

UNIVERSIDADE FEDERAL DE ALFENAS

INSTITUTO DE CIÊNCIAS EXATAS

Núcleo de Ciência da Computação

Mateus Neves Barreto

**UMA ABORDAGEM HÍBRIDA PARA IDENTIFICAÇÃO DE
PADRÕES DE ESCOAMENTO BIFÁSICO**

Alfenas, 29 de Novembro de 2011

UNIVERSIDADE FEDERAL DE ALFENAS

INSTITUTO DE CIÊNCIAS EXATAS

Núcleo de Ciência da Computação

**UMA ABORDAGEM HÍBRIDA PARA IDENTIFICAÇÃO DE
PADRÕES DE ESCOAMENTO BIFÁSICO**

Mateus Neves Barreto

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Alfenas como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Ricardo Menezes Salgado

Alfenas, 29 de Novembro de 2011

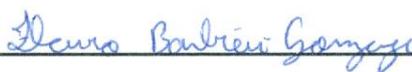
Mateus Neves Barreto

UMA ABORDAGEM HÍBRIDA PARA IDENTIFICAÇÃO DE PADRÕES DE ESCOAMENTO BIFÁSICO

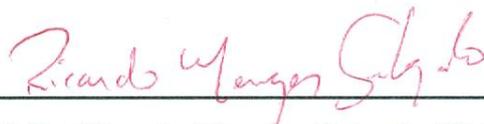
A Banca examinadora abaixo-assinada aprova a monografia apresentada como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação pela Universidade Federal de Alfenas.



Prof. Msc. Luiz Eduardo da Silva
Universidade Federal de Alfenas



Prof. Msc. Flavio Barbieri Gonzaga
Universidade Federal de Alfenas



Prof. Dr. Ricardo Menezes Salgado (Orientador)
Universidade Federal de Alfenas

Alfenas, 29 de Novembro de 2011

Dedico este trabalho ao meu orientador e AMIGO Ricardo Menezes Salgado, que através de minha trajetória acadêmica me apoiou em todos os meus projetos. Cobrou-me em momentos certos e possibilitou de forma direta a finalização deste trabalho. Obrigado!

AGRADECIMENTO

Agradeço as pessoas que me ajudaram a atingir mais um objetivo da vida. Agradeço aos meus Pais, Adilson e Elisabete, por terem me oferecido todo o suporte familiar, afetivo e financeiro. Aos meus irmãos, Mariana, Maiara e Pedro, por me apoiarem. A minha namorada, Talita, por estar comigo em todos os momentos difíceis e alegres.

Agradeço todo o corpo docente do Curso de Bacharelado de Ciência da Computação da Universidade Federal de Alfenas. Aos professores que marcaram esta etapa de minha vida: Ricardo Menezes, Mariane, Flavio, Paulo Bressan, Humberto, Luiz Eduardo, Rodrigo, Thomás, Ricardo Marques, Nelson e Eliseu. A minha amiga e Professora, Rejane Siqueira, que por toda a minha Graduação me ajudou.

Também sou grato aos meus amigos e companheiros Fernando, Marcelo, Douglas, Gustavo, Rodrigo, Tadeu e Mateus. Escrevemos histórias que nunca vamos esquecer!

Ao Laboratório de Inteligência Computacional (LInC), não tenho como descrever o meu agradecimento por ter me acolhido durante todos estes anos.

Aos meus AMIGOS que me acompanharam e “sofreram” comigo durante toda a Graduação, deixo o meu muito Obrigado!

Resumo

Sistemas de suporte à decisão ou *Decision Support Systems* (DSS) são sistemas capazes de auxiliar grandes linhas de produções além de tarefas que apresentam riscos e muitas outras atividades que requerem alto teor de confiabilidade. A utilização dos DSS é feita em setores que necessitam de otimizações e também em setores que possuem acesso limitado. Um exemplo que pode ser citado é a extração do petróleo, geralmente feita em alto mar em grandes profundidades. Seria impossível o exercício desta atividade sem a utilização de um DSS.

A proposta deste trabalho foi desenvolver um Sistema de classificação de padrões acoplado a um DSS específico para o escoamento bifásico vertical e horizontal, baseado em um combinador de algoritmos inteligentes (*ensemble*). As técnicas de Inteligência Artificial (IA) utilizadas no combinador foram: as Redes Neurais Artificiais (RNAs); Algoritmo de K-Vizinhos e o algoritmo *Support Vector Machines* (SVM). A base de dados utilizada foi fornecida pelo Laboratório de Fenômenos Multifásicos da Faculdade de Engenharia Mecânica (FEM) da Unicamp.

A classificação dos dados verticais e horizontais, utilizando a metodologia ensemble, teve entre 96 a 100% de acerto quando vistos separadamente e aproximadamente 91% de acerto para os conjuntos compostos pelos dois tipos de dados, o que demonstra a alta confiabilidade do DSS proposto. A confiabilidade observada está relacionada também ao fato de que um sistema bem projetado e com interfaces amigáveis, fornece base e subsídios para a obtenção de lucros e maior segurança no setor de sua aplicação.

Palavras-Chave: Sistema de Suporte à Decisão. Escoamento Bifásico. Modelos Inteligentes. *Ensembles*. DSS. Framework Inteligente para Escoamento.

ABSTRACT

Decision Support Systems (DSS) are systems that can help large production lines as well as tasks that have risks and many other activities that require high levels of reliability. The use of DSS is made in sectors that require optimizations and also in areas that have limited access. An example can be cited is the extraction of oil, usually made on the high seas at great depths. It would be impossible exercise this activity without the use of a DSS.

