|^2}{\|x_i - c_k\|^2}^{m-1}} \quad (3)$$

$$C_j = \frac{\sum_{i=1}^N u_{ij}^m * x_i}{\sum_{i=1}^N u_{ij}^m} \quad (4)$$

O procedimento iterativo é finalizado quando $\max_{ij}\{|u_{ij}^{(q+1)} - u_{ij}^{(q)}|\} < \zeta$, onde ζ é um critério de parada entre 0 e 1 e q representa os passos de iteração. Este procedimento converge para um mínimo local ou um ponto de sela de J_m (VELMURUGAN; SANTHANAM, 2010). No FCM, os dados são alocados a cada grupo por meio de uma função de pertinência, que representa o comportamento *fuzzy* do algoritmo (AL-ZOUBI *et al.*, 2007; JAIN; DUBES, 1998; JAIN *et al.*, 1999). Para isso, o algoritmo constrói uma matriz U cujos fatores são números entre 0 e 1, que representam o grau de associação entre os dados e centros de *clusters* (JAIN *et al.*, 1999; YONG *et al.*, 2004). Assim, o algoritmo é composto pelos seguintes passos (VELMURUGAN; SANTHANAM, 2010; KRIL; YUAN, 1995):

Etapa 1: Inicializar a matriz $U = [u_{ij}]$, $U^{(0)}$

Etapa 2: No passo q , calcular os centros dos vetores $C^{(q)} = [c_j]$ em $U^{(q)}$

Etapa 3: Atualizar $U^{(q)}$ com $U^{(q+1)}$

Etapa 4: Se $\|U^{(q+1)} - U^{(q)}\| < \zeta$, PARAR, caso contrário, voltar ao passo 2.

A clusterização FCM se comporta de maneira similar ao *K-Means*, porém é baseada no comportamento difuso dos dados e proporciona uma técnica natural para a produção de agrupamentos nos quais a adesão dos pesos tem uma interpretação natural (VELMURUGAN; SANTHANAM, 2010).

3.2.3. Métodos *Kernel*

Bancos de dados reais muitas vezes não são separáveis linearmente no espaço euclidiano, fazendo com que uma distância euclidiana entre os pontos de dados no espaço de entrada não seja representativa para determinada classe de dados. Para contornar esta dificuldade, uma função *Kernel* pode ser utilizada para transformação dos dados. Métodos de agrupamento baseados em *Kernel* usam uma matriz *Kernel* (explicada abaixo) para particionar os dados em um novo espaço característico (FAUBER; SCHWENKER, 2013).

Para entender como os métodos *Kernel* trabalham, considere N vetores de observações; sejam \mathbf{x}_i e \mathbf{x}_j dois vetores desse conjunto. Define-se a matriz quadrada $H_{ij} = h(\mathbf{x}_i, \mathbf{x}_j)$, de ordem $(N \times N)$, cujas entradas representam produtos internos entre as observações definidos por uma função *Kernel*. Essa matriz é denominada Matriz *Kernel*. A função $h(\mathbf{x}_i, \mathbf{x}_j)$ gera uma matriz positiva definida; desta forma, se h é positivo definido, existe um mapa ϕ , onde $h(\mathbf{x}_i, \mathbf{x}_j) = [\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)]$ (SCHOLKOPF; SMOLA, 2002). Porém, para calcular o produto interno no espaço característico, pode-se usar uma função *Kernel* sem explicitamente utilizar a função de mapeamento $\phi(x)$ (PARK *et al.*, 2006). Para tanto, é necessário definir o tipo de

função $H(\mathbf{x}_i, \mathbf{x}_j)$ que admite a representação do produto interno no espaço característico. Para isso, considere o Teorema de Mercer (GIROLAMI, 2002; FILIPPONE *et al.*, 2008): seja H uma função simétrica, isto é, $\forall \mathbf{x}_i, \mathbf{x}_j \in X, X \subseteq \mathbb{R}$:

$$H(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \quad (5)$$

onde, $\varphi(x): X \rightarrow F$ é uma função de mapeamento não linear dentro do espaço característico F . A função H pode ser representada em termos da equação (5). Neste artigo as funções representadas pelas equações (6) a (9) na Tabela 3.1 são usadas para substituir o produto interno e testá-las no algoritmo de clusterização *Fuzzy C-Means*.

Tabela 3.1– Funções *Kernel*

<i>Kernel</i>	Função	
Linear	$h(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i, \mathbf{x}_j)$ (6)	
Polinomial	$h(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i, \mathbf{x}_j)^d$ (7)	d= grau do polinômio
Gaussiano	$h(x_i, x_j) = e^{\left[-\frac{ x_i - x_j ^2}{\sigma^2}\right]}$ (8)	$\sigma > 0$
Sigmóide	$h(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\alpha \mathbf{x}_i^T \mathbf{x}_j + \gamma)$ (9)	α e $\gamma \subseteq \mathbb{R}$. Este último não é positivo definido, porém apresentou resultados interessantes em Park <i>et al.</i> (2006) e Yen <i>et al.</i> (2009)

A combinação de Métodos *Kernel* com o algoritmo FCM tem sido recentemente proposta pela literatura. Ao aplicar funções *Kernel* ao FCM, o algoritmo de clusterização irá melhor identificar a distância entre os pontos de dados através de um mapeamento não linear com relações em espaços característicos apropriados (HUANG *et al.*, 2012).

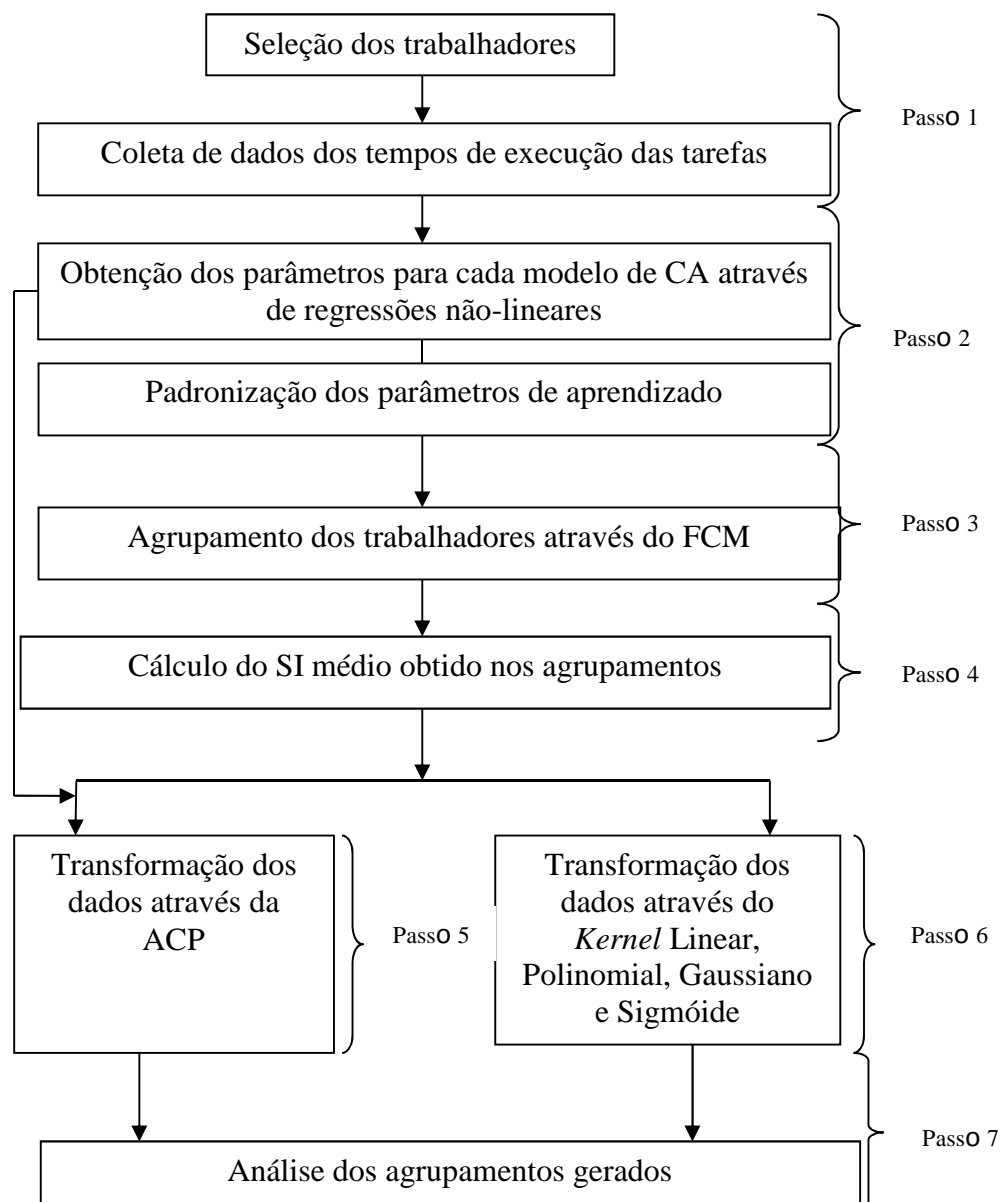
Em termos de suas aplicações, Huang *et al.* (2012) propuseram um algoritmo que estende o FCM a métodos *Kernel* através da incorporação de centros de *clusters* múltiplos que ajustam automaticamente os pesos do *Kernel* mais apropriado. Park *et al.* (2006) propuseram a combinação do FCM com o *Kernel* Gaussiano; o algoritmo proposto é formado pela incorporação de métodos *Kernel* ao FCM para trabalhar com limites de separação não-lineares entre os *clusters*. Os resultados demonstram uma acurácia de 88,12% nas classificações testadas pelos autores. Yen *et al.* (2009) testaram diferentes combinações de algoritmos de agrupamento com métodos *Kernel*. Os *Kernels* Gaussiano e Sigmóide foram combinados com o método de agrupamento hierárquico de Ward, *K-Means* e com o FCM; em três bancos de dados testados pelos autores, o *Kernel* Sigmóide apresentou melhor desempenho. Já Heo *et al.* (2009) propõem o método de clusterização *Robust kernel fuzzy PCA*, que estende o método *Kernel PCA* a associações da lógica *Fuzzy*. Segundo os autores,

os resultados experimentais sugerem que o algoritmo é um eficiente método não linear de redução dimensão e é menos sensível a ruídos que o original *Kernel PCA*.

3.3. Metodologia Proposta

O método proposto visa agrupar trabalhadores de acordo com seus perfis de aprendizado em ambientes de CM. O tempo de processamento das tarefas depende da velocidade de aprendizado dos trabalhadores e, por consequência, impacta na formação de equipes. A abordagem proposta apresenta duas etapas capitais na sua operacionalização: transformação dos dados através de métodos *Kernel* e ACP e formação dos agrupamentos através da técnica de clusterização FCM, conforme a Figura 3.1. Os passos são detalhados na sequência.

Figura 3.1- Fluxograma do Método



Passo 1: Seleção dos trabalhadores

Este passo inicia com a seleção de I ($i=1, \dots, I$) trabalhadores que terão seus processos de aprendizado monitorados. No momento em que o trabalhador é apresentado a uma nova tarefa, deve-se cronometrar o tempo demandado para execução de cada repetição. Os dados de desempenho devem ser coletados no início da operação, sendo armazenados até que não sejam observadas maiores alterações no padrão de desempenho. A coleta de dados é realizada através da contagem do número de unidades produzidas em determinado intervalo de operação.

Passo 2: Obtenção dos parâmetros para cada modelo de CA

Os dados de desempenho dos trabalhadores são modelados através dos dez modelos de CAs apresentadas na Tabela 3.2. Os parâmetros das CAs podem ser estimados através de rotinas não-lineares disponíveis em aplicativos de análise estatística. Assim, cada trabalhador i é caracterizado por um conjunto de parâmetros estimados a partir dos dados de desempenho obtidos em cada CA, totalizando 29 parâmetros de aprendizado que serão utilizados como variáveis de agrupamento. Tais parâmetros são apresentados na Tabela 3.2.

Tabela 3.2- Modelos de CAs (adaptado de Wong et al., 2007)

Modelo	Equação	Parâmetros	Aplicações	Referências
Modelos Potenciais				
Wright	$Y = Y_1 x^b$ (10)	Y = tempo demandado para a repetição x da tarefa; Y_1 = tempo de execução da primeira repetição; b = expoente de aprendizagem, onde $-1 < b < 0$, $b = \frac{\log \phi}{\log 2}$ e ϕ indica a taxa de aprendizado expressa percentualmente.	Expressar o efeito da aprendizagem no decrescimento do custo médio (ou tempo) ao longo das repetições de tarefas.	Wright (1936).
Stanford-b	$Y = Y_1(x + B)^b$ (11)	B = unidades equivalentes à experiência prévia do trabalhador. *	Incorporar a experiência prévia do trabalhador no início de um ciclo de produção.	Anzanello e Fogliatto (2011a); Yeh e Rubin (2012).
DeJong	$Y = Y_1[M + (1 - M)x^b]$ (12)	M = fator de incompressibilidade, onde $0 < M < 1$, representa a proporção do tempo total de operação constituído por procedimentos automatizados. Quando $M = 0$, o modelo reduz-se ao modelo de Wright. *	Incorporar a influência de maquinário no processo de aprendizado.	Anzanello e Fogliatto (2011a).
Curva S	$Y = Y_1[M + (1 - M)(x + B)^b]$ (13)	Os parâmetros assumem o mesmo significado dos modelos em que a equação foi baseada.	Descrever o processo de aprendizagem quando há a intervenção de maquinário.	Nembhard e Uzumeri (2000b); Anzanello e Fogliatto (2011a).
Plateau	$Y = C + Y_1 x^b$ (14)	C = constante aditiva que descreve o estado estacionário do aprendizado.*	Descrever o estado estacionário do processo de aprendizado.	Anzanello e Fogliatto (2011a).
Modelos Exponenciais				
Knecht	$Y = Y_1 x^b e^{c'x}$ (15)	C' = segunda constante. *	Modelar processos produtivos que apresentam elevado número de repetições.	Anzanello e Fogliatto (2011a).

