

INFERÊNCIA DAS PROPRIEDADES DE ALGODÃO ATRAVÉS DE TÉCNICAS DE *DATA MINING*

Rosa M.Vasconcelos , Luís Amaral & Pedro Bastos

RESUMO

Nos últimos anos, o desenvolvimento tecnológico tornou possível a medição das diferentes propriedades físicas das fibras de algodão, tais como, comprimento, micronaire, grau de uniformidade, alongamento, resistência, cor e grau de impureza. Estes dados são obtidos através da utilização de instrumentos HVI¹, proporcionando resultados rápidos e fiáveis. No entanto, no que diz respeito ao estudo das propriedades químicas, os resultados são obtidos usando métodos laboratoriais mais demorados e dispendiosos.

Este conhecimento é bastante importante, pois as propriedades químicas afectam fortemente o processo de transformação das referidas fibras.

Neste estudo, são usadas várias técnicas de prospecção de dados (*data mining*) disponíveis na ferramenta *Clementine*^[1]. A prospecção de dados consiste num dos passos do processo de extracção de conhecimento em bases de dados (DCBD), processo que tem como principal função a descoberta de relações e associações entre conjuntos de dados. A ferramenta inclui técnicas avançadas de modelação baseadas em inteligência artificial, extraíndo dos dados possíveis relações complexas existentes, bem como regras de associação entre eles.

Este artigo descreve o estudo efectuado utilizando o *Clementine* na análise de dados relativos a propriedades físicas e químicas da fibra de algodão. Os resultados atingidos apontam a forma como as técnicas de prospecção de dados podem ser usadas para assim estabelecer, de forma eficiente, relações que caracterizem as propriedades físicas e químicas das fibras.

PALAVRAS-CHAVE: Propriedades das fibras, Algodão, Descoberta de Conhecimento em Bases de Dados, *Data Mining*.

1. INTRODUÇÃO

O algodão é tradicionalmente a matéria-prima mais utilizada na indústria têxtil, representando cerca de 50%^[2] da produção mundial de fibras têxteis.

Devido ao facto de o algodão ser uma fibra natural, este é um material intrinsecamente heterogéneo, apresentando variações significativas nas propriedades das fibras do mesmo fardo, planta, cápsula e até semente. A variabilidade das suas propriedades são determinadas por factores genéticos e ambientais. O potencial genético de determinado genótipo define-se como o rendimento da colheita ou qualidade das fibras que podem ser atingidos sob condições óptimas de crescimento^[3]. O grau que o potencial genético atinge varia de acordo com alterações ambientais provocadas por flutuações inevitáveis de temperatura, quantidade de luz diurna e insolação ou então pela utilização da rega, de

¹ High Volume Instruments

fertilizantes ou de insecticidas. Outros factores que influenciam as propriedades são: o método utilizado na colheita ou a forma como é efectuado o descaroçamento por exemplo.

Hoje em dia, a maior parte das propriedades físicas já podem ser objectivamente avaliadas, recorrendo a aparelhos e instrumentos de medida, dos quais se destaca o aparelho HVI – *High Volume Instruments* – e o sistema AFIS – *Uster Advanced Fiber Information System* – que, de forma rápida^[4], permitem a caracterização das propriedades físicas das fibras de algodão, substituindo uma série de avaliações subjectivas e ensaios morosos. No entanto, apesar desta avaliação ser extremamente importante não fornece qualquer tipo de indicação relativamente às características químicas das fibras. Para a sua determinação são necessários ensaios laboratoriais morosos e dispendiosos. Devido a este facto, não existem ainda estudos que permitem uma análise mais profunda sobre a influência destas propriedades nem a forma como se podem relacionar as propriedades físicas e químicas das fibras. Estudos nestas area demonstraram que amostras com elevado teor de ceras dão origem a fios e tecidos mais resistentes^[5]. Nickerson demonstrou que variações de 0,15 ou superiores no grau de amarelo afectam o tingimento da fibra^[6]

A análise das propriedades das fibras de algodão tem sido efectuada através de técnicas de análise estatística, as quais permitem a identificação de correlações entre os dados analisados. Neste artigo é apresentada uma nova abordagem, na qual técnicas de *Data Mining* permitem a identificação de relacionamentos implícitos nos dados e a detecção de modelos passíveis de serem utilizados em tarefas de previsão.

As propriedades das fibras recolhidas, armazenadas numa base de dados, integram um conjunto de atributos que são analisados simultaneamente pelos algoritmos de *Data Mining*, permitindo a identificação de relacionamentos que emergem dos dados. Esta abordagem não condiciona ou limita os resultados que podem ser obtidos.