The purpose of this study was to develop a classification system coupled to a DSS specific to the vertical and horizontal two-phase flow, based on a combined intelligent algorithms (ensemble). The Artificial Intelligence (AI) techniques were used in the combiner: Artificial Neural Networks (ANN), algorithm k-Nearest Neighbor (KNN) and the algorithm Support Vector Machines (SVM). The database used was provided by the Laboratory of Multiphase Phenomena, Faculty of Mechanical Engineering (FEM) of Unicamp.

The classification of vertical and horizontal data, using the methodology ensemble, had between 96 to 100% correct when viewed separately and approximately 91% of accuracy for the sets composed of two types of data, which demonstrates the high reliability of the DSS proposed. The observed reliability is also related to the fact that a system well designed with user-friendly interfaces, provides the basis and subsidies for obtaining profits and the increased security in the sector of your application.

Keywords: Decision Support System. Two-phase flow. Intelligent models. Ensembles. DSS. Framework Intelligent Flow.

LISTA DE FIGURAS

FIGURA 1 – Regimes de escoamento gás/líquido vertical.....	18
FIGURA 2 – Padrões Cheio e Vazio.....	19
FIGURA 3 – Padrões de Escoamento Horizontal gás/líquido.....	20
FIGURA 4 – Modelo Relacional do Sistema <i>SupportSystemFlow</i>	26
FIGURA 5 – Relação Aquisição.....	29
FIGURA 6 – Relação Padrão.....	30
FIGURA 7 – Relação Aparato Experimental.....	30
FIGURA 8 – Relação Fase Líquida.....	31
FIGURA 9 – Relação Fase Gás.....	31
FIGURA 10 – Relação Responsável.....	32
FIGURA 11 – Relação Estação.....	32
FIGURA 12 – Relação Dado.....	33
FIGURA 13 – Relação Momentos Estatísticos.....	34
FIGURA 14 – Relação Usuário.....	35
FIGURA 15 – Sinal Original Escoamento Bifásico.....	36
FIGURA 16 – Redução de dimensionalidade I.....	37
FIGURA 17 – Redução de dimensionalidade II.....	37
FIGURA 18 – Sinal Reduzido Escoamento Bifásico.....	38
FIGURA 19 – PDFs dos Sinais Original e Reduzido.....	39
FIGURA 20 – Layout da Interface Gráfica com Usuário.....	40
FIGURA 21 – Tela Principal <i>SupportSystemFlow</i>	41
FIGURA 22 – Inicialização <i>SupportSystemFlow</i>	42
FIGURA 23 – Tela <i>Login SupportSystemFlow</i>	43
FIGURA 24 – Tela Principal Liberada <i>SupportSystemFlow</i>	43
FIGURA 25 – Apresentação dos Dados <i>SupportSystemFlow</i>	44
FIGURA 26 – Componente de Importação <i>SupportSystemFlow</i>	45
FIGURA 27 – Processo de Importação de Dados <i>SupportSystemFlow</i>	46
FIGURA 28 – Módulo de configuração de Conexão do BD <i>SupportSystemFlow</i>	47
FIGURA 29 – Menu de Gerenciamento do BD <i>SupportSystemFlow</i>	47
FIGURA 30 – Nomeando Estudo Inteligente - <i>SupportSystemFlow</i>	48
FIGURA 31 – Módulo Criação Estudo Inteligente - <i>SupportSystemFlow</i>	49

FIGURA 32 – Configuração Conjuntos Manualmente - <i>SupportSystemFlow</i>	50
FIGURA 33 – Configuração Conjuntos Aleatoriamente - <i>SupportSystemFlow</i>	50
FIGURA 34 – Treinamento Estudo - <i>SupportSystemFlow</i>	51
FIGURA 35 – Exportando Treinamento - <i>SupportSystemFlow</i>	52
FIGURA 36 – Busca Exaustiva - <i>SupportSystemFlow</i>	52
FIGURA 37 – Rede Neural Artificial.....	54
FIGURA 38 – Estratégia de Treinamento Rede Neural.....	55
FIGURA 39 – Classificação SVM.....	57
FIGURA 40 – <i>Ensemble</i>	58
FIGURA 41 – Rede Neural Artificial – Saída Única.....	67
FIGURA 42 – Rede Neural Artificial – Saída Múltipla.....	67
FIGURA 43 – Classificação K-Vizinhos.....	68
FIGURA 44 – <i>Ensemble</i> Híbrido.....	70
FIGURA 45 – Gráfico Tempo de Consulta I.....	72
FIGURA 46 – Gráfico Tempo de Consulta II.....	72
FIGURA 47 – Gráfico de Análise das Repetições.....	78

LISTA DE TABELAS

TABELA 1 – Tabela de Multiplicidade dos Relacionamentos.....	35
TABELA 2 – Funcionamento <i>Ensemble</i>	60
TABELA 3 – Única Saída Escoamento Bifásico Horizontal e Vertical.....	62
TABELA 4 –Múltiplas Saídas Escoamento Bifásico Horizontal e Vertical.....	62
TABELA 5 – Busca Exaustiva - RNAs.....	65
TABELA 6 – Busca Exaustiva – SVM.....	65
TABELA 7 – Tabela de Resultados Sobre o Conjunto de Teste [0 1].....	74
TABELA 8 – Porcentagem Erro Padrões de Escoamento I.....	75
TABELA 9 – Tabela de Resultados Sobre o Conjunto de Teste [0 1][2 5].....	76
TABELA 10 – Porcentagem Erro Padrões de Escoamento II.....	77
TABELA 11 – Médias de Erros dos Padrões de Escoamento.....	79