Exponencia 1 de 3 Parâmetros	$Y = k(1 - e^{-(x+p)/r})$ (16)	Y= desempenho do trabalhador em itens produzidos; x = tempo de operação na tarefa; k = patamar máximo de desempenho a ser atingido (expresso em número de unidades produzidas por tempo de operação); p= experiência prévia do trabalhador na execução da tarefa; r= a taxa de aprendizado do trabalhador, onde: Y, x, k ≥ 0 e x, p e r são dados em unidades de tempo.	Adequa-se bem a cenários onde os trabalhadores possuem experiência prévia na tarefa.	Mazur e Hastie (1978); Anzanello e Fogliatto (2011a).
Tempo constante	$Y = y_c + y_f(1 - e^{-t/\tau})$ (17)	y _c = desempenho inicial; y _f = patamar máximo de aprendizado; t= tempo acumulado de operação, onde y _c e y _f são dados em unidades de tempo, sendo que t possui o mesmo significado do número de unidades produzidas adotado pelos demais modelos de CAs.	Expressar os efeitos do aprendizado, após a adaptação do trabalhador à tarefa.	Towill (1990); Dardan et al. (2006); Anzanello e Fogliatto (2011a).
Modelos Hiperbólicos				
Hiperbólico de 2 Parâmetros	$Y = k\left(\frac{x}{x+r}\right)$ (18)	Y= o número de unidades produzidas em um dado intervalo de tempo; x = intervalo de tempo considerado; r = parâmetro de aprendizado, expresso em unidades de tempo; k = nível máximo de aprendizado, expresso em número de unidades produzidas por tempo de operação.	Modelar a CA de acordo o número de unidades produzidas em conformidade com o número total de unidades.	Mazur e Hastie (1978).
Hiperbólico de 3 Parâmetros	$Y = k\left(\frac{x+p}{x+r+p}\right)$ (19)	p = parâmetro que representa experiência prévia do trabalhador na execução da tarefa, expresso em unidades de tempo*.	Possui a mesma aplicação que o modelo anterior, porém considera a experiência prévia do trabalhador.	Mazur e Hastie (1978);

*Os demais parâmetros apresentam as mesmas definições anteriores.

Passo 3: Agrupamento dos trabalhadores através do FCM

Os parâmetros gerados pela modelagem no passo 2 são normalizados para evitar efeitos de escala no processo de agrupamento. Na sequência, aplica-se a clusterização FCM sobre os dados padronizados, conforme descrito na seção 3.2.1. Neste artigo é mantido o número de *clusters* $K= 2$, pois o banco de dados utilizado nesta análise é o mesmo utilizado em Azevedo e Anzanello (2013), onde a formação de dois *clusters* apresentou o melhor desempenho em relação ao SI utilizando a sistemática de clusterização sugerida pelos autores, a qual realiza os agrupamentos utilizando os escores dos componentes principais oriundos da ACP. Desta forma, é possível realizar comparações diretas com os resultados do método aplicado pelos autores. Alternativamente, pode-se realizar previamente a clusterização hierárquica (ver Fávero *et al.*, 2009) dos dados e identificar, através do dendograma gerado, o número recomendado de *clusters* a ser formado.

Passo 4: Cálculo e análise do SI médio obtido nos agrupamentos

A qualidade dos resultados obtidos com a clusterização é avaliada através do índice Silhouette (SI). O SI que mede o grau de similaridade de uma observação em relação às demais em seu próprio grupo e em relação às observações alocadas ao grupo mais próximo a ela (ROUSSEEUW, 1987). O SI é calculado através da equação (20), onde $a(i)$ é média da distância da i -ésima observação a todas as demais pertencentes ao grupo no qual a observação foi alocada e $b(i)$ é a média da distância entre a i -ésima observação a todas as observações do grupo vizinho mais próximo. Os valores de SI estão no intervalo de -1 a 1, sendo que valores próximos a -1 indicam que a observação foi erroneamente inserida no grupo de destino. A qualidade dos agrupamentos formados é obtida pela média dos SIs resultantes de cada trabalhador.

$$SI = \frac{b(i)-a(i)}{\max\{b(i),a(i)\}} \quad (20)$$

Passo 5: Transformação dos dados através da ACP

Uma forma de transformação dos dados prévia à clusterização consiste em aplicar ACP sobre os parâmetros (variáveis originais de clusterização), transformando-os em variáveis latentes z (componentes principais). A ACP calcula os autovalores e autovetores da matriz de covariância dos dados arranjados de modo que a primeira linha seja o autovetor correspondente ao maior autovalor, e assim sucessivamente, até que a última linha corresponda ao menor autovalor (RENCHER, 1995; JOLLIFFE, 2002). Os componentes principais (CPs) resultantes, z , são combinações lineares das variáveis originais, x , $z_{ni} = w_{1n} x_{1i} + w_{2n} x_{2i} + \dots + w_{pn} x_{pi}$, e são mutuamente ortogonais (RENCHER, 1995). O número de CPs a ser retido pode ser definido com base no valor de SI obtido nas clusterizações. Assim, testam-se clusterizações retendo-se diferentes números de componentes principais e comparam-se os resultados obtidos. Neste artigo o banco de dados utilizado é o mesmo de Azevedo e Anzanello (2013), no qual este mesmo critério de retenção de CPs demonstrou que apenas o primeiro CP deve ser retido para a clusterização. Na sequência, executa-se a clusterização FCM e calcula-se o SI médio dos agrupamentos conforme descrito no Passo 4.

Passo 6: Transformação dos dados através do *Kernel* Linear, Polinomial, Gaussiano e Sigmóide

Alternativamente à transformação dos dados via ACP, neste passo testam-se transformações dos parâmetros de aprendizado através dos métodos *Kernel* apresentados na

Tabela 3.1. A transformação tem por objetivo inserir no contexto da análise eventuais relações não lineares. Assim, operacionalizam-se as matrizes *Kernel* conforme descrito na Seção 3.2.3; ao final deste passo, haverá quatro matrizes quadradas de ordem 21x21 (número de observações coletadas). A primeira matriz é composta pelo *Kernel* Linear conforme a equação (6); a segunda matriz é composta pelo *Kernel* Polinomial, utilizando $d=2$, conforme a equação (7); a terceira matriz é composta pelo *Kernel* Gaussiano, conforme a equação (8) e, por fim, a quarta matriz é composta pelo *Kernel* Sigmóide, utilizando $\alpha =1$ e $\gamma=-1$, conforme a equação (9). Ao aplicar as diferentes funções *Kernel*, o algoritmo FCM irá formar os agrupamentos através de um mapeamento não-linear, o que presumidamente facilitará a separação dos dados em espaços característicos. Na sequência, as quatro matrizes devem ser normalizadas para evitar efeitos de escala. Por fim, executa-se a clusterização FCM com o número de agrupamentos $K=2$ e calcula-se o SI médio dos agrupamentos conforme descrito no Passo 4.

Passo 7: Análise dos agrupamentos gerados

Por fim, comparam-se os valores de SI_j obtidos nos passos 4, 5 e 6. Haverá 6 resultados diferentes, oriundos dos dados originais e das 5 transformações propostas. O método responsável pelo maior SI deve ser considerado.

3.4. Resultados e Discussão

O método proposto foi aplicado na indústria calçadista. Este setor apresenta uma crescente CM em seus processos evidenciada pela redução no tamanho dos lotes de produção e alternância frequente de modelos (ANZANELLO; FOGLIATTO, 2011b).

Dados de desempenho de 21 trabalhadores foram coletados e modelados de acordo com as equações da Tabela 3.2. Foram gerados 29 parâmetros que representam os perfis de aprendizado (ver Apêndice). Tais parâmetros foram padronizados e utilizados como variáveis de agrupamento no método FCM, gerando-se $K=2$ clusters. Em seguida foram testadas cinco transformações nos dados: (i) ACP; (ii) *Kernel* Linear; (iii) *Kernel* Polinomial; (iv) *Kernel* Gaussiano e (v) *Kernel* Sigmóide.

Na transformação (i) foi retido 1 CP na análise, pois o banco de dados utilizado neste artigo é o mesmo utilizado em Azevedo e Anzanello (2013), no qual a utilização do primeiro CP apresentou o melhor desempenho em relação ao SI, explicando 75% da variância observada. Assim, os dados foram transformados em 1 vetor coluna contendo 21 linhas de

escores que representam os trabalhadores analisados. Em seguida, os dados foram agrupados em 2 *clusters* através do FCM. A extração das variáveis latentes z dos parâmetros de aprendizado reduz o efeito de correlação das variáveis sobre a clusterização. As transformações (ii), (iii), (iv) e (v) utilizando métodos *Kernel* geraram matrizes de ordem 21x21, as quais foram padronizadas para evitar efeitos de escala no processo de agrupamento. Em todos os casos foram formados dois agrupamentos ($K=2$).

Após a formação dos agrupamentos, foi calculado o SI médio de cada método. A Tabela 3.3 apresenta a média dos valores de SIs obtidos pelos 21 trabalhadores agrupados nos métodos analisados. *V* indica o *cluster* o trabalhador foi alocado e *SI* o respectivo Índice Silhouette. As duas últimas colunas são resultados de Azevedo e Anzanello (2013), onde as clusterizações foram realizadas através do *K-Means*.

Tabela 3.3- Resultados dos agrupamentos gerados pelos métodos propostos

Trabalhador	FCM aplicado aos dados originais		ACP FCM		Kernel Linear FCM		Kernel Gaussiano FCM		Kernel Sigmóide FCM		Kernel Polinomial FCM		<i>K-Means</i> aplicado aos dados originais		ACP <i>K-Means</i>	
	V	SI	V	SI	V	SI	V	SI	V	SI	V	SI	V	SI	V	SI
1	2	0,14	1	0,98	2	0,14	2	0,07	1	0,09	1	0,98	1	0,54	1	0,98
2	1	0,22	2	0,99	1	0,22	2	0,07	2	0,53	2	0,58	4	1,00	2	0,99
3	1	0,41	1	0,96	1	0,41	1	0,84	1	0,03	1	0,97	3	0,54	1	0,96
4	2	0,35	1	0,98	2	0,35	2	0,12	2	0,44	1	0,98	2	0,42	1	0,98
5	2	0,47	1	0,94	2	0,47	2	0,08	2	0,58	1	0,9	2	0,26	1	0,94
6	2	0,33	1	0,98	2	0,33	2	0,08	1	0,19	1	0,98	1	0,26	1	0,98
7	2	0,44	1	0,97	2	0,44	2	0,17	2	0,62	1	0,97	2	0,54	1	0,97
8	1	0,42	1	0,98	1	0,42	1	0,69	1	0,07	1	0,97	3	0,51	1	0,98
9	1	0,45	1	0,96	1	0,45	1	0,73	1	0,06	1	0,97	3	0,56	1	0,96
10	1	0,2	1	0,98	1	0,20	2	0,23	1	-0,44	1	0,98	2	-0,08	1	0,98
11	2	0,01	1	0,98	2	0,01	2	0,21	1	-0,11	1	0,98	2	0,25	1	0,98
12	1	0,28	1	0,98	1	0,28	2	0,11	2	0,52	1	0,98	3	0,10	1	0,98
13	1	0,39	1	0,97	1	0,39	1	0,86	2	0,6	1	0,97	3	0,54	1	0,97
14	2	0,36	1	0,98	2	0,36	2	0,22	2	0,43	1	0,97	2	0,51	1	0,98
15	1	0,15	1	0,96	1	0,15	2	0,15	2	0,55	1	0,98	2	-0,02	1	0,96
16	1	0,41	1	0,98	1	0,41	1	0,8	2	0,48	1	0,97	3	0,47	1	0,98
17	1	0,5	1	0,94	1	0,50	2	0,07	2	0,46	1	0,98	3	0,31	1	0,94
18	2	0,16	1	0,98	2	0,16	2	0,11	1	-0,2	1	0,98	2	0,12	1	0,98
19	2	0,46	1	0,86	2	0,46	2	0,07	2	0,26	1	0,63	2	0,60	1	0,86
20	2	0,24	2	0,98	2	0,24	2	0,07	2	0,44	2	0,82	2	0,51	2	0,98
21	2	0,23	1	0,98	2	0,23	2	0,09	2	0,6	1	0,98	2	0,29	1	0,98
SI médio		0,32		0,97		0,32		0,28		0,29		0,93		0,39		0,97

Pode-se verificar que, semelhantemente aos resultados obtidos em Azevedo e Anzanello (2013), a transformação nos dados utilizando a ACP previamente à clusterização aumenta significativamente a qualidade dos agrupamentos formados. Enquanto a clusterização FCM apresentou um SI médio de 0,32, a utilização da ACP prévia ao FCM gerou um SI médio de 0,97, mesmo resultado obtido em Azevedo e Anzanello (2013) através do método *K-Means*. Este resultado evidencia a robustez da ACP em relação aos demais métodos, pois permite que se considere apenas uma variável na clusterização (caso haja a retenção de 1 CP) .

Em relação às transformações *Kernel*, o método que apresentou o melhor desempenho foi os *Kernel* Polinomial, com SI médio de 0,93. Os *Kernels* Gaussiano e Sigmóide tiveram o menor desempenho entre as transformações: 0,28 e 0,29, respectivamente. Já o *Kernel* Linear obteve um SI de 0,32, resultado igual à clusterização realizada diretamente nos dados (sem transformação). Percebe-se que as duas últimas funções não remapeiam os dados de maneira a favorecer a clusterização. Verifica-se ainda, com base nos resultados deste estudo, que a transformação nos dados originais promove maior impacto nos resultados do que a técnica utilizada para clusterização (FCM versus *K-Means*). Neste experimento, a escolha pelo FCM ou pelo *K-Means* não impactou nos resultados, conduzindo a SIs médios semelhantes, conforme Tabela 3.3.

3.5. Considerações Finais

A adoção de estratégias de personalização de produtos e serviços tornou a CM uma realidade que não pode ser negligenciada pelas corporações. Produtos customizados, no entanto, tipicamente requerem processos de manufatura manuais que expõem o trabalhador à produção de novos modelos de produtos ou serviços com frequência. Tal exposição tende a reduzir os níveis de produção nos ciclos iniciais por conta do processo de aprendizado a que cada trabalhador é submetido. Assim, a formação de equipes de trabalhadores com perfis semelhantes de aprendizado representa em uma forma eficiente de equilibrar tais processos distintos de adaptação a novas tarefas.

Este artigo apresentou uma sistemática que integra ferramentas de análise multivariada - Análise de Componentes Principais (ACP) e Métodos *Kernel* - ao algoritmo *Fuzzy C-Means* (FCM), de forma a transformar os dados a serem agrupados e aumentar o desempenho do mesmo. Curvas de Aprendizado são incluídas na sistemática com o objetivo de formar grupos de trabalhadores com seus perfis homogêneos de aprendizado. Os parâmetros originados pela

modelagem das curvas informam as peculiaridades dos perfis de aprendizado dos trabalhadores; tais parâmetros são tratados como variáveis de agrupamento, as quais são transformadas com vistas ao aprimoramento da qualidade da clusterização. A qualidade dos agrupamentos formados foi avaliada através do Índice Silhouette (SI). Dados transformados pela ACP e *Kernel* Polinomial geraram agrupamentos mais apropriados, com SI=0.97 e SI=0.93, respectivamente. O alto desempenho da ACP neste estudo corrobora os resultados obtidos em Azevedo e Anzanello (2013), evidenciando a robustez do método.