2. PROPRIEDADES DAS FIBRAS DE ALGODÃO

A classificação do algodão segundo o grau, o carácter e o comprimento das fibras utilizando os métodos clássicos de caracterização encontra-se desadequada devido aos padrões de qualidade exigidos pelo sector produtivo. Actualmente, os laboratórios de controlo, quer de produtores, quer de empresas foram re-equipados para que se consiga obter num curto período de tempo os dados sobre as propriedades físicas das fibras de algodão. Os parâmetros de qualidade da fibra de algodão são definidos pelas propriedades físicas mensuradas que são habitualmente o Comprimento, o Índice de Uniformidade, a Resistência, o Índice Micronaire, a Cor, traduzida pela Reflectância (Rd) e Grau de Amarelo (+b) e a Percentagem de Impurezas, todas quantificadas pelo sistema HVI^[7]. A evolução sentida nesta área tem contribuído para uma melhoria significativa dos processos têxteis. No entanto, as variáveis controladas não explicam convenientemente o comportamento destas fibras nos processos de preparação, ultimateção e acabamentos, pelo que as propriedades das mesmas podem ser utilizadas para explicar ou compreender este comportamento. Consideram-se características químicas do algodão, as que se apresentam na tabela 1 e que podem influenciar os tratamentos a molhado utilizados na indústria têxtil explicar algum tipo de comportamento anómalo no processo têxtil e no meio ambiente.

Constituinte	Empire	Wilds	Stoneville	Desconhec Madura	Desconhec Imatura	Desconhec (Mebane)
Celulose (%)	95,30	93,42	95,58	96,41	92,44	94,93
Teor de Açúcares (%)	0,07	0,25	0,18	0,17	0,13	0,11
Açúcares redutores (%)	0,01	0,17	0,04	0,02	0,03	0,02
Azoto (%)	0,17	0,26	0,18	0,16	0,32	0,19
Ceras (%)	0,73	1,07	0,69	0,45	1,14	0,57
Cinzas (%)	0,86	1,32	1,00	0,79	1,32	1,16
Alcalinidade das cinzas	13,97	20,52	14,81	11,75	21,75	16,46
PH do extracto aquoso	7,16	6,23	7,50	7,58	5,84	6,46
Extracção com água (%)	2,86	3,51	2,33	2,00	4,62	3,02

Tabela 1 – Características de algodões de diferentes origens, em matéria seca ^[8].

Embora todas estas características influenciem o processo têxtil, a sua determinação envolve períodos extremamente morosos. Neste estudo, o objecto de análise recaiu unicamente sobre o teor de cinzas e dos seus constituintes.

O algodão retirado directamente do descaroçador apresenta uma percentagem de cinzas entre 2 a 3%, provavelmente devido a fragmentos da cápsula, bem como a poeiras do solo que se aderem às fibras ^[9]. Análises às cinzas demonstram que os seus principais constituintes são o magnésio, cálcio, carbonato de potássio, fosfatos, sulfatos ou cloretos. Estes carbonatos são resíduos de metais originalmente presentes ou de ácidos orgânicos. As operações de fervura e branqueamento reduzem as cinzas para menos de 0,1%. As cinzas são fortemente alcalinas, uma propriedade típica das cinzas das plantas. Cerca de 85% do teor das cinzas podem ser removidas da fibra por extracção com água, mas a maior parte dos sais de cálcio, ferro e alumínio permanece na fibra ^[8].

3. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Extrair conhecimento de bases de dados pode envolver, entre outras coisas, a utilização de algoritmos de aprendizagem máquina capazes de generalizar os factos encontrados num grande repositório de dados com a forma de regras de alto nível, compreensíveis ao ser humano ^[10]

As técnicas de aprendizagem máquina parecem estimular bastante os investigadores da área e talvez seja a parte do processo de DCBD que contenha, substancialmente, os maiores conteúdos científicos estudados. Todavia, quando se tenta realizar tarefas de descoberta de conhecimento, em aplicações do mundo real, percebe-se que as mesmas podem ser extremamente complexas e que a tarefa de prospecção de dados representa apenas uma porção (importante) do processo global ^[11].

O processo de descoberta de conhecimento em bases de dados é interactivo e iterativo, envolvendo inúmeros passos com várias decisões a serem tomadas pelo utilizador. Brachaman & Arnand ^[12], dão uma visão prática e sobre todo o processo de DCBD, dando ênfase à natureza interactiva do processo (figura 1).

(figura 1).

Assim, o processo de DCBD deve ser visualizado como um conjunto de várias etapas interligadas entre si.

O processo de descoberta de conhecimento em bases de dados é interactivo, envolvendo diversas etapas, como é demonstrado na figura 1, embora esta visão seja um pouco generalista.

Assim, se analisarmos em pormenor todo o processo de DCBD, podemos deslumbrar este mesmo processo de uma forma um pouco distinta e mais aprofundada do que está descrito e visualizado na figura 1.

Considera-se os seguintes passos de forma a efectuar a extracção de conhecimento ^[13]:

1. Compreensão do domínio de aplicação;
2. Selecção de dados;
3. Pré-processamento dos dados;
4. *Data mining*;
5. Interpretação e validação do conhecimento;
6. Consolidação do conhecimento.

As etapas do processo de DCBD, normalmente, são realizadas de forma sequencial, ou seja, é necessário compreender o domínio de aplicação, seleccionar e pré-processar os dados por forma a se encontrar padrões nesses mesmos dados. E por se tratar de um processo interactivo, as pessoas envolvidas na sua realização podem possuir um canal de comunicação que viabilize uma boa troca de informações. Estes utilizadores possuem diferentes especialidades e necessidades e cada um possui o seu papel dentro do processo. As classes dos usuários envolvidos no processo de DCBD são ^[14]:

- Especialista do domínio: possui ampla compreensão do domínio da aplicação. Por exemplo, durante a análise de um conjunto de dados sobre matéria-prima de uma organização, o especialista poderá ser o gestor de stocks.
- Analista: executa o processo de DCBD e, portanto, deve possuir um amplo conhecimento das etapas que compõem esse processo. Por exemplo, analistas de sistemas, estatísticos e administradores de bases de dados.
- Utilizador final: pode ou não possuir amplo conhecimento do domínio. É frequentemente aquele que utiliza o conhecimento extraído no processo de DCBD para auxiliá-lo num processo de tomada de decisão.