LISTA DE ABREVIações

BD	Banco de Dados
BFGS	<i>Broyden-Fletcher-Goldfarb-Shanno</i>
BPM	<i>Back Propagation</i> com termo momentum
CG	<i>Conjugate Gradient</i>
DFP	<i>Davidon-Fletcher-Powell</i>
DSS	<i>Decision Support System</i>
FDP	Função Densidade Probabilidade
FEM	Faculdade de Engenharia Mecânica da Unicamp
FR	Fletcher & Reeves
GRAD	<i>Gradient</i>
GUIs	<i>Graphical User Interfaces</i>
IA	Inteligência Artificial
IDE	<i>Integrated Development Environment</i>
KNN	<i>k-Nearest Neighbor</i>
MR	Modelo Relacional
OSS	<i>One-Step Secant</i>
PDF	<i>Probability Density Function</i>
PR	Polak-Ribière
QN	<i>quasi-Newton</i>
RNA	Redes Neurais Artificiais
SCG	<i>Scaled Conjugate Gradient</i>
SSF	<i>SupportSystemFlow</i>
SSF	<i>SupportSystemFlow</i>
SVM	<i>Support Vector Machines</i>
TCC	Trabalho de Conclusão de Curso

SUMÁRIO

1 Introdução	13
1.1. Justificativa e Motivação	13
1.2. Problematização	14
1.3. Objetivos	14
1.3.1. Gerais	14
1.3.2. Específicos	15
1.4. Organização da Monografia	15
2 Descrição do Problema e Revisão Bibliográfica	17
2.1. Escoamento Vertical e Horizontal	17
2.2. Tratamentos dos Sinais	20
2.3. Revisão Bibliográfica	21
2.3.1. Revisão - Classificação de Padrões	21
2.3.2. Revisão - Sistemas de Suporte à Decisão	22
3 Sistema de Suporte à Decisão	24
3.1. Breve História	25
3.2. Base de Dados	25
3.2.1. Banco de Dados - Sistema SSF	26
3.2.2. Tratamento dos Sinais - SSF	35
3.3. Interface/Fronteira com o Usuário	40
3.3.1. Tela Principal – SSF	41
3.3.2. Acesso ao SSF	42
3.3.3. Dados – SSF	43
3.3.4. Banco de Dados – SSF	46
3.3.5. Componentes Inteligentes – SSF	48
4 Modelos Inteligentes	53
4.1. Redes Neurais Artificiais	53

4.2.	Algoritmo K-Vizinhos.....	56
4.3.	Algoritmo <i>Support Vector Machines</i> (SVM).....	56
4.4.	Ensemble.....	58
5	Metodologia Utilizada	61
5.1.	Classificação de Aquisições	61
5.2.	Busca Exaustiva	64
5.2.1.	Busca Exaustiva - RNAs	64
5.2.2.	Busca Exaustiva - SVM.....	65
5.3.	Modelos Inteligentes	65
5.3.1.	Redes Neurais Artificiais	66
5.3.2.	Algoritmo K-Vizinhos	67
5.3.3.	Algoritmo <i>Support Vector Machines</i> (SVM).....	68
5.4.	<i>Ensemble</i>	69
6	Resultados	71
6.1.	Base de Dados.....	71
6.2.	Componentes Inteligentes e <i>Ensemble</i>	73
6.2.1.	Normalização no intervalo [0 1]	73
6.2.2.	Sinal original - intervalos [0 1] e [2 5] – Caso 1	76
6.2.3.	Sinal original - intervalos [0 1] e [2 5] – Caso 2	77
7	Análises e Conclusões	80
7.1.	Melhores Componentes Inteligentes	80
7.2.	<i>Ensemble</i> x Componentes Inteligentes	81
7.3.	Conclusões	81
7.4.	Trabalhos Futuros	83
	REFERÊNCIAS	84

1

Introdução

Este capítulo apresenta a motivação e os objetivos para realização deste trabalho. No item 1.1. seguem a justificativa e motivação. A problematização é apresentada no item 1.2.. Nos itens 1.3.1. e 1.3.2. são apresentados os objetivos gerais e específicos respectivamente. Finalmente, no item 1.4. é apresentada a organização desta monografia.

1.1. Justificativa e Motivação

O uso dos Sistemas de Suporte à Decisão (DSS) vem sendo cada vez mais amplo devido a grandes quantidades de dados geradas e armazenadas por todos os tipos de setores, a fim de extrair alguma informação importante. Diante deste dilema, a pesquisa com relação ao DSS vem se aprimorando em buscas de novas técnicas e modelos que possam solucionar os diversos problemas de forma eficiente.

O escoamento bifásico vem a ser um fluxo, através de uma tubulação, composto por duas substâncias em fases distintas, homogêneas ou heterogêneas, que não se misturam. Esse regime de escoamento gerar padrões que podem ser classificados de formas distintas. Segundo (Corrêa,2009) *“Escoamentos bifásicos gás-líquido são comumente encontrados na natureza e em diferentes atividades industriais, entre elas na indústria química, de petróleo e de geração de energia elétrica”*. Através destes escoamentos os desafios são: modelar padrões aos mesmos e tomar decisões a partir destes padrões.

Na literatura são encontrados softwares que realizam a classificação do escoamento bifásico tratando o problema na maioria das vezes de forma generalizada. Entretanto, este tipo de setor trabalha com a menor taxa de erro possível, pois uma decisão tomada de forma incorreta pode gerar grandes catástrofes e acidentes como a ocorrida no Golfo do México (G1,2010), por exemplo. Uma das maneiras possíveis de diminuir o risco de cometer erros críticos é através de sistemas especialistas, tendo entre eles o DSS.