Trabalhos futuros visam à construção de uma sistemática para seleção das variáveis (parâmetros de aprendizado) a serem utilizadas na clusterização, permitindo avaliar as variáveis que efetivamente impactam nos agrupamentos formados. Outras pesquisas visam testar a clusterização através de algoritmos de Redes Neurais e de Redes Complexas com o objetivo de comparar os resultados obtidos e escolher o método mais adequado ao conjunto de dados.

3.6. Referências

- Al-Zoubi, M.B., Hudaib, A., Al-Shboul, B., 2007a, A fast fuzzy clustering algorithm, Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, February 2007, Corfu Island, Greece, 28-32.
- Anzanello, M. J., Fogliatto, F.S., 2007. Curvas de aprendizado: estado da arte e perspectivas de pesquisa. *Revista Gestão e Produção*, São Carlos, v. 14, n. 1, 109-123.
- Anzanello, M. J., Fogliatto, F.S., 2011a. Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics* 41, 573-583.
- Anzanello, M. J., Fogliatto, F.S., 2011b. Programação de Tarefas Baseada em Curvas de Aprendizado para Linhas de Produção Customizadas. *Revista Produção Online*. Florianópolis, SC, v.11, n. 3, 851-870.
- Arrow, K.J., 1962, The economic implications of learning by doing. *Review of Economic Studies*, Vol. 29, 166-70.
- Azevedo, B.B., Anzanello, M.J., 2013. Agrupamento de Trabalhadores Através de Perfis de Aprendizagem. Artigo a submeter!
- Chang, C.T., Lai, J.Z.C, Jeng, M.D., 2011. A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement. *Journal Of Information Science And Engineering* 27, 995-1009.
- Cohen, Y.; Vitner, G.; Sarin, S., 2008. Work allocation to stations with varying learning slopes and without buffers. *European Journal of Production Research*, Netherlands v. 184, 797-801.

- Daaboul, J., Da Cunha, C., Bernard, A., Laroche, F., 2011. Design for mass customization: product variety vs. process variety. *CIRP Annals - Manufacturing Technology* 60, 169–174.
- Da Silveira, G., Borestein, D., Fogliatto, F.S., 2001. Mass customization: literature review and research directions. *International Journal of Production Economics* 72, 1–13.
- Dardan, S., Busch, D., Sward, D., 2006. An application of the learning curve and the nonconstant-growth dividend model: IT investment valuation at Intel Corporation. *Decision Support Systems* 41 (4), 688 - 697.
- Fauber, S., Schwenker, F., Semi-supervised clustering of large data sets with kernel methods, 2013. *Pattern Recognition Letters* , doi: <http://dx.doi.org/10.1016/j.patrec.2013.01.007>.
- Fávero, L. P.; Belfiore, P.; da Silva, F.L.; Chan, B. L., 2009. *Análise de Dados*. Rio de Janeiro. Editora Campus.
- Filippone, M., Camastra, F., Masulli, F., Rovetta, S., 2008. A survey of kernel and spectral methods for clustering. *Pattern Recognition* 41, 176 – 190.
- Fogliatto, F.S., Da Silveira, G., Borestein, D., 2011. The mass customization decade: An updated review of the literature. *International Journal of Production Economics* 138, 14–25.
- Girolami, M., 2002. Mercer Kernel-Based Clustering in Feature Space. *IEEE Trans. On Neural Networks* 13(3), 780-784.
- Heo, G., Gader, P., Frigui, H., 2009. RKF-PCA: Robust kernel fuzzy PCA. *Neural Networks* 22,642-650.
- Huang, H.C., Chuang, Y.Y, Chen, C.S., 2012. Multiple Kernel Fuzzy Clustering *IEEE Transactions On Fuzzy Systems*, Vol. 20, n°. 1, February 2012.
- Jain, A.K., Dubes, R.C., 1988. *Algorithms for Clustering Data*, Prentice Hall Inc., Englewood Cliffs, New Jersey. ISBN: 0-13-022278-X.
- Jain, A.K., M.N. Murty and P.J. Flynn, 1999, “Data Clustering: A review”, *ACM Computing Surveys*, Vol. 31, No. 3, September 1999. DOI:10.1.1.18.2720&rep=rep1&type=pdf.
- Jolliffe, I.T., 2002. *Principal Component Analysis*. Springer Series in Statistics, New York, 2nd-ed.
- Kril, G.J., Yuan, B., 1995. *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Prentice Hall, New Jersey.
- MacCarthy, B., Brabazon, Philip G., Bramham, J., 2003. Fundamental modes of operation for mass customization.
- Mazur, J.E., Hastie, R., 1978. Learning as accumulation: a reexamination of the learning curve. *Psychology Bulletin* 85 (6), 1256–1274.
- Nembhard, D.A., Uzumeri, M.V., 2000. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* 25, 315-326.

- Park, D-C., Tran, C. N., Park, S., 2006. J. Wang et al. (Eds.): ISNN 2006, LNCS 3971, 1038–1043. Springer-Verlag Berlin Heidelberg .
- Plaza, M. Rohlf, K., 2008. Learning and performance in ERP implementation projects: A learning-curve model for analyzing and managing consulting costs. *International Journal of Production Economics* 115, 72-85.
- Rencher, A. C., 1995. *Methods of Multivariate Analysis*. John Wiley & Sons, Inc., New York.
- Ross, T.J., 2004. *Fuzzy Logic With Engineering Applications*. Second Edition. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
- Rousseeuw, P.J., 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* 20, 53 – 65.
- Scholkopf, B., Smola, J., 2002. *Learning with Kernels*. England: The MIT Press.
- Stroieke, R. E.; Fogliatto, F.S, Anzanello, M.J. 2011. Formação de Agrupamentos Homogêneos de Trabalhadores Através de Curvas de Aprendizado. Pré-anais XLIII Simpósio Brasileiro de Pesquisa Operacional.
- Towill, D.R., 1990. Forecasting learning curves. *International Journal of Forecasting* 6 (1), 25 - 38.
- Uzumeri, M.; Nembhard, D., 1998. A Population of Learners: A New Way to Measure Organizational Learning. *Journal of Operations Management*, Netherlands, v. 16, n. 5, 515-528.
- Velmurugan, T., Santhanam, T., 2010. Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points. *European Journal of Scientific Research*, ISSN 1450-216X Vol.46 No.3, pp.320-330.
- Wong, P.S.P., Cheung, S.O., HardCastle, C., 2007. Embodying learning effect in performance prediction. *Journal of Construction Engineering and Management*, ASCE 133 (6), 474 – 482.
- Wong, P.S.P., Cheung, S.O., Wu, R.T.H., 2010. Learning from project monitoring feedback: A case of optimizing behavior of contractors. *International Journal of Project Management* 28, 469 – 481.
- Wright, T. P., 1936. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3 (4), 122-128.
- Yeh, S., Rubin, E., 2012. A review of uncertainties in technology experience curves. *Energy Economics* 34, 762–771.
- Yen, L., Fouss, F., Decaestecker, C., Francq, P., Saerens, M., 2009. Graph nodes clustering with the sigmoid commute-time kernel: A comparative study. *Data & Knowledge Engineering* 68, 338–361.
- Yong.Y, Chongxun, Z. , Pan, L., 2004. A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding, *Measurement Science Review*, Volume 4(1).

Apêndice - Parâmetros CAs

Trabalhador	Wright	Stanford-b	DeJong	Curva S	Plateau	Knecht	Exponencial de 3 parâmetros	Tempo constante	Hiperbólico de 2 Parâmetros	Hiperbólico de 3 Parâmetros
1	Y_1 119.66 $b=-$ -0.05	Y_1 175.10 $B=$ 335.67 $b=-$ -0.09	Y_1 127.93 M 0.75 $b=-$ -0.56	Y_1 102.62 M 1.00 $B=$ 0.00 $b=$ 0.00	$C=$ 95.51 Y_1 31.42 $b=-$ -0.56	Y_1 130.14 $b=-$ -0.13 $c=$ 0.01	$k=$ 9.26 $\rho=$ 100.98 $r=$ 143.63	y_c 4.67 y_l 4.58 $t=$ 143.62	$k=$ 7.40 $r=$ 7.43	$k=$ 12.33 $\rho=$ 88.53 $r=$ 145.18
2	Y_1 62.06 $b=-$ -0.06	Y_1 1013.87 $B=$ 219.45 $b=-$ -0.54	Y_1 62.06 M 0.00 $b=-$ -0.06	Y_1 1370.71 M 1.00 $B=$ 241.54 $b=-$ -0.58	$C=$ 0.00 Y_1 61.98 $b=-$ -0.06	Y_1 51.20 $b=-$ -0.01 $c=$ 0.00	$k=$ 17.00 $\rho=$ 58.84 $r=$ 88.10	y_c 8.28 y_l 8.72 $t=$ 88.10	$k=$ 15.06 $r=$ 9.19	$k=$ 18.91 $\rho=$ 29.27 $r=$ 43.68
3	Y_1 59.40 $b=-$ -0.11	Y_1 59.43 $B=$ 0.01 $b=-$ -0.11	Y_1 191.92 M 0.42 $b=-$ -0.28	Y_1 40.67 M 1.00 $B=$ 1.00 $b=$ 0.00	$C=$ 26.31 Y_1 35.95 $b=-$ -0.28	Y_1 63.28 $b=-$ -0.16 $c=$ 0.00	$k=$ 16.74 $\rho=$ 17.78 $r=$ 22.71	y_c 9.09 y_l 7.65 $t=$ 22.71	$k=$ 17.43 $r=$ 5.08	$k=$ 18.71 $\rho=$ 10.53 $r=$ 7.17
4	Y_1 191.92 $b=-$ -0.17	Y_1 265.47 $B=$ 5.24 $b=-$ -0.24	Y_1 191.92 M 0.00 $b=-$ -0.17	Y_1 108.74 M 1.00 $B=$ 1.00 $b=$ 0.00	$C=$ 0.00 Y_1 191.26 $b=-$ -0.16	Y_1 162.00 $b=-$ -0.12 $c=$ 0.00	$k=$ 10.06 $\rho=$ 77.80 $r=$ 195.36	y_c 3.30 y_l 6.76 $t=$ 195.46	$k=$ 7.29 $r=$ 18.68	$k=$ 12.14 $\rho=$ 54.52 $r=$ 156.25
5	Y_1 202.08 $b=-$ -0.12	Y_1 324.31 $B=$ 15.03 $b=-$ -0.22	Y_1 202.08 M 0.00 $b=-$ -0.12	Y_1 324.28 M 1.00 $B=$ 15.03 $b=-$ -0.22	$C=$ 0.00 Y_1 201.58 $b=-$ -0.12	Y_1 185.00 $b=-$ -0.09 $c=$ 0.00	$k=$ 7.15 $\rho=$ 168.55 $r=$ 280.79	y_c 3.23 y_l 3.93 $t=$ 280.79	$k=$ 5.25 $r=$ 12.06	$k=$ 8.06 $\rho=$ 105.20 $r=$ 167.02
6	Y_1 155.83 $b=-$ -0.11	Y_1 155.83 $B=$ 0.00 $b=-$ -0.11	Y_1 170.87 M 0.55 $b=-$ -0.50	Y_1 155.83 M 0.00 $B=$ 0.00 $b=-$ -0.11	$C=$ 94.30 Y_1 76.57 $b=-$ -0.50	Y_1 183.00 $b=-$ -0.19 $c=$ 0.00	$k=$ 11.43 $\rho=$ 117.75 $r=$ 281.09	y_c 3.91 y_l 7.52 $t=$ 281.16	$k=$ 6.48 $r=$ 8.82	$k=$ 11.38 $\rho=$ 72.52 $r=$ 146.82
7	Y_1 205.84 $b=-$ -0.08	Y_1 278.82 $B=$ 15.09 $b=-$ -0.14	Y_1 205.84 M 0.00 $b=-$ -0.08	Y_1 279.97 M 0.27 $B=$ 8.85 $b=-$ -0.16	$C=$ 0.00 Y_1 205.75 $b=-$ -0.08	Y_1 196.00 $b=-$ -0.06 $c=$ 0.00	$k=$ 5.08 $\rho=$ 115.37 $r=$ 142.29	y_c 2.82 y_l 2.26 $t=$ 142.29	$k=$ 4.19 $r=$ 6.31	$k=$ 6.45 $\rho=$ 94.85 $r=$ 122.73
8	Y_1 101.43 $b=-$ -0.18	Y_1 101.43 $B=$ 0.00 $b=-$ -0.18	Y_1 108.81 M 0.27 $b=-$ -0.35	Y_1 108.81 M 0.27 $B=$ 1.00 $b=-$ -0.35	$C=$ 29.69 Y_1 79.12 $b=-$ -0.35	Y_1 109.57 $b=-$ -0.24 $c=$ 0.00	$k=$ 12.60 $\rho=$ 0.00 $r=$ 12.30	y_c -2.19 y_l 14.84 $t=$ 12.32	$k=$ 14.87 $r=$ 11.54	$k=$ 13.63 $\rho=$ -6.00 $r=$ 5.10
9	Y_1 35.97 $b=-$ -0.13	Y_1 35.97 $B=$ 1.00 $b=-$ -0.13	Y_1 43.13 M 0.43 $b=-$ -0.53	Y_1 43.13 M 0.43 $B=$ 1.00 $b=-$ -0.53	$C=$ 18.37 Y_1 24.76 $b=-$ -0.53	Y_1 44.00 $b=-$ -0.21 $c=$ 0.00	$k=$ 37.27 $\rho=$ 40.60 $r=$ 5.53	y_c 19.93 y_l 17.34 $t=$ 53.05	$k=$ 33.32 $r=$ 5.22	$k=$ 43.09 $\rho=$ 24.55 $r=$ 30.83
10	Y_1 184.77 $b=-$ -0.28	Y_1 220.81 $B=$ 1.13 $b=-$ -0.33	Y_1 184.77 M 0.00 $b=-$ -0.28	Y_1 220.81 M 0.00 $B=$ 1.13 $b=-$ -0.33	$C=$ 0.00 Y_1 183.13 $b=-$ -0.28	Y_1 163.00 $b=-$ -0.25 $c=$ 0.00	$k=$ 15.34 $\rho=$ 13.85 $r=$ 82.91	y_c 2.36 y_l 12.98 $t=$ 82.91	$k=$ 15.03 $r=$ 38.08	$k=$ 19.70 $\rho=$ 8.35 $r=$ 79.91
11	Y_1 246.67 $b=-$ -0.27	Y_1 250.20 $B=$ 0.08 $b=-$ -0.28	Y_1 257.57 M 0.12 $b=-$ -0.36	Y_1 250.19 M 0.00 $B=$ 0.08 $b=-$ -0.28	$C=$ 30.45 Y_1 227.11 $b=$ 0.36	Y_1 221.00 $b=-$ -0.24 $c=$ 0.00	$k=$ 8.46 $\rho=$ 17.23 $r=$ 61.18	y_c 2.08 y_l 6.38 $t=$ 61.18	$k=$ 9.28 $r=$ 28.98	$k=$ 10.19 $\rho=$ 4.50 $r=$ 45.09
12	Y_1 158.75 $b=-$ -0.24	Y_1 312.88 $B=$ 6.84 $b=-$ -0.41	Y_1 158.75 M 0.00 $b=-$ -0.24	Y_1 126.76 M 0.00 $B=$ 18.02 $b=-$ -0.12	$C=$ 0.00 Y_1 156.86 $b=-$ -0.23	Y_1 135.00 $b=-$ -0.19 $c=$ 0.00	$k=$ 11.45 $\rho=$ 7.61 $r=$ 41.62	y_c 1.91 y_l 9.54 $t=$ 41.62	$k=$ 13.48 $r=$ 23.93	$k=$ 15.12 $\rho=$ 5.10 $r=$ 42.58
13	Y_1 60.70 $b=-$ -0.09	Y_1 70.78 $B=$ 5.58 $b=-$ -0.12	Y_1 60.70 M 0.00 $b=-$ -0.09	Y_1 70.78 M 0.00 $B=$ 5.58 $b=-$ -0.12	$C=$ 0.00 Y_1 60.67 $b=-$ -0.09	Y_1 56.00 $b=-$ -0.06 $c=$ 0.00	$k=$ 16.37 $\rho=$ 49.31 $r=$ 50.96	y_c 10.15 y_l 6.22 $t=$ 50.96	$k=$ 15.75 $r=$ 4.97	$k=$ 18.66 $\rho=$ 29.55 $r=$ 26.28
14	Y_1 237.80 $b=-$ -0.15	Y_1 246.11 $B=$ 10.70 $b=-$ -0.13	Y_1 239.17 M 0.07 $b=-$ -0.17	Y_1 246.11 M 0.00 $B=$ 10.70 $b=-$ -0.13	$C=$ 16.91 Y_1 222.26 $b=-$ -0.17	Y_1 242.09 $b=-$ -0.17 $c=$ 0.00	$k=$ 6.72 $\rho=$ 58.17 $r=$ 135.31	y_c 2.35 y_l 4.37 $t=$ 135.29	$k=$ 5.85 $r=$ 22.46	$k=$ 7.04 $\rho=$ 20.43 $r=$ 59.63
15	Y_1 190.75 $b=-$ -0.25	Y_1 314.89 $B=$ 4.28 $b=-$ -0.38	Y_1 190.75 M 0.00 $b=-$ -0.25	Y_1 314.89 M 0.00 $B=$ 4.28 $b=-$ -0.38	$C=$ 0.00 Y_1 187.53 $b=$ 4.28	Y_1 85.49 $b=-$ -0.02 $c=$ 0.00	$k=$ 10.49 $\rho=$ 14.91 $r=$ 60.57	y_c 2.29 y_l 8.20 $t=$ 60.57	$k=$ 11.30 $r=$ 29.57	$k=$ 14.70 $\rho=$ 12.98 $r=$ 71.85
16	Y_1 90.49 $b=-$ -0.15	Y_1 97.90 $B=$ 1.20 $b=-$ -0.17	Y_1 90.49 M 0.00 $b=-$ -0.15	Y_1 97.90 M 0.00 $B=$ 1.20 $b=-$ -0.17	$C=$ 0.00 Y_1 90.39 $b=-$ -0.15	Y_1 90.49 $b=-$ -0.15 $c=$ 0.00	$k=$ 14.78 $\rho=$ 36.10 $r=$ 35.09	y_c 7.11 y_l 7.68 $t=$ 55.09	$k=$ 13.96 $r=$ 8.40	$k=$ 17.60 $\rho=$ 22.86 $r=$ 36.76
17	Y_1 39.06 $b=-$ -0.08	Y_1 358.82 $B=$ 132.60 $b=-$ -0.48	Y_1 39.06 M 0.00 $b=-$ -0.08	Y_1 358.82 M 0.00 $B=$ 132.60 $b=-$ -0.48	$C=$ 0.00 Y_1 39.05 $b=-$ -0.08	Y_1 37.94 $b=-$ -0.07 $c=$ 0.00	$k=$ 24.44 $\rho=$ 53.18 $r=$ 46.25	y_c 16.70 y_l 7.74 $t=$ 46.25	$k=$ 22.49 $r=$ 2.58	$k=$ 27.08 $\rho=$ 29.31 $r=$ 19.32
18	Y_1 162.04 $b=-$ -0.13	Y_1 162.04 $B=$ 0.00 $b=-$ -0.13	Y_1 173.97 M 0.33 $b=-$ -0.29	Y_1 173.09 M 0.33 $B=$ 1.00 $b=-$ -0.29	$C=$ 58.09 Y_1 115.88 $b=-$ -0.29	Y_1 175.00 $b=-$ -0.01 $c=$ 0.00	$k=$ 7.84 $\rho=$ 75.88 $r=$ 105.09	y_c 4.03 y_l 3.81 $t=$ 105.09	$k=$ 7.42 $r=$ 12.67	$k=$ 8.83 $\rho=$ 37.58 $r=$ 52.22
19	Y_1 375.31 $b=-$ -0.17	Y_1 385.59 $B=$ 3.29 $b=-$ -0.15	Y_1 375.31 M 0.00 $b=-$ -0.17	Y_1 385.59 M 0.00 $B=$ 3.29 $b=-$ -0.15	$C=$ 0.00 Y_1 375.31 $b=$ -0.17	Y_1 340.00 $b=-$ -0.14 $c=$ 0.00	$k=$ 4.00 $\rho=$ 50.00 $r=$ 122.00	y_c 1.30 y_l 1.60 $t=$ 121.90	$k=$ 3.57 $r=$ 24.35	$k=$ 4.98 $\rho=$ 45.39 $r=$ 120.08
20	Y_1 334.95 $b=-$ -0.26	Y_1 1050.23 $B=$ 11.68 $b=-$ -0.53	Y_1 334.95 M 0.00 $b=-$ -0.26	Y_1 1050.23 M 0.00 $B=$ 11.68 $b=-$ -0.53	$C=$ 0.00 Y_1 328.05 $b=-$ -0.25	Y_1 298.00 $b=-$ -0.22 $c=$ 0.00	$k=$ 7.35 $\rho=$ 38.54 $r=$ 157.89	y_c 1.59 y_l 5.58 $t=$ 157.89	$k=$ 6.99 $r=$ 56.61	$k=$ 11.15 $\rho=$ 37.12 $r=$ 220.62
21	Y_1 118.18 $b=-$ -0.08	Y_1 144.26 $B=$ 10.40 $b=-$ -0.12	Y_1 118.18 M 0.00 $b=-$ -0.08	Y_1 144.26 M 0.00 $B=$ 10.40 $b=-$ -0.12	$C=$ 0.00 Y_1 118.13 $b=-$ -0.08	Y_1 117.01 $b=-$ -0.07 $c=$ 0.00	$k=$ 9.14 $\rho=$ 161.14 $r=$ 180.05	y_c 5.40 y_l 3.73 $t=$ 180.03	$k=$ 7.62 $r=$ 6.35	$k=$ 9.76 $\rho=$ 77.90 $r=$ 68.67