Durante a realização do processo de DCBD, a interacção e comunicação entre o analista, o especialista do domínio e o utilizador final é muito importante, uma vez que o sucesso da extracção de conhecimento depende da comunicação entre eles ^[12]. Nesse sentido, é possível melhorar a troca de informações entre os envolvidos no processo, utilizando-se técnicas de visualização de dados adequados para auxiliar na execução das diferentes etapas do processo ^[15].

Estas etapas sugeridas, podem não cobrir exhaustivamente todos os passos existentes, no entanto, esta sequência de tarefas pode ser considerada como consistindo nos requisitos mínimos de forma a se conseguir extrair conhecimento de bases de dados.

4. TRABALHO REALIZADO

Na realização deste estudo utilizou-se a ferramenta *Clementine*^[1], que integra um conjunto de algoritmos de modelação e aprendizagem, com uma interface de programação visual que torna compreensível todo o processo de extracção de conhecimento a partir de diversas fontes de dados. Numa primeira fase, iniciou-se a construção do Armazém de Dados propriamente dito. O processo de DCBD^[16] pode ser assim dividido em dois níveis: um primeiro nível, do qual fazem parte os passos relativos à selecção e pré-processamento dos dados, em que a partir do ambiente operacional são capturados e transformados os dados que depois constituem um Armazém de Dados; e um segundo nível, em que a partir dos dados no Armazém de Dados podem ocorrer novas transformações e em que é efectuada a captura de padrões de relacionamento

O objectivo do Armazém de Dados é o de proporcionar as infra-estruturas necessárias de armazenamento, acesso e manipulação de dados, organizando os dados operacionais de forma a otimizar o processo de extracção de conhecimento, funcionando como uma plataforma de suporte para as aplicações de prospecção de dados

No entanto, independentemente do tamanho da tabela referente aos dados químicos ser menor, efectuou-se a divisão das bases de dados, para treino e para validação, numa proporção de 60% por 40%. Para se proceder ao início do trabalho de prospecção de dados tornou-se necessário agrupar os dados químicos existentes, com os todos os dados físicos recolhidos. É de notar o menor número de dados químicos existentes, o que poderá levar a que os resultados finais não atinjam os níveis de certeza desejados.

Todo o armazém de dados é mantido em uma base de dados *Access*. A tabela relativa aos dados físicos é composta por 5936 registos de várias origens: USA, Costa do Marfim, Moçambique, Sudão Rússia, Brasil, Egipto, Zimbabwe, Camarões, Madagáscar, Uganda, China, Benin e Tchad. No entanto e devido aos avultados custos financeiros que acarreta a elaboração de testes químicos, bem como da morosidade dos referidos testes, a todos os fardos de algodão provenientes das diferentes origens, só foi possível efectuar testes sobre oito dessas origens, tendo sido criada uma nova tabela referente aos dados químicos, com um total de 197 registos, referentes a fardos provenientes do Brasil, Rússia, Egipto, Zimbabwe, Camarões, Madagáscar, Marrocos e Uganda. A figura 2 mostra um excerto da tabela referente aos dados das propriedades físicas das fibras de algodão.

figura 2

O processo de junção das duas tabelas iniciais, contendo os registos referentes aos dados físicos e químicos, bem como a criação das duas novas tabelas com os registos agrupados, está descrito na figura 3. Como resultado foram obtidas duas novas tabelas, sendo uma designada de *Treino* e outra de *Validação*. A tabela para treino vai ser usada para criar e refinar escalões que serão gerados e para aplicar os modelos considerados como os mais adequados. Após o treino, será efectuada a validação, utilizando a tabela de validação, com o objectivo de se validar os modelos obtidos.

A característica fundamental desta *stream* foi a necessidade de aplicação de um gráfico de distribuição dos registos por origem, após a junção de todos os registos efectuada através de uma operação de *merge* e assumindo-se como chave primária o campo origem. Com um gráfico de distribuição gerou-se um ícone de balanceamento (*balance*) para assim garantir a presença de registos relativos a todas as origens nas duas tabelas geradas, Treino e Validação.

figura 3.

As propriedades químicas em estudo são: percentagem de cinzas (%cinzas), Potássio (K), o Magnésio (Mg), o Fósforo (P), Cálcio (Ca) e a percentagem de materiais extraídos (ME). Assim, nesta parte do estudo pretende-se prever o comportamento das várias variáveis de entrada para as variáveis de saída definidas.

É importante salientar que as variáveis Potássio, Magnésio, Fósforo e Cálcio, são obtidas após análise das cinzas dos diferentes fardos de algodão, provenientes das diferentes origens em estudo. Convém também salientar que os elementos constituintes das cinzas constituem um espectro de propriedades muito mais alargados, no entanto, estes são os elementos existentes em maior percentagem e comuns a todos os algodões, sendo assim, dentro desse conjunto alargado de elementos, considerados como os mais relevantes.