Nesta conjuntura, a proposta deste trabalho é apresentar um DSS específico para classificação de padrões do escoamento bifásico vertical e horizontal. Onde o mesmo possua uma boa navegabilidade, interatividade com o usuário, interface amigável, rápido tempo de resposta e outros atributos que o classifique como um bom software. No decorrer desta monografia ele será citado como “*SupportSystemFlow*” ou SSF.

1.2. Problematização

Construir um DSS seguro e confiável para o escoamento bifásico é um grande desafio. Pois além de possuir todos os atributos de um bom Sistema, interface amigável, boa navegabilidade e ser bem estruturado, o mesmo requer um banco de dados robusto que suporte grandes quantidades de dados e possua consultas instantâneas, pois geralmente este tipo de *software* é utilizado em tempo real; Também requerem algoritmos inteligentes capazes de classificar os padrões de escoamento bifásico, apresentando todos estes dados estatísticos com intuito de prover uma fácil compreensão ao usuário final.

Em ambientes reais, este suporte deve ser feito sem interrupções com alta velocidade de resposta e com uma alta taxa de confiabilidade. Segundo (Franco, 2010) “*A dificuldade em capturar e processar uma grande quantidade de dados, além das mudanças contínuas de padrões, faz com que a tomada de decisão seja um processo complexo, principalmente pela quantidade de características que o problema pode apresentar*”.

Portanto a questão a ser feita é: “É possível construir um DSS, integrado com componentes inteligentes, para o escoamento bifásico vertical e horizontal a fim de auxiliar usuários a tomarem suas decisões de forma segura?”.

1.3. Objetivos

1.3.1. Gerais

Este trabalho possui como objetivo geral a proposta de um Sistema de classificação de escoamento bifásico embarcado em um DSS, o Sistema *SupportSystemFlow* (SSF), que possui um combinador de algoritmos inteligentes (*Ensemble*). A proposta do *Ensemble* é fornecer opções seguras, com alto teor de acerto, a possíveis usuários do setor de escoamento.

1.3.2. Específicos

Os objetivos específicos deste trabalho são:

- Criar uma abordagem híbrida para identificação de padrões de escoamento bifásico;
- Encontrar formas viáveis e eficientes na utilização de um combinador de algoritmos (*Ensemble*), com intuito de atingir melhores resultados;
- Compreender o funcionamento completo de um DSS;
- Criar um banco de dados (BD) específico e eficiente para o problema de escoamento bifásico;
- Verificar, através de métodos estatísticos, que a utilização de representações reduzidas dos dados de escoamento tem a mesma validade que todo o conjunto;
- Defender este Trabalho de Conclusão de Curso (TCC), almejando a conclusão da Graduação.

1.4. Organização da Monografia

No próximo capítulo, Capítulo 2, são apresentados detalhes sobre o escoamento bifásico vertical e horizontal, de que forma o escoamento ocorre, informações sobre seus padrões e de que maneira é possível identificá-los. No mesmo Capítulo, se encontra revisão bibliográfica, que correlaciona trabalhos já realizados com o intuito de enriquecer o conhecimento do leitor para comparações posteriores.

O Capítulo 3 aborda sobre Sistemas de Suporte à Decisão ou DSS, onde inicia descrevendo a história dos DSS e sua evolução. Comenta sobre os tipos de DSS e seus limites. E em seguida detalha a organização do Sistema proposto neste trabalho.

Os Modelos Inteligentes são descritos no Capítulo 4. Todos os algoritmos utilizados para a classificação dos padrões de escoamento bifásico são explicados; São eles: Algoritmos de Redes Neurais Artificiais (RNAs), algoritmo de K-Vizinhos e o algoritmo de *Support Vector Machines* (SVM). Ainda no mesmo Capítulo, encontra-se o detalhamento do combinador de algoritmos inteligentes, o *Ensemble*, apresentando os critérios de escolhas e combinação dos algoritmos a fim de atingir resultados ótimos.

Em seguida, no Capítulo 5 é exposta a metodologia proposta para a execução deste trabalho, Classificação de Aquisições, Modelos Inteligentes e *Ensemble*.

Resultados deste trabalho encontram-se no Capítulo 6, onde é apresentada a porcentagem de acerto do modelo, assim como suas principais variáveis estatísticas obtidas durante os testes de validação da base de dados. Também são apresentados objetivos atingidos, como a criação de um DSS, composto de componentes inteligentes, capaz de suprir os requisitos do escoamento bifásico.

Finalmente as análises e conclusões obtidas durante o desenvolvimento e finalização deste trabalho se encontram no Capítulo 7.

2

Descrição do Problema e Revisão Bibliográfica

Segundo (Franco, 2010) “*Escoamento é o fluxo de uma ou mais substâncias com fases distintas, homogêneas ou heterogêneas, através de uma tubulação, sendo que, para cada tipo de sistema, há um conjunto de padrões ou regimes de escoamento que o compõe*”. Vale ressaltar que fases distintas, não estão relacionadas com exclusivamente um tipo de estado para cada substância. Conseqüentemente, duas substâncias podem se apresentar no estado líquido e serem imiscíveis. Existem quatro possibilidades de tipos do escoamento bifásico: o gás-líquido, gás-sólido, líquido-líquido e líquido sólido. Sendo o primeiro citado o mais complexo devido às possibilidades de variações de pressão entre as duas substâncias.