4. TERCEIRO ARTIGO: SELEÇÃO DE VARIÁVEIS PARA FORMAÇÃO DE GRUPOS HOMOGÊNEOS DE TRABALHADORES BASEADA EM CURVAS DE APRENDIZADO

Bárbara Brzezinski Azevedo, Michel José Anzanello

Universidade Federal do Rio Grande do Sul (UFRGS), Av. Osvaldo Aranha, 99, Porto Alegre, CEP 90.035-190, Rio Grande do Sul, Brasil.

Resumo: A manufatura de produtos personalizados implica no aumento da variedade de modelos e redução no tamanho dos lotes de produção. Tarefas que dependem da habilidade humana são especialmente afetadas nesse contexto, visto que os trabalhadores precisam se adaptar às características de um novo modelo. Esse processo de adaptação pode ocorrer de forma distinta dentro do grupo de trabalhadores, o que pode ser monitorado através de diversas variáveis de aprendizado. Neste artigo foram inicialmente identificadas 29 variáveis oriundas da modelagem por curvas de aprendizado. Porém, a retenção de variáveis irrelevantes e/ou ruidosas pode prejudicar o procedimento de clusterização, alocando inapropriadamente trabalhadores a grupos. Assim, propõe-se um método para seleção das variáveis mais relevantes para formação de grupos homogêneos de trabalhadores, o qual integra curvas de aprendizado, análise de componentes principais (ACP) e análise envoltória de dados (AED). Os procedimentos de clusterização *K-Means Fuzzy C-Means* (FCM) são testados no método proposto. Parâmetros oriundos da modelagem das curvas, os quais caracterizam o processo de aprendizado de cada trabalhador, são analisados pela ACP e hierarquizados através de um índice de importância de variável proposto. As variáveis irrelevantes são iterativamente removidas do conjunto de dados. A qualidade dos agrupamentos formados é avaliada através do Índice Silhouette (SI). A clusterização utilizando apenas a variável mais relevante apresentou SIs superiores às demais clusterizações, 0,70 e 0,65 para o *K-Means* e FCM, contra um SI=0,31, utilizando todas as variáveis disponíveis. A AED corroborou a aplicabilidade do método demonstrando eficiência superior com a exclusão das variáveis irrelevantes na clusterização.

Palavras-chave: Clusterização, Curvas de Aprendizado, Seleção de Variáveis, Análise de Componentes Principais, Análise Envoltória de Dados

4.1. Introdução

O consumidor exige cada vez mais produtos personalizados, com elevada qualidade, funcionalidade, ofertados a baixos custos e prazos de entrega reduzidos. O acesso à informação e o aumento no número de empresas alterou consideravelmente o perfil de consumo, fazendo que as mesmas tenham de adotar novas estratégias de produção que satisfaçam os requisitos do consumidor. Nesse sentido, a Customização em Massa (CM) permitiu às empresas projetar, produzir e ofertar produtos diferenciados que atendem às necessidades específicas dos clientes em tempo hábil e com preços semelhantes aos da produção em massa. O objetivo da CM é garantir capacidade de resposta ao cliente, obter eficiência em custos e produzir em alto volume produtos customizados (TU *et al.*, 2001; DONG, 2012).

Diversos produtos customizados, no entanto, requerem processos de manufatura manuais (FOGLIATTO; DA SILVEIRA, 2012), sendo fabricados em lotes menores (ANZANELLO; FOGLIATTO, 2011a). Assim, na CM os processos alteram-se frequentemente, expondo o trabalhador à manufatura de novos modelos de produtos constantemente. Tal exposição tende a reduzir os níveis de produção nos ciclos iniciais por conta do processo de adaptação (aprendizado) a que o trabalhador é submetido quando da execução de procedimentos associados a um novo produto (NEMBHARD; UZUMERI, 2000a). Este processo de aprendizado pode ocorrer de maneira distinta em cada trabalhador a ser alocado a linhas de produção. De tal forma, a formação de equipes de trabalhadores seguindo perfis semelhantes de aprendizado pode ser uma maneira eficiente para reduzir gargalos nas linhas de produção, mitigar níveis de ociosidade e proporcionar maior flexibilidade para rotação de trabalhadores dentro de cada equipe (ANZANELLO; FOGLIATTO, 2007).

A modelagem do perfil de aprendizado do trabalhador ocorre através da sua Curva de Aprendizado (CA). CAs têm sido utilizadas como ferramentas de gestão, permitindo prever e monitorar o desempenho de trabalhadores, grupos e organizações (JABER; GLOCK, 2013). Há diversos modelos de CAs disponíveis na literatura (WRIGHT, 1936; ANZANELLO; FOGLIATTO, 2011a; NEMBHARD; UZUMERI, 2000; MAZUR; HASTIE, 1978; TOWILL, 1990); todos estes modelos sugerem que o desempenho aumenta com a repetição das tarefas. Cada modelo, descrito por uma equação, utiliza parâmetros distintos para avaliar o processo de aprendizado de cada trabalhador. Neste artigo foram identificados 29 parâmetros, oriundos de 10 modelos de CAs, capazes de informar como o processo de aprendizado ocorre em cada trabalhador.

Um elevado número de variáveis (parâmetros oriundos das CAs), no entanto, pode reduzir a qualidade dos agrupamentos de trabalhadores formados, caso o procedimento de clusterização inclua variáveis irrelevantes e ruidosas (ANZANELLO; FOGLIATTO, 2011b; SATO-ILIC, 2011). A estruturação de agrupamentos depende de um número reduzido de variáveis e algumas podem atrapalhar a identificação dos grupos (MAUGIS *et al.*, 2007), sendo necessário selecionar previamente as variáveis que possuem maior peso na formação dos grupos. A seleção de variáveis preliminarmente à operacionalização de procedimentos de clusterização tem sido objeto de pesquisa na literatura (GNANADESIKAN; KETTENRING, 1995; GESU; MACCARONE, 1996; LIU; ONG, 2008; MAUGIS *et al.*, 2009a; MAUGIS *et al.*, 2011; BREABAN; LUCHIAN, 2011; VINH; BAILEY, 2013; BOUVEYRON; BRUNET-

SAUMARD, 2013), porém, não foram encontrados estudos aplicados à seleção de variáveis de aprendizado com vistas à formação de grupos homogêneos de trabalhadores.

Este artigo propõe um método para seleção de variáveis de aprendizado mais relevantes em procedimentos de clusterização utilizando Análise de Componentes Principais (ACP). Para tanto, a ACP é aplicada sobre os parâmetros das CAs que descrevem o aprendizado de trabalhadores (tidas como variáveis de clusterização), e um índice de importância das variáveis é obtido através dos pesos gerados pela ACP. Este índice visa hierarquizar as variáveis para identificar as mais importantes para o procedimento de clusterização. Na sequência, uma sistemática iterativa é iniciada valendo-se da ordem sugerida pelo índice de importância: a cada rodada, retira-se a variável com menor peso (menos relevante), agrupam-se os trabalhadores com base nas variáveis remanescentes, e avalia-se a qualidade dos agrupamentos formados através do Índice Silhouette (SI), conforme proposto por Rousseeuw (1987). Duas técnicas de clusterização são testadas: *K-Means* e *Fuzzy C-Means*. Por fim, aplica-se a Análise Envoltória de Dados (AED) para analisar a eficiência de cada técnica de clusterização, avaliando-se o número de variáveis utilizadas no agrupamento e o SI médio obtido.

O restante do artigo está organizado como segue. A seção 4.2 apresenta uma breve revisão de métodos para seleção de variáveis em clusterização, bem como uma introdução aos métodos de clusterização *K-Means* e *Fuzzy C-Means*, ACP e AED. Na seção 4.3 é apresentado o método para formação de equipes caracterizadas pelo aprendizado dos trabalhadores. A seção 4.4 apresenta a aplicação da sistemática em um estudo de caso na indústria calçadista. Por fim, a seção 4.5 traz a conclusão do artigo e sugestões para pesquisas futuras.