Analisando os registos das propriedades químicas existentes, verificou-se que são definidas por um conjunto alargado de valores, o que tornou necessária a criação de escalões que tornassem possível aplicação de modelos necessários para o desenvolvimento do estudo. Assim e através da *stream cria_esc_quim.str* (figura 4), utilizando histogramas (*histogram*), foram criados vários nós de derivação (*derive nodes*), referentes, cada um deles, à propriedade analisada em cada histograma. Como resultado final foram obtidas novas tabelas com novos campos: *esc_%cinzas*, *esc_ME²*, *esc_K*, *esc_Mg*, *esc_P* e *esc_Ca*.

figura 4

A *stream* aplicou-se, quer aos dados de treino, quer aos dados de validação, tendo como resultado final, duas novas tabelas de nome: *treino_esc_quimicos* e *validacao_esc_quimicos*.

Os escalões criados relativamente às propriedades químicas podem ser analisados na tabela 2.

Escalões	A	B	C	D
% Cinzas	< 1.23	[1.23;1.37 [[1.37;1.46 [>= 1.46
% Material Extraído	< 7.28	[7.28;7.82 [[7.82;10.02 [>= 10.02
Potássio (K)	< 55.64	[55.64;67.04 [[67.04;69.57 [>= 69.57
Magnésio (Mg)	< 4.58	[4.58;5,00 [[5,00;6.32 [>= 6.32
Cálcio (Ca)	< 15.02	[15.02;17.49 [[17.49;23.29	>= 23.29
Fósforo (P)	< 2.53	[2.53;3,00 [[3,00;4.30 [>= 4.30

Tabela 2 – Escalões químicos criados

De seguida, e para melhor compreensão da forma como as fibras provenientes das diferentes origens estão relacionadas com as propriedades químicas em estudo, gerou-se a *stream relacao.str* (figura 5).

figura 5.

A primeira fase da análise consistiu na utilização de Web Nodes para a identificação de associações entre os atributos. Os Web Nodes são gráficos disponibilizados pelo Clementine que permitem explorar os dados e verificar se existem associações entre os mesmos. Nesta fase de exploração analisaram-se os atributos *origem* (origem e

² Materiais Extraídos

percentagem de cinzas (escalao_cinzas). A figura 6 evidencia o Web Node obtido. Nestes gráficos, as linhas mais grossas representam relações fortes entre os atributos e as linhas mais finas ou tracejadas representam relações mais fracas.

figura 6

Após a exploração dos dados procedeu-se à análise dos dados com técnicas de *Data Mining*. A primeira técnica de *Data Mining* utilizada foi a indução de **árvores de decisão**. O algoritmo utilizado foi o C5.0, o qual permitiu a obtenção de um conjunto de regras que descrevem relacionamentos entre as diversas propriedades das fibras armazenadas na base de dados. Diversos modelos foram gerados, com o objectivo de identificar regras que permitissem prever cada uma das propriedades em estudo. A figura 7 apresenta parte das regras obtidas no modelo para a previsão do atributo origem.

figura 7

5 DISCUSSÃO DE RESULTADOS

Após a análise de todos os resultados obtidos através da aplicação das *streams* criadas no *Clementine*, foram elaboradas um conjunto de tabelas por forma a verificar as possíveis relações existentes entre cada propriedade física considerada relevante, com cada uma das propriedades químicas em estudo.

Na tabela 3, apresenta-se a título de exemplo, uma síntese dos resultados obtidos para a determinação da origem das fibras em função das propriedades químicas. Salienta-se que apenas foram consideradas como relevantes as regras com maior significância estatística.

Origem	Escalões					
	% Cinzas	ME	K	P	Ca	Mg
Brasil	C, D	B	C	B	B	A
Zimbabwe	A, B	B	D	B	A	B
Rússia	B, C	A	B	C	C	C
Camarões	B, C	C	D	A	B	A
Madagáscar	C, D	B	C	B	C	C
Uganda	B, C, D	A	B	C	B	A
Marrocos	A	D	A	D	D	D
Egipto	A, B	A	A	D	D	D

Tabela 3 – Escalões químicos por origem.

A tabela 4 apresenta-se uma síntese das propriedades físicas mais relevantes na determinação das propriedades químicas das fibras de algodão.

	+b	RD	SL2.5	SL50	UI	ST	PSI	EL	Micronaire
%Cinzas	x	x							x
Ca	x	x	x		x	x			
K		x			x				
ME	x	x				x	x		x
Mg			x				x		
P	x	x		x			x	x	

Tabela 4 – Influência entre as propriedades físicas e químicas

Através da análise da tabela anterior, podemos verificar que as propriedades físicas grau de amarelo (+b) e a reflectância (RD) influenciam quase a totalidade das propriedades químicas, à excepção do Potássio (K) em relação ao grau de amarelo e do Magnésio (Mg) em relação a ambas.

6 CONCLUSÕES E TRABALHO FUTURO

Com este trabalho, abrangendo duas áreas de investigação muito distintas, procurou-se mostrar a aplicação de técnicas de prospecção de dados na análise de relações entre as propriedades físicas e químicas das fibras de algodão, procurando-se prever as características químicas a partir das características físicas das mesmas.