Segundo (Corrêa, 2009) “*As classificações dos padrões de escoamento são concordantes entre diferentes pesquisadores (Hewitt, 1982; Ishii, 1975; Mishima e Ishii, 1984; Taitel et al., 1980; Wallis, 1969), e podem ser escoamento vertical e horizontal*”. O fato de detectar esses padrões é primordial para vários processos industriais que utilizam do escoamento bifásico ou multifásico. Indústrias químicas, petrolíferas e nucleares utilizam destes padrões para configurarem suas redes em tempo real ou/e *online*. Sensores de observação são acoplados em locais estratégicos da rede de escoamento, com o objetivo de controlar os parâmetros de pressão, velocidade de vazão, medição das vazões volumétricas transportadas, determinação da queda de pressão ao longo das linhas de escoamento, gerenciamento da produção, que evitam acidentes e viabilizam o escoamento.

2.1. Escoamento Vertical e Horizontal

Para o escoamento bifásico vertical, foi adotada a classificação de padrões presente na dissertação de mestrado de (Corrêa, 2009) dividindo-se em três grupos:

- Bolhas
 - Bolhas
 - Capa esférica (*spherical cap*)
- Intermitente
 - Golfadas
 - Agitado
- Anular
 - Semi-Anular
 - Anular

Respectivamente Anular, Semi-Anular, Agitado, Golfadas, Capa Esférica e Bolhas, Figura 1.

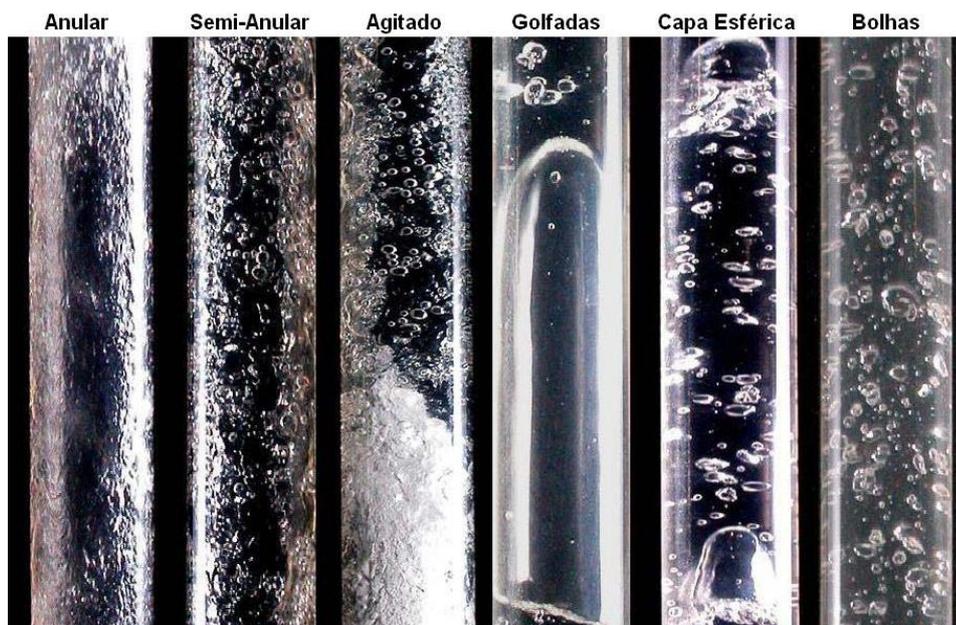


Figura 1 - Regimes de escoamento gás/líquido vertical.

Fonte: (Corrêa, 2009).

“O subpadrão Anular é caracterizado por um núcleo com gás e gotas de líquido em alta velocidade envolto por um filme de líquido co-corrente que escoia junto à parede”.

O Semi-Anular *“é considerado uma forma degenerada do sub-padrão anular com grandes ondas interfaciais onde o filme de líquido adjacente as parede pode apresentar mudança de direção”.*

O Agitado “*instável, como o nome sugere. O líquido apresenta um movimento oscilatório ou agitado*”.

Já o subpadrão Golfadas “*é caracterizado por uma sucessão de pistões de líquido aerados seguidos por bolhas de gás alongadas, frequentemente denominadas por bolhas de Taylor, escoando verticalmente ao redor de um filme de líquido junto à parede*”.

O subpadrão Capa Esférica ou *Spherical Cap* “*é caracterizado por bolhas que podem tomar parcialmente ou completamente a seção transversal do tubo na forma de uma "capa" ou "tampa" de gás*”.

Finalmente o subpadrão Bolhas “*representa distribuições com bolhas monodispersas com tamanho uniforme e também cluster de bolhas frequentemente denominadas por bolhas discretas*”.

Para o escoamento vertical, além destes seis padrões existem dois padrões que são os limites; Os padrões Cheio ou *Full* e Vazio ou *Empty*, Figura 2.

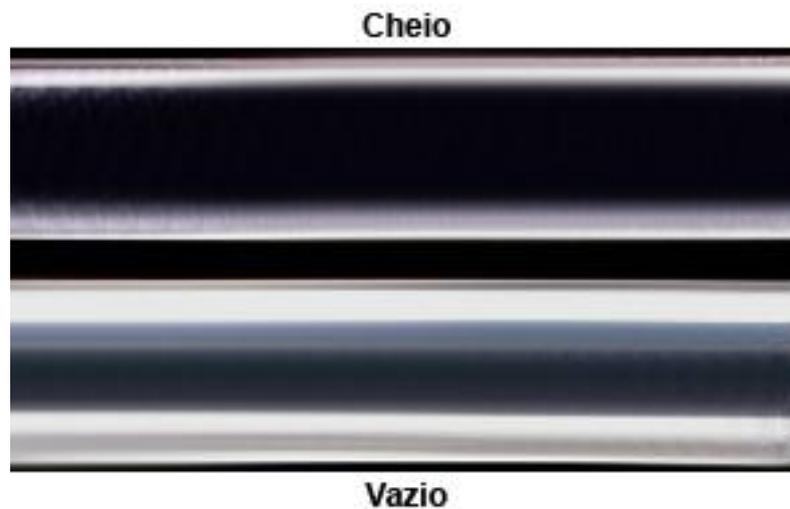


Figura 2 – Padrões Cheio e Vazio.