4.2. Fundamentação Teórica

Esta seção inicialmente versa sobre métodos de seleção de variáveis para clusterização. Na sequência são apresentados os fundamentos analíticos da Análise de Componentes Principais e aplicações da técnica para seleção de variáveis propostas pela literatura. Por fim, é apresentado embasamento teórico da Análise Envoltória de Dados e exemplos de aplicações na literatura.

4.2.1. Métodos de Seleção de Variáveis para Clusterização

Bases de dados de grande dimensão podem comprometer o desempenho de diversas ferramentas multivariadas, dentre as quais técnicas de clusterização (BOUYEYRON;

BRUNET-SAUMARD, 2013). Variáveis de agrupamento irrelevantes são independentes ou estão ligadas a variáveis de agrupamento mais importantes (MAUGIS *et al.*, 2009a), fazendo com que espaços multidimensionais que possuem características específicas possam ser melhor exploradas através de técnicas de seleção de variáveis.

Em Gnanadesikan *et al.* (1995), a identificação de variáveis relevantes para algoritmos de clusterização apoia-se em pesos que quantificam a importância de cada variável; tais pesos foram gerados utilizando ferramentas de análise multivariada. Gesu e Maccarone (1996) selecionam as variáveis significativas combinando abordagens probabilísticas através da lógica *Fuzzy*, enquanto que Breaban e Luchian (2011) sugerem uma função objetivo para selecionar as variáveis significativas e, simultaneamente, minimizar a variância dentro dos grupos e maximizar a distância entre os grupos. Já Sotoca e Pla (2010) propõem uma abordagem para a seleção de variáveis usando Informação Mútua; Vinh e Bailey (2013), utilizando o mesmo método, verificam, no entanto, que, as premissas do método não são aplicáveis. Com propósitos semelhantes, Liu *et al.* (2003), Maugis *et al.* (2009b), Raftery e Dean (2006) e Lian (2010) focam-se no problema de seleção de variáveis para clusterização utilizando um modelo de estrutura Bayesiana.

Maugis *et al.* (2007) realizam a seleção de variáveis através de modelos de mistura Gaussiana e generalizam o modelo de Raftery e Dean (2006) para identificar o papel de cada variável no conjunto de dados. Galimberti *et al.* (2009) também aplicaram modelos de mistura Gaussiana com o mesmo objetivo; a abordagem proposta foi investigada através de uma simulação de Monte Carlo. Há diversos estudos valendo-se desta abordagem, como Pan e Shen (2007), Wang e Zhou (2008), Xie *et al.* (2008), Zhang *et al.* (2009) e mais recentemente Poon *et al.* (2013). Zhang *et al.* (2009), em particular, introduziram um termo de penalidade na função de log-verossimilhança a fim de produzir dispersão nas características dos dados.

Já Liu e Ong (2008) apoiam-se em algoritmos genéticos para selecionar variáveis e determinar o número ideal de *clusters* em um estudo de segmentação de mercado. Os resultados mostraram que a seleção de variáveis através de algoritmos genéticos aumentou a qualidade da clusterização realizada através do procedimento *K-Means*.

4.2.2. Análise de Componentes Principais

A Análise de Componentes Principais (ACP) é um método utilizado para reduzir a dimensionalidade de uma matriz de dados, preservando a maior parte da variabilidade existente no conjunto de dados original (RENCHE, 1995). É considerado um método popular na literatura para reduzir dimensões de conjuntos de dados (BOUYEYRON;

BRUNET-SAUMARD, 2013; CADIMA *et al.*, 2004), no entanto, é pouco utilizada para reduzir o número de variáveis originais, pois seu objetivo inicial é abrigar as variáveis originais em componentes principais. Tais componentes (CPs), Z_1, Z_2, \dots, Z_n , são combinações lineares das variáveis originais, X_1, X_2, \dots, X_p , $z_{ni} = w_{1n} x_{1i} + w_{2n} x_{2i} + \dots + w_{pn} x_{pi}$, obtidos através dos autovalores e autovetores da matriz de covariância dos dados originais (RENCHER, 1995, JOLLIFFE, 2002). z são escores que representam a composição dos CPs em relação aos pontos de dados, enquanto que os pesos (w_{pn}) representam a mesma composição em relação às variáveis, pesos estes que decorrem da maximização da variância entre as variáveis latentes z (JOLLIFFE, 2002). O número de CPs retidos na análise pode ser definido com base no percentual de variância explicado por cada componente. Assim, somam-se os autovalores – arranjados do maior ao menor – até atingir o valor esperado de variância (RENCHER, 1995).

Alguns exemplos de estudos que aplicaram a ACP como técnica de redução de dimensões são: Campíns-Falcó *et al.* (2000) que utilizam um método semelhante ao utilizado neste artigo para selecionar variáveis através dos escores e dos pesos extraídos da ACP; Abdul-Wahab *et al.* (2005) que realizaram uma rotação dos componentes principais para selecionar as variáveis adequadas para um modelo de regressão com o objetivo de minimizar o efeito de multicolinearidade na estimativa dos coeficientes, e Sahmer e Qannari (2008), que selecionaram um subconjunto de variáveis com o objetivo de reduzir a lista de atributos em um painel sensorial.

Na literatura recente são pouco usuais estudos valendo-se da ACP com o objetivo de selecionar variáveis relevantes para a clusterização. Os estudos identificados utilizam a técnica com outros propósitos. Segundo Pacheco *et al.* (2013) a ACP não fornece uma redução real de dimensionalidade em termos das variáveis originais, uma vez que todas as variáveis são necessárias para definir um único CP. No entanto, será visto na sequência que, através das informações extraídas da ACP, é possível reduzir o número de variáveis originais através de um procedimento simples quando comparado a técnicas robustas como algoritmos genéticos e modelos Gaussianos.

4.2.3. A Análise envoltória de dados

A Análise envoltória de dados (DEA) é uma técnica matemática baseada em programação linear. Foi introduzida por Charnes *et al.* (1978) para determinar a eficiência de um grupo de unidades de tomada de decisão (DMU), quando medida ao longo de um conjunto de variáveis de entrada (*input*) e a saída (*output*) (GUO, 2009; ROSS; DROGE, 2004). Para

um dado conjunto de variáveis de entrada e saída, a DEA produz uma única medida de desempenho global (escore de eficiência) para cada DMU.

Os modelos de DEA mais conhecidos são: CRS – *Constant Returns to Scale* proposto por Charnes *et al.* (1978) e o modelo VRS – *Variable Returns to Scale*, proposto por Banker *et al.* (1984). Neste artigo é utilizado o modelo CRS, que consiste em um método de programação linear no qual a eficiência da DMU avaliada é dada pela otimização da divisão entre a soma ponderada das saídas e a soma ponderada das entradas sujeito à condição de que a proporção de cada DMU não seja maior do que 1. Matematicamente, é descrito como segue:

$$\text{Max } Eff_o = \frac{\sum_{j=1}^s g_j y_{lo}}{\sum_{i=1}^r v_i x_{io}} \quad (1)$$

$$\text{Sujeito a } \frac{\sum_{j=1}^s g_j y_{lk'}}{\sum_{i=1}^r v_i x_{ik'}} \leq 1, \forall k' \quad (2)$$

$$v_i, u_l \geq 0, \forall i, l$$

onde, Eff_o é a eficiência da DMU $_o$ em análise; v_i e g_j são os pesos de *inputs* i , $i=1, \dots, r$, e *outputs* l , $l=1, \dots, s$ respectivamente; i_k x e l_k y são os *inputs* i e *outputs* l da DMU k' , $k'=1, \dots, n$; x_{io} e y_{lo} são os *inputs* i e *outputs* l da DMU $_o$.

O modelo permite que cada DMU escolha os pesos para cada variável (entrada ou saída), desde que esses pesos aplicados às outras DMUs não gerem uma razão superior a 1 (FARREL, 1957). O CRS pode ser transformado em um problema de programação linear. Para tal, obriga-se que o denominador da função objetivo deva ser igual a uma constante, normalmente 1. Nesse modelo as variáveis de decisão são os pesos v_i e u_j , conforme a equação (4).

$$\text{Max } Eff_o = \sum_{j=1}^s g_j y_{lo} \quad (3)$$

$$\text{Sujeito a } \sum_{i=1}^r v_i x_{io} = 1; \quad (4)$$

$$\sum_{l=1}^s g_j y_{lk'} - \sum_{i=1}^r v_i x_{ik'} \leq 0, \forall k' \quad (5)$$

$$v_i, g_j \geq 0, \forall i, l$$

O modelo CRS teve diferentes aplicações na literatura recente. Hamdan e Rogers (2008) valeram-se dele para avaliar a eficiência de prestadores terceirizados de serviços logísticos. Os autores determinaram o impacto de cada variável de entrada e saída sobre a eficiência de armazéns e propuseram um conjunto de alterações para auxiliar gerentes e engenheiros na concepção de operações mais eficientes. Shen *et al.* (2012) o utilizaram para avaliar a segurança rodoviária em países da União Européia. Variáveis de entrada (como exposição ao risco) e de saída (mortes nas estradas) foram inseridas no modelo para ranquear os países de acordo com o escore de eficiência de segurança rodoviária e auxiliar no

desenvolvimento de políticas baseadas nas diferenças entre cada país. Por fim, Chiou *et al.* (2012) propõem rotas de transportes baseados no escores de eficiência fornecidos pelo DEA. Um estudo de caso desenvolvido pelos autores demonstra a superioridade dos modelos propostos para identificar os trajetos ineficientes e reduzir folgas nas variáveis de entrada.

4.3. Método proposto

O método proposto visa selecionar as variáveis mais relevantes oriundas das CAs com vistas ao agrupamento de trabalhadores, o qual é demonstrado na Figura 4.1. Os passos são detalhados na sequência.

Passo 1: Seleção dos trabalhadores

O método inicia com a seleção de I ($i=1, \dots, I$) trabalhadores a serem agrupados e que terão seus perfis de aprendizado modelados. No instante em que o trabalhador é exposto à nova tarefa, deve-se cronometrar o tempo utilizado para execução de cada repetição. Os dados de desempenho devem ser coletados no início da operação, sendo armazenados até que não sejam observadas maiores alterações no padrão de desempenho. A coleta de dados é realizada através da contagem do número de peças produzidas em determinado intervalo de operação.

Passo 2: Obtenção dos parâmetros para cada modelo de CA através de regressões não-lineares

Os dados de desempenho dos trabalhadores são modelados através dos dez modelos de CAs apresentadas na Tabela 4.1. Os parâmetros das CAs podem ser estimados através de rotinas disponíveis em aplicativos de análise estatística. Cada trabalhador i é caracterizado por um conjunto de parâmetros estimados a partir dos dados de desempenho obtidos em cada CA, totalizando 29 parâmetros de aprendizado que serão utilizados como variáveis de agrupamento. Assim, ao final deste passo haverá uma matriz, na qual as linhas representam os trabalhadores e as colunas representam cada parâmetro estimado.

Tabela 4.1- Modelos de CAs (adaptado de Wong et al., 2007)

Modelo	Equação	Parâmetros	Referências
Wright	$Y = Y_1 x^b$ (6)	Y = tempo demandado para a repetição x da tarefa; Y_1 = tempo de execução da primeira repetição; b = expoente de aprendizagem, onde $-1 < b < 0$, $b = \frac{\log \emptyset}{\log 2}$ e \emptyset indica a taxa de aprendizado expressa percentualmente.	Wright (1936).
Stanford-b	$Y = Y_1(x + B)^b$ (7)	B = unidades equivalentes à experiência prévia do trabalhador. *	Anzanello e Fogliatto (2011); Yeh e Rubin

			(2012).
DeJong	$Y = Y_1[M + (1 - M)x^b]$ (8)	$M =$ fator de incompressibilidade, onde $0 < M < 1$, representa a proporção do tempo total de operação constituído por procedimentos automatizados. Quando $M = 0$, o modelo reduz-se ao modelo de Wright.*	Anzanello e Fogliatto (2011).
Curva S	$Y = Y_1[M + (1 - M)(x + B)^b]$ (9)	Os parâmetros assumem o mesmo significado dos modelos em que a equação foi baseada.	Nembhard e Uzumeri (2000b); Anzanello e Fogliatto (2011).
Plateau	$Y = C + Y_1x^b$ (10)	$C =$ constante aditiva que descreve o estado estacionário do aprendizado.*	Anzanello e Fogliatto (2011).
Knecht	$Y = Y_1x^b e^{c'x}$ (11)	$C' =$ segunda constante. *	Anzanello e Fogliatto (2011).
Exponencial de 3 Parâmetros	$Y = k(1 - e^{-(x+p)/r})$ (12)	$Y =$ desempenho do trabalhador em itens produzidos; $x =$ tempo de operação na tarefa; $k =$ patamar máximo de desempenho a ser atingido (expresso em número de unidades produzidas por tempo de operação); $p =$ experiência prévia do trabalhador na execução da tarefa; $r =$ a taxa de aprendizado do trabalhador, onde: $Y, x, k \geq 0$ e x, p e r são dados em unidades de tempo.	Mazur e Hastie (1978); Anzanello e Fogliatto (2011).
Tempo constante	$Y = y_c + y_f(1 - e^{-t/\tau})$ (13)	$y_c =$ desempenho inicial; $y_f =$ patamar máximo de aprendizado; $t =$ tempo acumulado de operação, onde y_c e y_f são dados em unidades de tempo, sendo que t possui o mesmo significado do número de unidades produzidas adotado pelos demais modelos de CAs.	Towill (1990); Dardan et al. (2006); Anzanello e Fogliatto (2011).
Hiperbólico de 2 Parâmetros	$Y = k(\frac{x}{x+r})$ (14)	$Y =$ o número de unidades produzidas em um dado intervalo de tempo; $x =$ intervalo de tempo considerado; $r =$ parâmetro de aprendizado, expresso em unidades de tempo; $k =$ nível máximo de aprendizado, expresso em número de unidades produzidas por tempo de operação.	Mazur e Hastie (1978).
Hiperbólico de 3 Parâmetros	$Y = k(\frac{x+p}{x+r+p})$ (15)	$p =$ parâmetro que representa experiência prévia do trabalhador na execução da tarefa, expresso em unidades de tempo*.	Mazur e Hastie (1978).

*Os demais parâmetros apresentam as mesmas definições anteriores.

Passo 3: Obtenção dos Componentes Principais

Neste passo, aplica-se a ACP sobre a matriz de parâmetros de aprendizado das curvas, originada no passo 2. Dentre as informações geradas pela ACP, são de interesse os coeficientes das variáveis nos componentes (pesos) e o percentual de variância explicada por cada componente. O número de componentes a serem retidos é obtido com base neste percentual. Neste estudo, definiu-se que serão utilizados para a composição do IIV_p os N componentes principais responsáveis por explicar 90% da variância dos dados.