Com o resultado da análise à base de dados proveniente das diferentes origens dos fardos de algodão, contendo os dados físicos e químicos, reconheceu-se a importância que este tipo de ferramentas pode ter na descoberta de inferência das propriedades do algodão. Comparativamente a outras ferramentas genéricas para a análise das propriedades do algodão, quer de forma individual, quer de forma generalizada, os sistemas de prospecção permitem técnicas de análise de dados mais sofisticada e orientadas para a descoberta de conhecimento útil aos diferentes níveis decisionais

Extrair mais e melhor conhecimento exige um trabalho significativo de selecção e preparação dos dados. Neste aspecto, é importante realçar todo o trabalho realizado na preparação da base de dados que serviu de suporte a todo o trabalho elaborado, visto a enorme variedade de fontes relativas aos dados físicos das fibras e a discrepância entre o número de registos relativos as propriedades físicas e químicas.

Com o estudo das relações entre algumas propriedades do algodão determinadas ao longo da realização deste trabalho, foi possível obter um maior conhecimento acerca das técnicas de prospecção de dados, através da ferramenta utilizada, bem como sobre esta fibra natural e a consequente verificação que os algodões analisados possuem características próprias relativamente à sua composição física e química, que permitem efectuar uma classificação das fibras consoante a sua proveniência.

Os resultados atingidos demonstram que através do *Clementine* e aplicando técnicas de prospecção de dados, podemos identificar, analisar e prever relacionamentos entre as diferentes propriedades do algodão e a posterior inferência das propriedades químicas da fibra a partir das propriedades físicas existentes, já que os resultados obtidos com este estudo assentam na obtenção de um conjunto de regras uniformes e com um grau de confiança elevado, através da aplicação de diferentes modelos existentes na ferramenta utilizada.

REFERÊNCIAS

- (1) “Clementine Data Mining System” : Reference manual version 5. 1994-1998.
- (2) Vasconcelos, R. M. “Planeamento de uma encomenda”, Nova Têxtil. Nº8: Abril (1988), pp52-55.
- (3) Bradow, J. M. e G. H. Davidonis "Quantitation of Fiber Quality and the Cotton Production-Processing Interface: A Physiologist's Perspective." *Journal of Cotton Science* 4: 2000, pp 34-64.
- (4) Furter, R. e K. Douglas (1990). “A medição do comprimento da fibra de algodão, do diâmetro e do conteúdo e de neps com os instrumentos AFIS”. *Ramas de algodão: Propriedades Físicas e Químicas*. In S. Lucas, Engenharia Têxtil. Guimarães, Universidade do Minho. Tese de Mestrado, 2000
- (5) Taylor, R.A. “Natural Waxes on Cotton Contribute to yarn and fabric quality”. *Textile Chemist and Colorist*. Volume 29. June 1990, pp32-35.
- (6) Aguiar, A. “Efeito do branqueamento nas propriedades do algodão” Parte 1- A fibra de algodão. Nova Têxtil nº 46 Outubro 1997, pp71-77.
- (7) Vasconcelos, R. M. “Contribuição à aplicação de técnicas de inteligência artificial na tecnologia da fição”, Universidade do Minho - Tese de Doutoramento, 1993.
- (8) Lucas, F. S. “Ramas de algodão: Propriedades Físicas e Químicas”. Engenharia Têxtil. Guimarães, Universidade do Minho. Tese de Mestrado, 2000.
- (9) Hamby, S. D. “The American Cotton Handbook”, Volume one, third edition, 1965.
- (10) Dias, S; Vasconcelos, R. M.; Santos, M.; Amorim, T.; Amaral, L. – “Análise das Propriedades Físicas do Algodão Recorrendo a Técnicas de Data Mining”. Nova Têxtil. Nº68: 2º trimestre 2003, pp51-56.
- (11) Abony, J. e Szeifert, F. “Computational Intelligence in Data Mining”. Veszprem, Hungary, University of Veszprem, Department of Process Engineering. 2000.
- (12) Brachman, R. e T. Arnanand "The Process of Knowledge Discovery in Databases." AAI/MIT Press: 1996, pp37-58.
- (13) Fayyad, U., G. Piatetsky-Shapiro, et al. “Knowledge Discovery and Data Mining: Towards a Unifying Framework”. Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon. 1996.
- (14) Dias, S; Vasconcelos, R. M.; Santos, M.; Amorim, T.; Amaral, L. – “Using Clementine Data Mining System in the Process of Analysis of Cotton Fiber Properties”, 9th International Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems, Annecy, França: Julho 2002. ISBN 2-9516453-2-5. pp1881-1887.

- (15) Rezende, S. O., R. B. T. Oliveira, et al. "Visualization for Knowledge Discovery in Database." Ebecken: 1998, pp 81-95.
- (16) Baptista, P. e M. J. Silva "Prospecção de Dados de Acesso a um Servidor de Notícias na Web". Lisboa, Departamento de Informática. 2000.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela Fundação para a Ciência e Tecnologia, com o contrato nº POCTI / 1999 / CTM / 32993.