O padrão Cheio ou *Full* significa que o fluxo de escoamento se encontra completo da substância líquida, para o escoamento bifásico gás/líquido que é o tratado neste trabalho.

Para o padrão Vazio ou *Empty* significa que o fluxo de escoamento se encontra cheio de substância gasosa ou vazio de substância líquida como sugere o nome do padrão.

Para o escoamento bifásico horizontal a classificação de padrões adotada foi da seguinte forma:

- Estratificado;

- Pistonado;
- Bolhas Dispersas;
- Disperso;
- Ondular;
- Pistonado Aerado;
- Anular;



Figura 3 – Padrões de Escoamento Horizontal gás/líquido.

Fonte: (Corrêa, 2009).

Os padrões, Estratificado e Ondular, são observados com baixas vazões, ou seja, velocidades do gás e do líquido mais baixas.

O padrão Aerado é observado no aumento da vazão do gás, quando se formam bolhas maiores na parte superior do tubo.

Nas altas velocidades de vazão, podem-se observar os padrões: Disperso, Bolhas Dispersas e Anular.

Assim como no escoamento bifásico vertical, o horizontal também possui os padrões, de limite, Cheio ou *Full* e Vazio ou *Empty*, com as mesmas configurações.

2.2. Tratamentos dos Sinais

Padrões de escoamento podem ser armazenados de diversas formas. Após as aquisições é preciso verificar se não ocorreu nenhuma interferência externa que afetasse os dados, pois isso acarreta em alterações de padrões deixando os dados inválidos. “Os sinais adquiridos, tanto de pressão, como de impedância, não são diretamente utilizados pelas metodologias desenvolvidas. Esses sinais variam bastante devido ao arranjo espacial das fases de gás e líquido, de modo que se torna difícil identificar o padrão diretamente a partir deste sinal” (CORRÊA, 2009).

Frente a isto, existem vários tratamentos de sinais que podem ser encontrados na literatura:

- PDF (Função de Densidade de Probabilidade), (Mi et al., 1996), (Tsoukalas; Ishii; Mi, 1997), (Matsui, 1984) e (Tutu, 1982);
- PSDF (Função de Distribuição de Probabilidade), (Lee et al., 2004);
- FFT (Transformada Rápida de Fourier), (Yamashita, 1997);
- Média e Desvio Padrão, (Mi; Ishii; Tsoukalas, 1998);
- Transformada de Gabor, (Selli; Selegim, 2004);
- Transformada de Hilbert Huang, (Bin et al., 2006) e (Ding et al., 2007);
- Redução de Dimensionalidade (Amostragem Sistemática), (Bolfarine; Bussab, 2005), (Cochran, 1977) e (Thompson, 1992).

Vale ressaltar que existem vantagens e desvantagens para cada um dos processos citados. As características variam de um para o outro, tendo entre elas o custo computacional e a confiança na representação, além de outras observações e variações específicas de cada tratamento. A combinação dessas características que determinam a eficiência de cada processo.

O tratamento utilizado neste trabalho foi o de redução de dimensionalidade e é descrito no subitem de Capítulo 3.3.2..

2.3. Revisão Bibliográfica

A revisão bibliográfica deste trabalho é dividida em duas partes. Primeiramente, é apresentada a revisão da classificação de padrões, tema foco deste trabalho, em conjunto com os componentes inteligentes. A segunda parte da revisão bibliográfica apresenta os DSS de suporte ao escoamento bifásico.

2.3.1. Revisão - Classificação de Padrões

A classificação de escoamento bifásico pode ser encontrada na literatura em dois tipos de regimes principais, vertical e horizontal.

No trabalho de (Dailiang et al., 2004) utilizaram a lógica fuzzy, atingindo taxas de acertos de 85,2% a 96,4%, na proposta de classificação em tempo real.

As propostas de classificação de padrões mais encontradas são com a utilização de RNAs, conseqüente de sua fácil utilização.

Utilizando valores estatísticos da variação de tensão como entradas para as RNAs (Mi; Ishii; Tsoukalas, 2001) conseguiram, em tempo real, identificar o escoamento.

(Cai et al., 1994) utilizou a abordagem de Kohonen, (Lee et al., 2004) além desta abordagem optaram por acrescentar a abordagem de RNAs *Perceptron* de Múltiplas Camadas. Ainda com a utilização de RNAs *Perceptron* de Múltiplas Camadas (MLP) (Rosa et al., 2010) atingiram resultados de 92% de acertos em suas classificações.

Com a utilização do algoritmo *Backpropagation* (BP) (Wu; Zhou; Wu, 2001) atingiram acertos de 92% na identificação de padrões.

Com a combinação da Lógica Fuzzy e RNAs (Mi et al., 1996) atingiram até 100% de taxas de acertos em suas classificações.

Através de um Sistema Especialista (Corrêa, 2009) atingiu classificações com taxas de 70% a 100% de acerto.

Com técnicas de agrupamento para criar conjuntos de padrões de maneira otimizada (Salgado et al., 2010) obtiveram taxas de acertos entre 70% a 91% também utilizando um Sistema Especialista.

2.3.2. Revisão - Sistemas de Suporte à Decisão

DSS são encontrados na petrolífera com grande abrangência de mercado. Sistemas que possuem robustez e tempo curto de resposta são os mais indicados, porém estes quesitos geralmente são degradados em sistemas muito abrangentes.