Passo 4: Geração do Índice de Importância das variáveis

Neste passo, calcula-se o Índice de Importância da Variável, IIV_p , para as P variáveis de clusterização. Esse índice considera o peso w_{np} da variável em cada um dos n componentes principais e variância explicada por cada um dos n componentes, v_n , conforme a equação (16).

$$IIV_p = \sum |w_{np}| * v_n, \quad (16)$$

Uma vez calculado o IIV para todas as variáveis, as mesmas são ordenadas de forma decrescente em relação ao valor do índice. O objetivo é hierarquizar as variáveis de acordo com os pesos que as mesmas possuem dentro do conjunto de dados.

Passo 5: Exclusão das variáveis irrelevantes e agrupamento dos trabalhadores

Este passo inicia com a normalização dos dados para evitar distorções de escala das variáveis nos cálculos de distância entre as observações (trabalhadores). Após a normalização, é realizada a clusterização das observações utilizando todas as P variáveis, e calculando-se então o SI médio dos agrupamentos. Na sequência, a variável com menor importância pelo IIV é retirada do conjunto de variáveis, uma nova clusterização é executada utilizando as $P-1$ variáveis remanescentes, e o SI médio as observações é armazenado. Tal procedimento é mantido até que haja somente 1 variável remanescente. Este passo é replicado para dois procedimentos de clusterização: *K-Means* e FCM.

O método *K-Means* é um procedimento de clusterização não hierárquico que utiliza a distância euclidiana como critério para formação dos *clusters* (FÁVERO *et al.*, 2009). É operacionalizado com a partição inicial das observações em K *clusters*. Na sequência, calculam-se os centroides para cada *cluster* e a distância euclidiana dos centroides para cada observação. Por fim, é feito o agrupamento das observações aos *clusters* cujos centroides estejam mais próximos. O procedimento é finalizado quando não houver variação significativa na distância mínima de cada observação a cada um dos centroides dos K *clusters* (NALDI *et al.*, 2011).

Por sua vez, o *Fuzzy C-Means* (FCM) realiza os agrupamentos de acordo com a distância entre os pontos de dados e os centros de fragmentação formados pelos *clusters* e atribui uma probabilidade de cada observação pertencer a cada *cluster* (VELMURUGAN; SANTHANAM, 2010). Tal probabilidade é obtida através de uma matriz cujos elementos são números entre 0 e 1, que representam o grau de associação entre os dados e os centros dos *clusters* (JAIN *et al.*, 1999; YONG *et al.*, 2004). O método é baseado na minimização da função objetivo (17) (ROSS, 2004): $J_m = \sum_{j=1}^k \cdot \sum_{i=1}^N u_{ij}^m d_{ij}$ (17)

onde, c_j é o centro de dimensão d do *cluster* C_j e u_{ij} é o grau de adesão de x_{ij} no *cluster* C_j , sendo que o somatório de u_{ij} deve ser igual a 1, para $i=1$ a N ; N é o número de observações, m é o parâmetro *fuzzifier* (qualquer número real maior do que 1); k é o número de agrupamentos, e d_{ij} é a distância euclidiana entre o ponto de dados x_{ij} e o centro do *cluster* C_j . Iterativamente, a função objetivo (17) é otimizada com a atualização de u_{ij} e de c_j pelas equações (18) e (19):

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left[\frac{x_i - c_j}{x_i - c_k} \right]^{m-1}} \quad (18); \quad C_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (19)$$

Tal procedimento é finalizado quando $\max_{ij} \{ |u_{ij}^{(q+1)} - u_{ij}^{(q)}| \} < \zeta$, onde ζ é um critério de parada entre 0 e 1 e q representa os passos de iteração e converge para um mínimo local ou

um ponto de sela de J_m (VELMURUGAN; SANTHANAM, 2010). Para maior detalhamento ver Ross (2004) e Kril e Yuan (1995).

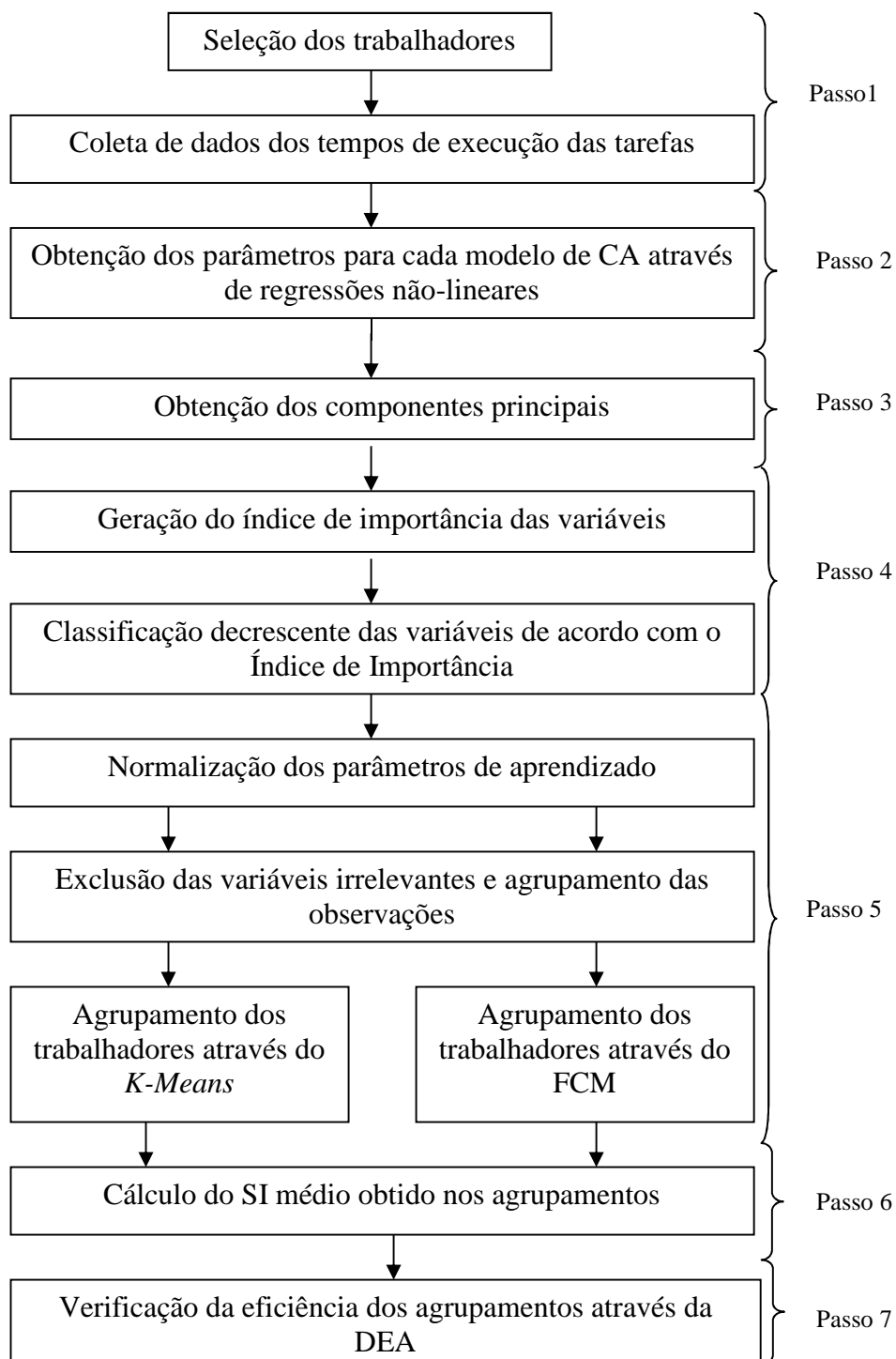
Passo 6: Cálculo e análise do SI médio obtido nos agrupamentos

A qualidade dos resultados obtidos com a clusterização é medida para posteriores comparações. Neste artigo utiliza-se o índice Silhouette (SI), que mede o grau de similaridade de uma observação em relação às demais em seu próprio *cluster*, comparada às observações alocadas ao *cluster* mais próximo a ela (ROUSSEEUW, 1987). O SI é calculado através da equação (20), onde $a(i)$ é média da distância da i -ésima observação a todas as demais pertencentes ao *cluster* no qual a observação foi alocada e $b(i)$ é a média da distância entre a i -ésima observação a todas as observações do *cluster* vizinho mais próximo. Os valores de SI estão no intervalo de -1 a 1, sendo que valores próximos a 1 indicam que a observação está corretamente alocada em seu *cluster*. A qualidade dos agrupamentos formados é obtida pela média dos SIs resultantes de cada trabalhador.

$$SI = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}} \quad (20)$$

Passo 7: Verificação da eficiência dos agrupamentos através da DEA

Após a obtenção dos SIs médios de cada clusterização pelos métodos *K-Means* e FCM, avalia-se a eficiência dos agrupamentos comparando a quantidade de variáveis utilizadas e o resultado do SI correspondente. Essa avaliação é feita através do modelo CRS da DEA. Neste artigo, as DMUs são tratadas como as clusterizações realizadas a cada retirada de variável, o *input* é o número de variáveis utilizada em cada clusterização e o *output* é o SI médio. Assim, cada clusterização apresentará um resultado entre 0 e 1 que representará seu escore de eficiência. A figura 4.1 sintetiza os passos descritos anteriormente.

Figura 4.1- Fluxograma do Método proposto

4.4. Resultados e Discussão

O método proposto foi aplicado no setor de costura de uma empresa calçadista. Utilizou-se o aplicativo Matlab 7.8 para geração da maioria dos resultados; a DEA foi operacionalizada no aplicativo SIAD (MEZA *et al.*, 2004).

Dados de desempenho de 21 trabalhadores foram coletados e modelados de acordo com as equações da Tabela 4.1. Foram gerados 29 parâmetros ($p=1,\dots,29$), apresentados no

Apêndice, que representam os perfis de aprendizado dos trabalhadores. Na sequência, aplicou-se a ACP sobre tais parâmetros. Para os dados analisados, os dois primeiros CPs explicaram 90% da variância presente nos dados, 75% o primeiro CP e 15% o segundo. A Tabela 4.2 apresenta o índice de importância, *IIV*, das variáveis de clusterização em ordem decrescente (ênfata-se o parâmetro de aprendizado que representam); o índice foi construído considerando 90% da variância explicada pelos dois primeiros componentes ($n = 2$), pois este era o critério para retenção dos CPs. Percebe-se que a variável p_1 (representando o tempo de execução da primeira repetição da tarefa no modelo de Wright) apresentou o maior *IIV*, 2,981.

Tabela 4.2- Hierarquia das variáveis de acordo com o *IIV*

Variável	Parâmetro	CA	<i>IIV_n</i>
p_1	Y_1	Wright	2,981
p_{16}	Y_1	Knecht	0,556
p_2	Φ	Wright	0,337
p_{15}	Φ	Plateau	0,187
p_4	B	Stanford-b	0,149
p_9	Y_1	Curva S	0,148
p_{21}	r	Exponencial de 3 parâmetros	0,138
p_{10}	M	Curva S	0,138
p_8	Φ	DeJong	0,136
p_5	Φ	Stanford-b	0,133
p_3	Y_1	Stanford-b	0,113
p_{13}	C	Plateau	0,106
p_{24}	t	Tempo constante	0,097
p_{26}	r	Hiperbólico de 2 Parâmetros	0,091
p_{28}	p	Hiperbólico de 3 Parâmetros	0,084
p_{17}	Φ	Knecht	0,069
p_{22}	y_c	Tempo constante	0,057
p_{25}	k	Hiperbólico de 2 Parâmetros	0,055
p_{19}	k	Exponencial de 3 parâmetros	0,047
p_{11}	B	Curva S	0,042
p_{23}	y_f	Tempo constante	0,039
p_{20}	p	Exponencial de 3 parâmetros	0,038
p_{12}	Φ	Curva S	0,031
p_{14}	Y_1	Plateau	0,029
p_{29}	r	Hiperbólico de 3 Parâmetros	0,028
p_{18}	c'	Knecht	0,024
p_{27}	k	Hiperbólico de 3 Parâmetros	0,024
p_6	Y_1	DeJong	0,023
p_7	M	DeJong	0,018

Tais parâmetros foram normalizados e utilizados como variáveis de clusterização. Após a normalização, foi realizada a clusterização das observações utilizando as 29 variáveis

em 2 grupos através do *K-Means* e do FCM. Na sequência, foi retirada a variável tida como menos importante pelo *IIV* (p_7) e realizada a clusterização dos trabalhadores novamente. Tal processo foi repetido até restar apenas a variável p_1 (tida como mais relevante pelo índice de importância de variável proposto).

A cada rodada, foram armazenados os SIs médios dos agrupamentos pelos procedimentos *K-Means* e FCM. A Tabela 4.3 apresenta o SI médio de cada clusterização, o desvio-padrão dos SIs, o número de variáveis utilizadas na clusterização, o procedimento empregado e a eficiência obtida através da DEA. Tanto nas clusterizações *K-Means* como nas FCM, a exclusão de variáveis menos importantes incrementou o SI médio dos agrupamentos, resultando em valores semelhantes nos dois procedimentos de clusterização. A clusterização utilizando todas as variáveis gerou SI médio de 0,31 no *K-Means* e no FCM. O SI médio aumenta consideravelmente ao utilizar-se apenas p_1 como variável de clusterização, resultando em 0,70 no *K-Means* e 0,65 no FCM. Por fim, a eficiência de cada clusterização foi medida através do modelo CRS da DEA. Tanto no procedimento *K-Means* quanto na FCM, a DEA evidenciou que a retenção do menor número de variáveis de clusterização aumenta a eficiência do processo de agrupamento, obtendo eficiência máxima ao reter-se apenas a p_1 .