Rosa M. Vasconcelos-Departamento de Engenharia Têxtil- Universidade do Minho, Campus de Azúrem 4800-058 Guimarães Portugal
 Tel. +351253510292 Fax +351253510293 - rosa@det.uminho.pt

Luis Amaral -Departamento de Sistemas de Informação Universidade do Minho, Campus de Azúrem 4800-058 Guimarães Portugal Tel. +351253510303 Fax +351253510300-
amaral@dsi.uminho.pt

Pedro Bastos –Instituto Politécnico de bragança - Campus de Sta Apolónia Apartado 1131 5301-057 Bragança Portugal Fax +351253510300- bastos@ipb.pt

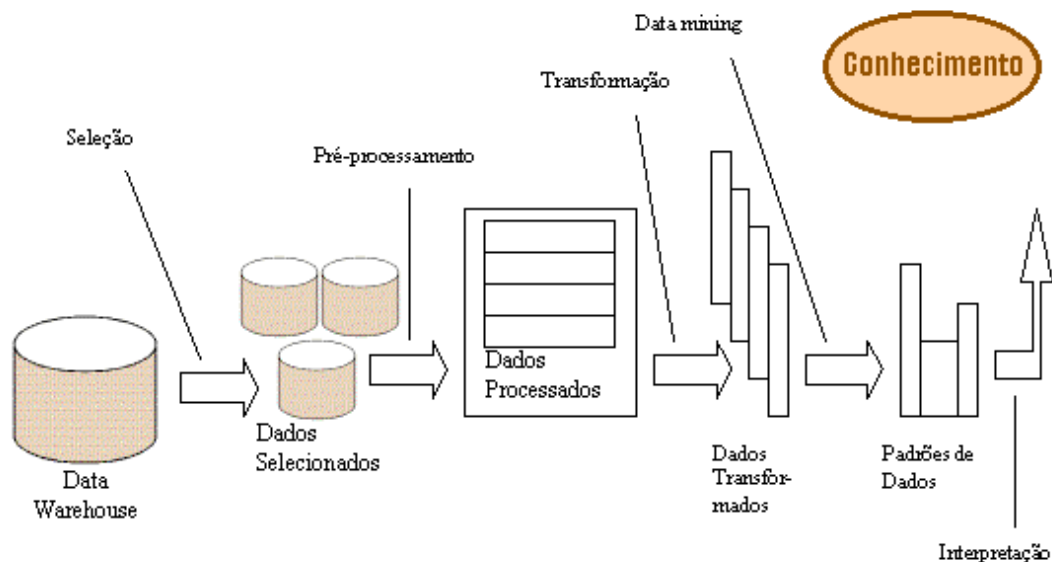


Figura 1 – Visão generalista do processo de DCBD (Abonyi e Szeifert, 2000).

The screenshot shows the Microsoft Access interface with a table named 'fisicos_v2' open. The table has the following columns and data:

ID_fisicos	origem	fardo	micronaire	UHM	SL50	SL2,5	UI	UR
1	usa	0	4.2	34.3	25.7	34.3	91.2	74.9
2	usa	1	3.8	36.3	21.6	36.1	87.6	60
3	usa	2	4.2	33.5	24.6	33.3	91.6	73.8
4	usa	3	3.9	33.8	21.1	33.8	86.5	62.6
5	usa	4	4.2	34	27.7	33.8	93.5	81.3
6	usa	5	4.1	35.3	26.9	35.1	93.3	76.9
7	usa	6	4.4	36.1	27.2	35.8	90.6	75.3
8	usa	7	4.5	33.3	22.4	33	86.3	67.5
9	usa	8	4.1	34	23.6	34	90.1	70
10	usa	9	4	33.8	20.3	33.3	87.5	61.1
11	usa	10	4	35.8	25.7	35.8	89.8	70.3
12	usa	11	4.1	34.3	22.9	34.8	89.3	65.5
13	usa	12	3.8	34	23.1	34	87.8	67.6
14	usa	13	4.1	35.3	27.9	35.1	93.5	79.6
15	usa	14	4.1	35.1	23.6	34.5	89.5	68.5
16	usa	15	4.5	34.5	25.4	34.3	93.1	73.8
17	usa	16	4.6	35.3	26.2	34.8	90.4	75.1
18	usa	17	4.5	35.6	26.4	35.3	91.6	74.5
19	usa	18	4.1	33.5	22.4	33.5	88.9	66.2
20	usa	19	3.8	34.5	26.4	34.3	91.3	77.2
21	usa	20	4.3	33.8	23.4	33	90	70.7
22	usa	21	4.4	36.6	25.9	36.8	89.3	70.9

Figura 2 – Excerto da tabela de dados referentes às propriedades físicas.