O *Gas Flow Rate Calculation* da EUCALYPT Systems é um software de suporte a vazão de gás que possui fácil interface gráfica para a entrada de parâmetros, suporte a usuários para reutilização de dados e configurações pré-definidas para novos modelos.

A QUORUM oferece um software, o *PGAS Flow Measurement*, que serve como um repositório central contendo dados de medição volumétrica e analítica, incluindo automação de testes sobre gás natural.

As Ferramentas da SUNRISE apresentam análises de fluxo de fluído. Possuem módulos para cálculos de fluxos instáveis, sistemas de proteção contra incêndio e um módulo para resolver problemas de fluxo em geral.

Nota-se que estas ferramentas apresentadas e outras que ainda podem ser encontradas nesta área, são geralmente de atuação abrangente. Porém, nenhuma das ferramentas encontradas trata especificamente sobre reconhecimento de escoamento bifásico. A proposta deste trabalho é suprir esta carência com o SSF, que apresenta características específicas para esta área.

3

Sistema de Suporte à Decisão

Neste Capítulo é apresentada uma breve história sobre os Sistemas de Suporte à Decisão (DSS) e toda a descrição do Sistema *SupportSystemFlow* (SSF) proposto por este trabalho com o objetivo de dar suporte as modelos inteligentes.

No item 3.1. será apresentada uma introdução sobre os DSS. No segundo item deste Capítulo, item 3.2., são descritos: a estrutura e importância da base de dados para um DSS. O último item deste Capítulo, item 3.3., expõe o conceito das interfaces do SSF que são componentes fundamentais para a composição de um Sistema.

Um Sistema de Suporte à Decisão é um conjunto de elementos combinados que tem o objetivo de oferecer suporte na tomada de decisões em um problema específico. Geralmente são construídos para solucionar problemas que possuem uma dificuldade elevada e dados indisponíveis (variáveis não conhecidas), característico de problemas não estruturados.

São compostos por três elementos principais: simulação, previsão e otimização. A simulação é responsável por encontrar todas as opções possíveis de decisão. A previsão seleciona as melhores decisões, com base em pontuações de execuções anteriores presente na base de dados. A otimização analisa as decisões, tomadas e executadas, e as pontua positivamente no caso de acerto ou negativamente no caso de erro.

Para a construção de um Sistema, de qualquer natureza, a mínima estrutura necessária sempre terá que ser composta, de alguma forma, de:

- Algoritmos;
- Base de Dados;
- Interface;

Onde os algoritmos são responsáveis pelo dinamismo e funcionamento do Sistema. A base de dados pelo fornecimento de dados aos algoritmos e persistência dos mesmos tratados. E finalmente a interface, que possibilita a interação do Sistema com o usuário através de dispositivos de entrada e saída.

A descrição dos Modelos Inteligentes, que praticamente compõem os algoritmos, é apresentada no Capítulo 4.

3.1. Breve História

Segundo (Bispo, 1998), *“Desde o início da civilização, o homem sempre procurou algo que lhe auxiliasse no seu processo decisório. Ele buscava nas divindades esse auxílio para que pudesse tomar decisões consideradas sábias”*. Com o início da era tecnológica o homem criou ferramentas que auxiliassem suas decisões, a fim de diminuir os erros. Nesta época, por volta de 1965, os auxílios dos sistemas eram apenas para cálculos matemáticos. *“No final dos anos 1960, um novo tipo de sistema de informação tornou-se prática - modelo orientado à DSS ou sistemas de decisão de gestão”* (POWER, 2003). Após este período os DSS evoluíram junto com os avanços da tecnologia; Chegando até os sistemas de tomadas de decisões “autônomas”, podendo citar alguns exemplos como:

- Sistema de pilotagem automática de aeronaves;
- Linhas de produções automobilísticas;
- Sistemas inteligentes de irrigação de plantações;
- E vários outros sistemas principalmente no setor de automação.

3.2. Base de Dados

A base de dados deve ser robusta e bem projetada, a fim de suportar grandes quantidades de dados.

Com o avanço da tecnologia, os dispositivos de observação de dados vêm se aprimorando. Decorrentes disto, grandes quantidades de dados estão sendo armazenadas. Este armazenamento não é viabilizado apenas para observações registradas como histórico de um trabalho. Mas se viabiliza pela possibilidade de estudos futuros, descrevendo modelos que de alguma forma venham contribuir gerando economias ou/e lucros.

A seguir, será apresentada a descrição do BD utilizado para armazenar as aquisições de escoamento bifásico e as entradas dos modelos inteligentes utilizados neste trabalho.

3.2.1. Banco de Dados - Sistema SSF

O Banco de Dados (BD) que será descrito, neste item, foi desenvolvido com o objetivo de compor o Sistema *SupportSystemFlow* (SSF) que, através do mesmo, controla as visualizações de dados e exportação de configurações de padrões. O BD deve gerenciar quaisquer manipulações referentes aos dados coletados nas estações de controle de padrões. Deve oferecer uma total segurança aos seus usuários, através de controle de contas, permissões e visões.

A seguir é apresentado, Figura 4, o Modelo Relacional (MR) do SSF precedido de sua descrição.