Tabela 4.3- SIs médio das clusterizações

Nº Variáveis retidas	<i>K-Means</i>			FCM		
	SI médio	Desvio-Padrão dos SIs	Eficiência	SI médio	Desvio-Padrão dos SIs	Eficiência
1	0,70	0,24	1,000	0,65	0,36	1,000
2	0,65	0,28	0,464	0,65	0,28	0,389
3	0,53	0,36	0,252	0,52	0,35	0,324
4	0,45	0,29	0,161	0,44	0,30	0,208
5	0,53	0,18	0,151	0,42	0,19	0,106
6	0,41	0,22	0,098	0,17	0,18	0,083
7	0,36	0,17	0,074	0,30	0,18	0,071
8	0,34	0,16	0,061	0,25	0,16	0,056
9	0,31	0,19	0,049	0,22	0,16	0,049
10	0,28	0,19	0,040	0,22	0,18	0,050
11	0,56	0,20	0,073	0,26	0,13	0,033
12	0,53	0,19	0,063	0,23	0,16	0,037
13	0,22	0,34	0,024	0,24	0,26	0,056
14	0,26	0,14	0,027	0,20	0,23	0,046
15	0,22	0,16	0,021	0,24	0,33	0,061
16	0,23	0,30	0,021	0,23	0,30	0,052

17	0,44	0,21	0,037	0,23	0,24	0,039
18	0,42	0,20	0,033	0,28	0,12	0,019
19	0,29	0,19	0,022	0,28	0,14	0,021
20	0,28	0,13	0,020	0,28	0,13	0,018
21	0,30	0,17	0,020	0,30	0,14	0,019
22	0,32	0,13	0,021	0,32	0,13	0,016
23	0,31	0,13	0,019	0,31	0,13	0,016
24	0,31	0,13	0,019	0,31	0,13	0,015
25	0,33	0,13	0,019	0,33	0,13	0,014
26	0,31	0,12	0,017	0,31	0,12	0,013
27	0,32	0,13	0,017	0,32	0,13	0,013
28	0,32	0,16	0,016	0,32	0,14	0,014
29	0,31	0,13	0,015	0,31	0,13	0,013

Com base nos resultados acima, verifica-se que é preferível reter um subconjunto reduzido de variáveis para procedimentos de clusterização. Testando-se a clusterização *K-Means* e FCM, os valores de SI médio obtidos com as variáveis selecionadas pelo método proposto são consideravelmente maiores do que quando a clusterização é realizada sobre todas as variáveis.

4.5. Conclusões

Este trabalho apresentou um método para seleção de variáveis de clusterização na formação de grupos homogêneos de trabalhadores de acordo com seus perfis de aprendizado. Para tanto, curvas de aprendizado são utilizadas para a geração de parâmetros que descrevem características dos trabalhadores; tais parâmetros são avaliados através da Análise de Componentes Principais (ACP) com vistas à geração de um índice de importância de variável (IIV). Dois procedimentos de clusterização são testados, *K-Means* e *Fuzzy C-Means* (FCM), para os quais é iterativamente removida a variável tida como menos importante a cada clusterização. A qualidade dos agrupamentos formados foi avaliada através do Índice Silhouette (SI), e a eficiência das clusterizações através da Análise Envoltória de Dados (AED).

Ao ser aplicado em um processo da indústria calçadista, o método apresentou um SI de 0,70 no procedimento *K-Means* e 0,65 no FCM, valores elevados quando comparados a SI=0,31 obtido com a clusterização retendo todas as variáveis originais. Para corroborar a eficácia do método, foi utilizada a DEA, que demonstrou que a utilização de um subconjunto de variáveis para a clusterização incrementa a eficiência dos agrupamentos tanto no *K-Means* quanto no FCM.

Trabalhos futuros visam à adaptação da sistemática proposta através de algoritmos de Redes Neurais e de Redes Complexas. Objetiva-se comparar os resultados obtidos e selecionar o método mais adequado para clusterizações futuras.

4.6. Referências

- Abdul-Wahab, S.A., Bakheit, C.S., Al-Alawia, S.M., 2005. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software* 20, 1263–1271.
- Angulo-Meza, L.; Biondi Neto, L.; Soares de Mello, J. C. C. B.; Gomes, E. G.; Coelho, P. H. G., 2004. *FSDA – Free Software for Decision Analysis (SLAD – Software Livre de Apoio a Decisão)*, Rio de Janeiro.
- Anzanello, M. J., Fogliatto, F.S., 2007. Curvas de aprendizado: estado da arte e perspectivas de pesquisa. *Revista Gestão e Produção*, São Carlos, v. 14, n. 1, 109-123.
- Anzanello, M. J., Fogliatto, F.S., 2011a. Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics* 41, 573-583.
- Anzanello, M.J.;Fogliatto,F.S.; 2011b. Selecting the best clustering variables fou grouping mass-customized involving worker’s learning. *International Journal of Production Economics* 130 (2), 268-276.
- Banker, R.D.; Charnes, A.; Cooper, W.W., 1984. Some models for estimating technical scale inefficiencies in data envelopment analysis. *Management Science*, v. 30, n. 9, 1078-1092.
- Bouveyron, C.; Brunet-Saumard, C., 2013. Model-based clustering of high-dimensional data: A review. *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2012.12.008.
- Breaban, M.; Luchian, H.; 2011. A unifying criterion for unsupervised clustering and feature selection. *Pattern Recognition* 44, 854–865.
- Cadima, J., Cerdeira, J.O., Minhoto, M., 2004. Computational aspects of algorithms for variable selection in the context of principal components. *Computational Statistics & Data Analysis* 47, 225 – 236.
- Campíns-Falcó, P., Blasco-Gómez, F., Bosch-Reig, F., Gallo-Martínez, L., 2000. Principal component analysis for the selection of variables in the application of the H-point and generalised H-point standard addition method. *Talanta* 53, 317–330.
- Chang, C.T., Lai, J.Z.C, Jeng, M.D.,2011. A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement. *Journal Of Information Science And Engineering* 27, 995-1009.
- Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring efficiency of decision making units. *European Journal of Operational Research* 2, 429–444.
- Chiou, Y-C., Lan, L.W.; Yen, B.T.H., 2012. Route-based data envelopment analysis models. *Transportation Research Part E* 48, 415–425.

- Dardan, S., Busch, D., Sward, D., 2006. An application of the learning curve and the nonconstant-growth dividend model: IT investment valuation at Intel Corporation. *Decision Support Systems* 41(4), 688 - 697.
- Dong, B.; Jia, H.; Li, Z.; Dong, K.;2012. Implementing Mass Customization in Garment Industry. *Systems Engineering Procedia* 3,372 – 380.
- Farrel, M.J.,1957.The measurement of productive efficiency. *Journal of the Royal Statistic Society, series A, part 3*, p. 253-290.
- Fogliatto, F.S.; Da Silveira, G.; Borestein, D., 2012. The mass customization decade: An updated review of the literature. *International Journal of Production Economics* 138, 14–25.
- Galimberti, G.; Montanari, A.; Viroli, C., 2009. Penalized factor mixture analysis for variable selection in clustered data. *Computational Statistics and Data Analysis* 53 , 4301-4310
- Gesu, V.D.; Maccarone, M.C.;1996. Features selection and ‘Possibility Theory’. *Pattern Recognition, Vol 19, N°1*, 63-72.
- Gnanadesikan, R.; Kettenring, J.R.; Tsao, S. L., 1995. Weighting and Selection of Variables for Cluster Analysis. *Journal of Classification* 12, 113-136.
- Guo, P. 2009. Fuzzy data envelopment analysis and its application to location problems. *Information Sciences* 179, 820–829.
- Hamdan, A.; Rogers, K.J., 2008. Evaluating the efficiency of 3PL logistics operations. *Int. J. Production Economics* 113, 235–244.
- Jaber, M. Y.;Glock, C.H.;2013. A learning curve for tasks with cognitive and motor elements. *Computers & Industrial Engineering* 64, 866–871.
- Jain, A.K.; Murty, M.N.; Flynn, P.J., 1999, “Data Clustering: A review”, *ACM Computing Surveys*, Vol. 31, No. 3, September 1999. DOI:10.1.1.18.2720&rep=rep1&type=pdf.
- Jolliffe, I.T., 2002. *Principal Component Analysis*.Springer Series in Statistics, New York, 2nd-ed.
- Kril, G.J., Yuan, B., 1995. *Fuzzy Sets and Fuzzy Logic, Theory and Applications*.Prentice Hall, New Jersey.
- Lian, H., 2010. Sparse Bayesian hierarchical modeling of high-dimensional clustering problems. *Journal of Multivariate Analysis* 101, 1728-1737.
- Liu, H-H.; Ong, C-S., 2008. Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications* 34, 502–510.
- Liu, J., Zhang, J.L., Palumbo, M.J., Lawrence, C.E., 2003. Bayesian clustering with variable and transformation selection. *Bayesian Statistics* 7, 249–276.
- Mazur, J.E., Hastie, R., 1978. Learning as accumulation: a reexamination of the learning curve. *Psychology Bulletin* 85 (6),1256–1274.

- Maugis, C.; Celeux, G.; Martin-Magniette, M.L.; 2007. Variable Selection for Clustering with Gaussian Mixture Models. Institut National De Recherche En Informatique Et En Automatique. Inria-00153057, version 2.
- Maugis, C.; Celeux, G.; Martin-Magniette, M.-L., 2009a. Variable selection in model-based clustering: A general variable role modeling. *Computational Statistics and Data Analysis* 53, 3872-3882.
- Maugis, C., Celeux, G., Martin-Magniette, M.-L., 2009b. Variable selection for clustering with Gaussian mixture models. *Biometrics* 65 (3), 701–709.
- Maugis, C.; Celeux, G.; Martin-Magniette, M.-L. 2011. Variable selection in model-based discriminant analysis. *Journal of Multivariate Analysis* 102 ,1374–1387.
- Nembhard, D.A., Uzumeri. M.V., 2000a. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* 25, 315-326.
- Nembhard, D.A., Uzumeri, M.V., 2000b. An individual-based description of learning within an organization. *IEEE Transactions on Engineering Management* 47 (3), 370-378.
- Pacheco, J., Casado, S., Porras, S., 2013. Exact methods for variable selection in principal component analysis: Guide functions and pre-selection. *Computational Statistics and Data Analysis* 57, 95–111.
- Pan, W., Shen, X., 2007. Penalized model-based clustering with application to variable selection. *Journal of Machine Learning Research* 8, 1145–1164.
- Poon, L.K.M.; Zhang, N. L.; Liu, T.; Liu; A.H., 2013. Model-based clustering of high-dimensional data: Variable selection versus facet determination. *International Journal of Approximate Reasoning* 54 (2013) 196–215.
- Raftery, A.E., Dean, N., 2006. Variable selection for model-based clustering. *Journal of the American Statistical Association* 101 (473), 168–178.
- Rencher, A. C., 1995. *Methods of Multivariate Analysis*. John Wiley & Sons, Inc., New York.
- Ross, A.D.; Droge, C., 2004. An analysis of operations efficiency in large-scale distribution systems. *Journal of Operations Management* 21, 673–688.
- Ross, T.J., 2004. *Fuzzy Logic With Engineering Applications*. Second Edition. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
- Rousseeuw, P.J., 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* 20, 53 – 65.
- Sahmer, K. Qannari, E.M., 2008. Procedures for the selection of a subset of attributes in sensory profiling. *Food Quality and Preference* 19, 141–145.
- Sato-Ilic, M., 2011. Symbolic Clustering with Interval-Valued Data. *Procedia Computer Science* 6 , 358–363.
- Shen, Y.; Hermans, E.; Brijs, T.; Wets, G.; Vanhoof, K., 2012. Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident Analysis and Prevention* 48, 430– 441.

- Sotoca, J.M.; Pla, F., 2010. Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition* 43, 2068–2081.
- Towill, D.R., 1990. Forecasting learning curves. *International Journal of Forecasting* 6 (1), 25 - 38.
- Tu, Q.; Vonderembse, M.A.; Ragu-Nathan, T.S., 2001. The impact of time-based manufacturing practices on mass customization and value to customer. *Journal of Operations Management* 19, 201–217.
- Vinh, N. X.; Bailey, J.; 2013. Comments on supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition* 46, 1220–1225.
- Velmurugan, T., Santhanam, T., 2010. Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points. *European Journal of Scientific Research*, ISSN 1450-216X Vol.46 No.3, pp.320-330.
- Wang, S., Zhou, J., 2008. Variable selection for model-based high dimensional clustering and its application to microarray data. *Biometrics* 64, 440–448.
- Wong, P.S.P., Cheung, S.O., HardCastle, C., 2007. Embodying learning effect in performance prediction. *Journal of Construction Engineering and Management*, ASCE 133 (6), 474 – 482.
- Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3 (4), 122-128.
- Xie, B., Pan, W., Shen, X., 2008. Penalized model-based clustering with cluster-specific diagonal covariance matrices and grouped variables. *Electrical Journal of Statistics* 2, 168–212.
- Yeh, S., Rubin, E., 2012. A review of uncertainties in technology experience curves. *Energy Economics* 34, 762–771.
- Yong, Y., Chongxun, Z., Pan, L., 2004. A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding, *Measurement Science Review*, Volume 4(1).
- Zhang, Z., Dai, G., Jordan, M.I., 2009. A flexible and efficient algorithm for regularized fisher discriminant analysis, In: *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, 632–647.