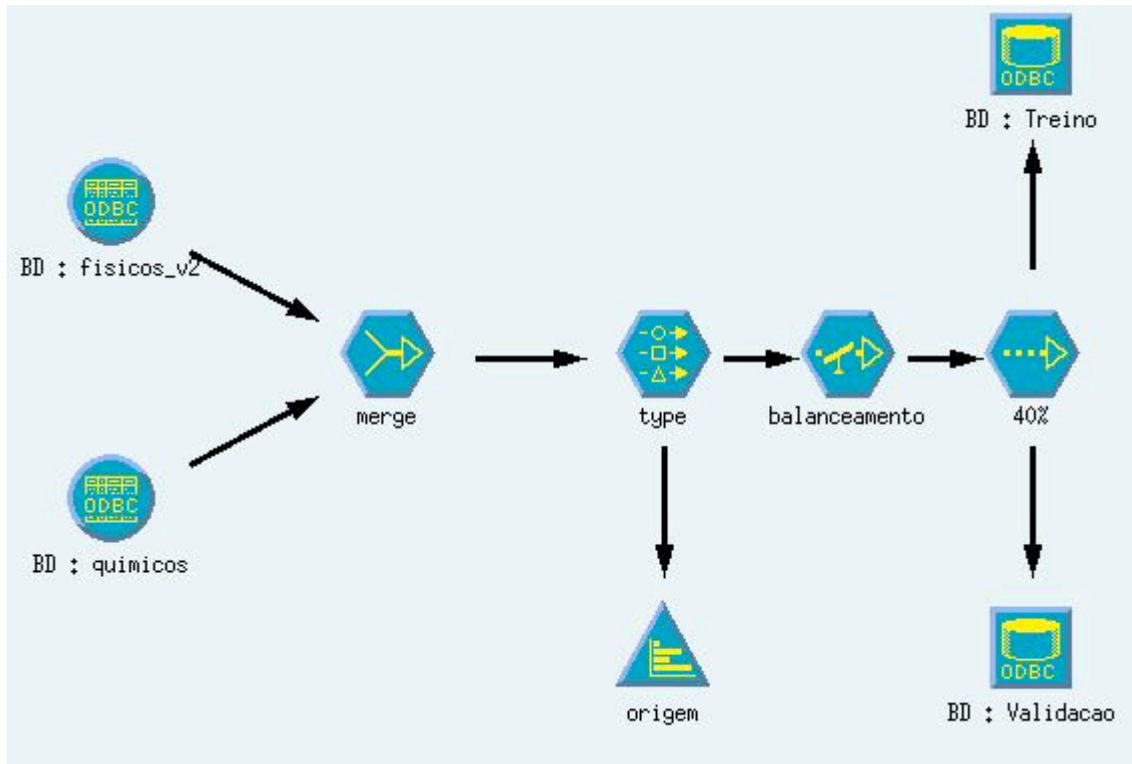


Figura 3 – Stream separacao.str

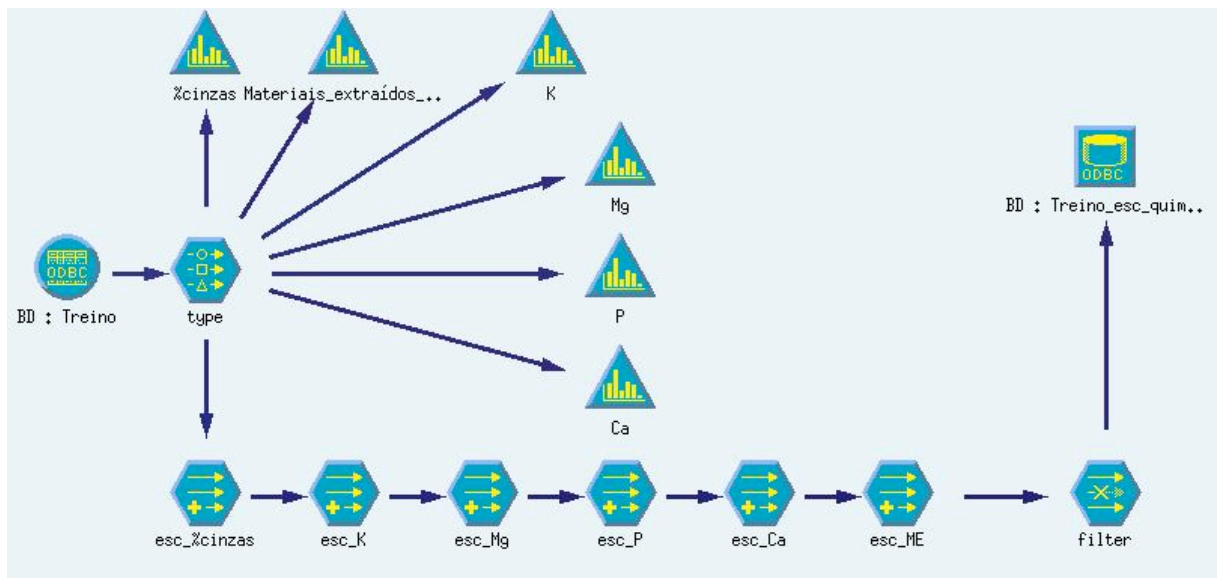


Figura 4– Stream cria_esc_quim.str

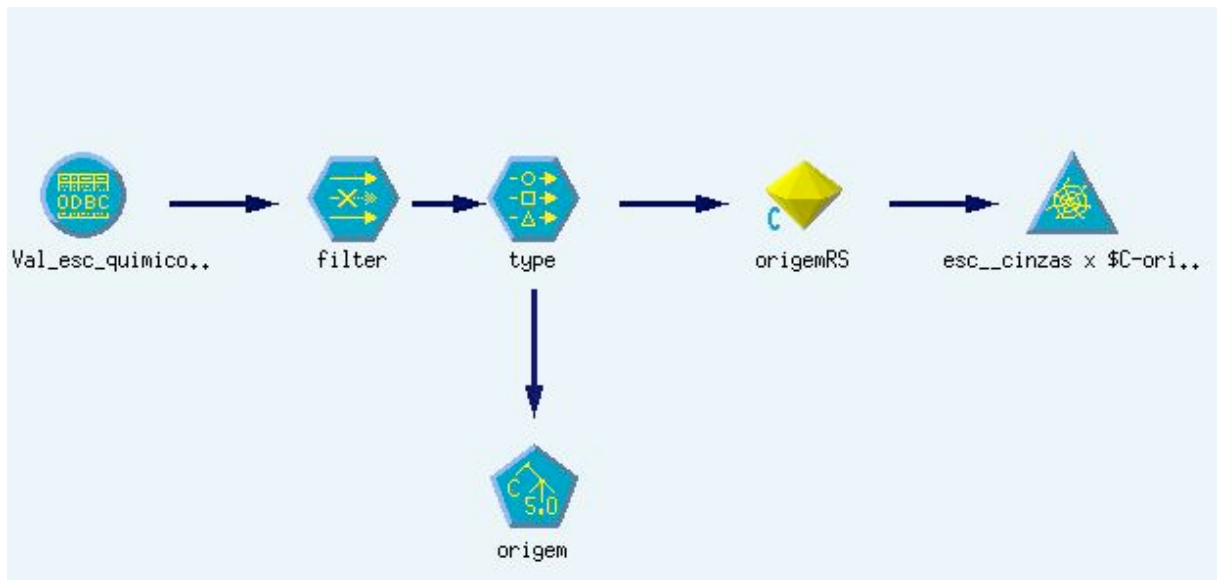


Figura 5 – Stream relacao.str (utilizando o modelo C5.0)

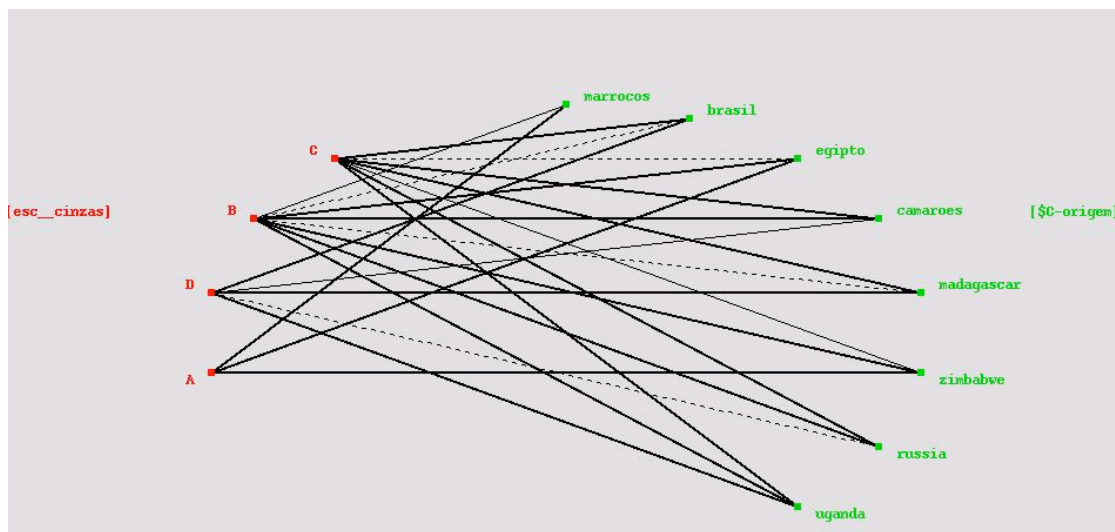


Figura 6 – Web node: análise da origem, versus % de cinzas

Origem vs. Esc_cinzas

Rule #1 for brasil:
 if esc_P == B
 and esc_Mg == A
 then -> brasil
 Rule #2 for brasil:
 if esc_P == B
 and esc_Mg == D
 then -> brasil

Rules for camaroes:
 Rule #1 for camaroes:
 if esc_P == A
 then -> camaroes
 Rules for egipto:
 Rule #1 for egipto:
 if esc_P == D
 and esc_ME == A
 then -> egipto

Rules for madagascar:
 Rule #1 for
 madagascar:
 if esc_P == B
 and esc_Mg == C
 then -> madagascar
 Rules for marrocos:
 Rule #1 for marrocos:
 if esc_P == D

```

    and esc_ME == [B
C] then -> marrocos
Rule #2 for marrocos:
  if esc_P == D
  and esc_ME == D
  then -> marrocos
Rules for russia:
  Rule #1 for russia:
    if esc_P == C

```

```

    and esc_Mg == [B
D] then -> russia
Rule #2 for russia:
  if esc_P == C
  and esc_Mg == C
  then -> russia
Rules for uganda:
  Rule #1 for uganda:
    if esc_P == C

```

```

    and esc_Mg == A
  then -> uganda
Rules for zimbabwe:
  Rule #1 for
zimbabwe:
  if esc_P == B
  and esc_Mg == B
  then -> zimbabwe

```

Figura 7 – Previsão do atributo origem : algumas regras