Apêndice - Parâmetros CAs

Trabalhador	Wright	Stanford-b	DeJong	Curva S	Plateau	Knecht	Exponencial de 3 parâmetros	Tempo constante	Hiperbólico de 2 Parâmetros	Hiperbólico de 3 Parâmetros
1	Y ₁ 119.66 b= -0.05	Y ₁ 175.10 B= 335.67 b= -0.09	Y ₁ 127.93 M 0.75 b= -0.56	Y ₁ 102.62 M 1.00 B= 0.00 b= 0.00	C= 95.51 Y ₁ 31.42 b= -0.56	Y ₁ 130.14 b= -0.13 c= 0.01	k= 9.26 p= 100.98 r= 143.63	y _c 4.67 y _f 4.58 t= 143.62	k= 7.40 r= 7.43	k= 12.33 p= 88.53 r= 145.18
2	Y ₁ 62.06 b= -0.06	Y ₁ 1013.87 B= 219.45 b= -0.54	Y ₁ 62.06 M 0.00 b= -0.06	Y ₁ 1370.71 M 1.00 B= 241.54 b= -0.58	C= 0.00 Y ₁ 61.98 b= -0.06	Y ₁ 51.20 b= -0.01 c= 0.00	k= 17.00 p= 58.84 r= 88.10	y _c 8.28 y _f 8.72 t= 88.10	k= 15.06 r= 9.19	k= 18.91 p= 29.27 r= 43.68
3	Y ₁ 59.40 b= -0.11	Y ₁ 59.43 B= 0.01 b= -0.11	Y ₁ 191.92 M 0.42 b= -0.28	Y ₁ 40.67 M 1.00 B= 1.00 b= 0.00	C= 26.31 Y ₁ 35.95 b= -0.28	Y ₁ 63.28 b= -0.16 c= 0.00	k= 16.74 p= 17.78 r= 22.71	y _c 9.09 y _f 7.65 t= 22.71	k= 17.43 r= 5.08	k= 18.71 p= 10.53 r= 7.17
4	Y ₁ 191.92 b= -0.17	Y ₁ 265.47 B= 5.24 b= -0.24	Y ₁ 191.92 M 0.00 b= -0.17	Y ₁ 108.74 M 1.00 B= 1.00 b= 0.00	C= 0.00 Y ₁ 191.26 b= -0.16	Y ₁ 162.00 b= -0.12 c= 0.00	k= 10.06 p= 77.80 r= 195.36	y _c 3.30 y _f 6.76 t= 195.46	k= 7.29 r= 18.68	k= 12.14 p= 54.52 r= 156.25
5	Y ₁ 202.08 b= -0.12	Y ₁ 324.31 B= 15.03 b= -0.22	Y ₁ 202.08 M 0.00 b= -0.12	Y ₁ 324.28 M 1.00 B= 15.03 b= -0.22	C= 0.00 Y ₁ 201.58 b= -0.12	Y ₁ 185.00 b= -0.09 c= 0.00	k= 7.15 p= 168.55 r= 280.79	y _c 3.23 y _f 3.93 t= 280.79	k= 5.25 r= 12.06	k= 8.06 p= 105.20 r= 167.02
6	Y ₁ 155.83 b= -0.11	Y ₁ 155.83 B= 0.00 b= -0.11	Y ₁ 170.87 M 0.55 b= -0.50	Y ₁ 155.83 M 0.00 B= 0.00 b= -0.11	C= 94.30 Y ₁ 76.57 b= -0.50	Y ₁ 183.00 b= -0.19 c= 0.00	k= 11.43 p= 117.75 r= 281.09	y _c 3.91 y _f 7.52 t= 281.16	k= 6.48 r= 8.82	k= 11.38 p= 72.52 r= 146.82
7	Y ₁ 205.84 b= -0.08	Y ₁ 278.82 B= 15.09 b= -0.14	Y ₁ 205.84 M 0.00 b= -0.08	Y ₁ 279.97 M 0.27 B= 8.85 b= -0.16	C= 0.00 Y ₁ 205.75 b= -0.08	Y ₁ 196.00 b= -0.06 c= 0.00	k= 5.08 p= 115.37 r= 142.29	y _c 2.82 y _f 2.26 t= 142.29	k= 4.19 r= 6.31	k= 6.45 p= 94.85 r= 122.73
8	Y ₁ 101.43 b= -0.18	Y ₁ 101.43 B= 0.00 b= -0.18	Y ₁ 108.81 M 0.27 b= -0.35	Y ₁ 108.81 M 0.27 B= 1.00 b= -0.35	C= 29.69 Y ₁ 79.12 b= -0.35	Y ₁ 109.57 b= -0.24 c= 0.00	k= 12.60 p= 0.00 r= 12.30	y _c -2.19 y _f 14.84 t= 12.32	k= 14.87 r= 11.54	k= 13.63 p= -6.00 r= 5.10
9	Y ₁ 35.97 b= -0.13	Y ₁ 35.97 B= 1.00 b= -0.13	Y ₁ 43.13 M 0.43 b= -0.53	Y ₁ 43.13 M 0.43 B= 1.00 b= -0.53	C= 18.37 Y ₁ 24.76 b= -0.53	Y ₁ 44.00 b= -0.21 c= 0.00	k= 37.27 p= 40.60 r= 5.53	y _c 19.93 y _f 17.34 t= 53.05	k= 33.32 r= 5.22	k= 43.09 p= 24.55 r= 30.83
10	Y ₁ 184.77 b= -0.28	Y ₁ 220.81 B= 1.13 b= -0.33	Y ₁ 184.77 M 0.00 b= -0.28	Y ₁ 220.81 M 0.00 B= 1.13 b= -0.33	C= 0.00 Y ₁ 183.13 b= -0.28	Y ₁ 163.00 b= -0.25 c= 0.00	k= 15.34 p= 13.85 r= 82.91	y _c 2.36 y _f 12.98 t= 82.91	k= 15.03 r= 38.08	k= 19.70 p= 8.35 r= 79.91
11	Y ₁ 246.67 b= -0.27	Y ₁ 250.20 B= 0.08 b= -0.28	Y ₁ 250.19 M 0.12 b= -0.36	Y ₁ 250.19 M 0.00 B= 0.08 b= -0.28	C= 30.45 Y ₁ 227.11 b= 0.36	Y ₁ 221.00 b= -0.24 c= 0.00	k= 8.46 p= 17.23 r= 61.18	y _c 2.08 y _f 6.38 t= 61.18	k= 9.28 r= 28.98	k= 10.19 p= 4.50 r= 45.09
12	Y ₁ 158.75 b= -0.24	Y ₁ 312.88 B= 6.84 b= -0.41	Y ₁ 158.75 M 0.00 b= -0.24	Y ₁ 126.76 M 0.00 B= 18.02 b= -0.12	C= 0.00 Y ₁ 156.86 b= -0.23	Y ₁ 135.00 b= -0.19 c= 0.00	k= 11.45 p= 7.61 r= 41.62	y _c 1.91 y _f 9.54 t= 41.62	k= 13.48 r= 23.93	k= 15.12 p= 5.10 r= 42.58
13	Y ₁ 60.70 b= -0.09	Y ₁ 70.78 B= 5.58 b= -0.12	Y ₁ 60.70 M 0.00 b= -0.09	Y ₁ 70.78 M 0.00 B= 5.58 b= -0.12	C= 0.00 Y ₁ 60.67 b= -0.09	Y ₁ 56.00 b= -0.06 c= 0.00	k= 16.37 p= 49.31 r= 50.96	y _c 10.15 y _f 6.22 t= 50.96	k= 15.75 r= 4.97	k= 18.66 p= 29.55 r= 26.28
14	Y ₁ 237.80 b= -0.15	Y ₁ 246.11 B= 10.70 b= -0.13	Y ₁ 239.17 M 0.07 b= -0.17	Y ₁ 246.11 M 0.00 B= 10.70 b= -0.13	C= 16.91 Y ₁ 222.26 b= -0.17	Y ₁ 242.09 b= -0.17 c= 0.00	k= 6.72 p= 58.17 r= 135.31	y _c 2.35 y _f 4.37 t= 135.29	k= 5.85 r= 22.46	k= 7.04 p= 20.43 r= 59.63
15	Y ₁ 190.75 b= -0.25	Y ₁ 314.89 B= 4.28 b= -0.38	Y ₁ 190.75 M 0.00 b= -0.25	Y ₁ 314.89 M 0.00 B= 4.28 b= -0.38	C= 0.00 Y ₁ 187.53 b= 4.28	Y ₁ 85.49 b= -0.02 c= 0.00	k= 10.49 p= 14.91 r= 60.57	y _c 2.29 y _f 8.20 t= 60.57	k= 11.30 r= 29.57	k= 14.70 p= 12.98 r= 71.85
16	Y ₁ 90.49 b= -0.15	Y ₁ 97.90 B= 1.20 b= -0.17	Y ₁ 90.49 M 0.00 b= -0.15	Y ₁ 97.90 M 0.00 B= 1.20 b= -0.17	C= 0.00 Y ₁ 90.39 b= -0.15	Y ₁ 90.49 b= -0.15 c= 0.00	k= 14.78 p= 36.10 r= 35.09	y _c 7.11 y _f 7.68 t= 55.09	k= 13.96 r= 8.40	k= 17.60 p= 22.86 r= 36.76
17	Y ₁ 39.06 b= -0.08	Y ₁ 358.82 B= 132.60 b= -0.48	Y ₁ 39.06 M 0.00 b= -0.08	Y ₁ 358.82 M 0.00 B= 132.60 b= -0.48	C= 0.00 Y ₁ 39.05 b= -0.08	Y ₁ 37.94 b= -0.07 c= 0.00	k= 24.44 p= 53.18 r= 46.25	y _c 16.70 y _f 7.74 t= 46.25	k= 22.49 r= 2.58	k= 27.08 p= 29.31 r= 19.32
18	Y ₁ 162.04 b= -0.13	Y ₁ 162.04 B= 0.00 b= -0.13	Y ₁ 173.97 M 0.33 b= -0.29	Y ₁ 173.09 M 0.33 B= 1.00 b= -0.29	C= 58.09 Y ₁ 115.88 b= -0.29	Y ₁ 175.00 b= -0.01 c= 0.00	k= 7.84 p= 75.88 r= 105.09	y _c 4.03 y _f 3.81 t= 105.09	k= 7.42 r= 12.67	k= 8.83 p= 37.58 r= 52.22
19	Y ₁ 375.31 b= -0.17	Y ₁ 385.59 B= 3.29 b= -0.15	Y ₁ 375.31 M 0.00 b= -0.17	Y ₁ 385.59 M 0.00 B= 3.29 b= -0.15	C= 0.00 Y ₁ 375.31 b= -0.17	Y ₁ 340.00 b= -0.14 c= 0.00	k= 4.00 p= 50.00 r= 122.00	y _c 1.30 y _f 1.60 t= 121.90	k= 3.57 r= 24.35	k= 4.98 p= 45.39 r= 120.08
20	Y ₁ 334.95 b= -0.26	Y ₁ 1050.23 B= 11.68 b= -0.53	Y ₁ 334.95 M 0.00 b= -0.26	Y ₁ 1050.23 M 0.00 B= 11.68 b= -0.53	C= 0.00 Y ₁ 328.05 b= -0.25	Y ₁ 298.00 b= -0.22 c= 0.00	k= 7.35 p= 38.54 r= 157.89	y _c 1.59 y _f 5.58 t= 157.89	k= 6.99 r= 56.61	k= 11.15 p= 37.12 r= 220.62
21	Y ₁ 118.18 b= -0.08	Y ₁ 144.26 B= 10.40 b= -0.12	Y ₁ 118.18 M 0.00 b= -0.08	Y ₁ 144.26 M 0.00 B= 10.40 b= -0.12	C= 0.00 Y ₁ 118.13 b= -0.08	Y ₁ 117.01 b= -0.07 c= 0.00	k= 9.14 p= 161.14 r= 180.05	y _c 5.40 y _f 3.73 t= 180.03	k= 7.62 r= 6.35	k= 9.76 p= 77.90 r= 68.67

5. Considerações Finais

Esta dissertação teve como principal objetivo o desenvolvimento de sistemáticas de formação de agrupamentos de trabalhadores com perfis homogêneos de aprendizado. Com a revisão bibliográfica realizada, objetivos específicos foram definidos. São eles: integrar a ACP às técnicas de clusterização com vistas ao aumento da qualidade dos agrupamentos formados; avaliar o uso de funções *Kernel* para aprimorar o desempenho das ferramentas de clusterização; verificar o desempenho do método proposto através de um experimento de simulação que abranja fatores relevantes para formação de agrupamentos; testar diferentes técnicas de clusterização em termos do seu desempenho para agrupamento de trabalhadores; e por fim, selecionar as variáveis de aprendizado mais relevantes para a clusterização através de um índice de importância gerado por meio de parâmetros oriundos da ACP.

O primeiro e segundo objetivos, (i) integrar a ACP a técnicas de clusterização com vistas ao aumento da qualidade dos agrupamentos formados e (ii) avaliar o uso de funções *Kernel* para aprimorar o desempenho das ferramentas de clusterização, foram atingidos no segundo artigo; o qual apresentou um método que integra CAs, ACP, funções *Kernel* e *Fuzzy C-Means* (FCM) para a formação de grupos de trabalhadores de acordo com seus perfis de aprendizado. As curvas são utilizadas para a geração dos parâmetros de aprendizado de cada trabalhador a ser agrupado. Tais parâmetros são tratados como variáveis de clusterização, as quais são transformadas através da ACP e de funções *Kernel*. A qualidade dos agrupamentos formados foi avaliada através do SI. Dados transformados pela ACP e *Kernel* Polinomial geraram agrupamentos mais apropriados, com SI=0.97 e SI=0.93, respectivamente.

O terceiro objetivo, (iii) verificar o desempenho do método proposto através de um experimento de simulação que abranja fatores relevantes para formação de agrupamentos, foi atingido no artigo 1, o qual utilizou as curvas para a geração de parâmetros que descrevem os perfis de aprendizado dos trabalhadores e remapeou os dados através da ACP com vistas à melhoria da precisão do algoritmo *K-Means*. A qualidade dos agrupamentos formados foi avaliada através do SI. Ao ser aplicado em um estudo de caso, o método na qual os parâmetros são transformados em componentes principais apresentou um índice SI=0,968, valor superior a SI=0,392 obtido com a clusterização tradicional. Para corroborar a eficácia do método, foi realizado um experimento de simulação, o qual comprovou o desempenho satisfatório do método proposto, mesmo com a inserção de fatores de ruído, correlação e diferentes proporções de variáveis por observações.

O quarto objetivo, (iv) testar diferentes técnicas de clusterização em termos do seu desempenho para agrupamento de trabalhadores, foi atingido no segundo artigo, o qual remapeou os dados através da ACP e de funções *Kernel* com o propósito de melhorar o desempenho do algoritmo FCM, os resultados foram comparados aos resultados obtidos no artigo 1 que realizou a clusterização das variáveis latentes através do algoritmo *K-Means*.

O último objetivo: (v) selecionar as variáveis de aprendizado mais relevantes para a clusterização através de um índice de importância gerado por meio de parâmetros oriundos da ACP foi atingido no terceiro artigo, o qual utilizou as CAs para a geração de parâmetros de aprendizado e os avaliou através da ACP com vistas à geração de um índice de importância de variável. Dois procedimentos de clusterização são testados, *K-Means* e FCM, para os quais é iterativamente removida a variável tida como menos importante a cada clusterização. A qualidade dos agrupamentos foi avaliada através do SI, e a eficiência das clusterizações através da Análise Envoltória de Dados (AED). O método apresentou um SI de 0,70 no procedimento *K-Means* e 0,65 no FCM, valores elevados quando comparados a SI=0,31 obtido com a clusterização retendo todas as variáveis disponíveis. Para corroborar a eficácia do método, foi utilizada a AED, que demonstrou que a utilização de um subconjunto de variáveis ou a retenção da variável mais relevante para a clusterização incrementa a eficiência dos agrupamentos tanto no *K-Means* quanto no FCM.

Sendo assim, a partir das sistemáticas propostas nos três artigos desenvolvidos, é possível formar equipes em setores industriais utilizando critérios mais eficientes que consideram o perfil de aprendizado dos trabalhadores envolvidos, como os apresentados nesta dissertação. Formar equipes apoiando-se apenas nos tempos médios de execução de tarefas, visando minimizar a variância entre estes pode não ser o melhor critério, principalmente se o cenário de produção for caracterizado como de Customização em Massa, no qual a produção é feita em lotes de tamanhos reduzidos e grande alternância de modelos. Neste caso, a formação das equipes pode ser mais eficiente considerando os perfis de aprendizado, os quais capturam o desempenho do trabalhador desde o início do processo analisado - e não somente quando o processo de aprendizado está concluído.

O desenvolvimento desta dissertação contribui academicamente com a proposição das sistemáticas apresentadas, porém pesquisas futuras podem ser desenvolvidas como extensões das proposições aqui descritas. São elas: propor métodos para formação de agrupamentos de trabalhadores que incluam o processo de esquecimento e a qualidade dos produtos com a

repetição das tarefas; testar as sistemáticas apresentadas nessa dissertação em outros cenários produtivos; e testar outros algoritmos de clusterização, como Redes Neurais.