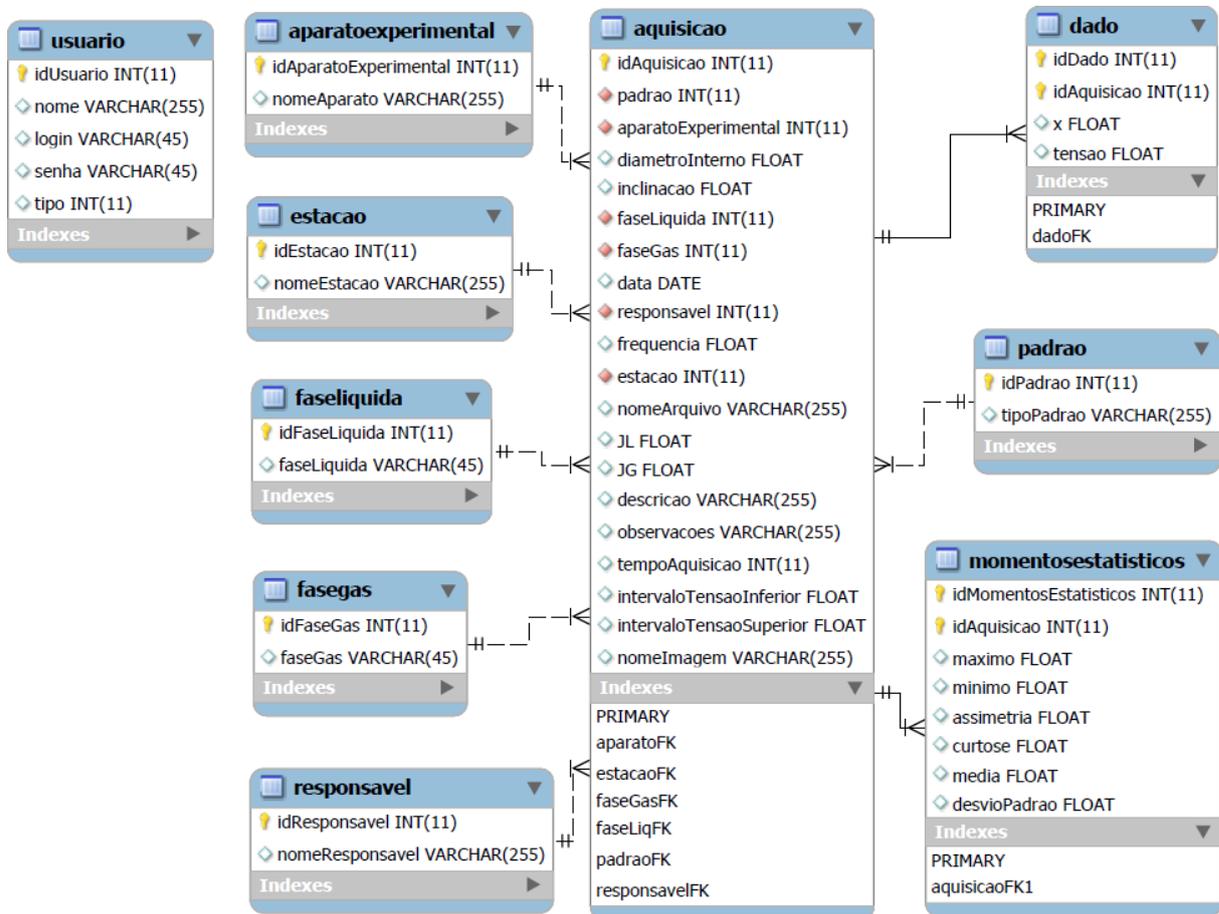


Figura 4 – Modelo Relacional do Sistema *SupportSystemFlow*.

Normalizações são utilizadas para minimizar as redundâncias e as anomalias que podem ser encontradas em um BD. Seguindo este pretexto, a MR do SSF, Figura 4, garante as três primeiras formas normais de BD encontradas na literatura.

“Uma forma normal é uma regra que deve ser obedecida por uma tabela para que esta seja considerada “bem projetada”. Há diversas formas normais, isto é, diversas regras, cada vez mais rígidas, para verificar tabelas relacionais” (HEUSER, 2001). No caso deste trabalho, vamos considerar três formas normais.

Como o problema modelado trata-se de aquisições de dados em uma rede de escoamento bifásico. Estas aquisições identificam padrões de escoamento e se relacionam com:

- Estação;
- Aparato experimental;
- Padrão identificado;
- Responsável;
- Fase líquida;
- Fase gás;
- Dado;
- Momentos estatísticos.

As multiplicidades de todos os relacionamentos serão descritas no fim deste item. Na sequência são descritos todas as relações desprezando o atributo identificador, a chave primária (*Primary Key*).

A partir da relação "aquisicao", Figura 4, é possível restaurar todos os dados presentes no momento do escoamento. Observa-se a existência de 19 atributos que caracterizam a estrutura. A seguir é apresentada a descrição de cada um dos atributos:

- padrao → Identifica o padrão da aquisição;
- aparatoExperimental → Identifica o aparato que foi utilizado no momento da aquisição;
- diametroInterno → Armazena o diâmetro da tubulação em que foi feita a aquisição;
- inclinacao → Armazena a inclinação da tubulação em que foi feita a aquisição;
- faseLiquida → Identifica a fase líquida no momento da aquisição;
- faseGas → Identifica a fase gás no momento da aquisição;
- data → Armazena a data em que foi feita a aquisição;

- `responsavel` → Identifica o responsável pela aquisição;
- `frequencia` → Armazena a frequência, em Hertz, que foram feitas as coletas de dados pelo aparato experimental;
- `estacao` → Identifica a estação que a aquisição foi feita;
- `nomeArquivo` → Representa o nome do arquivo gerado na aquisição, o que facilita um backup de segurança;
- `JL` → Armazena a velocidade o líquido no momento da aquisição;
- `JG` → Armazena a velocidade do gás no momento da aquisição;
- `descricao` → Armazena a descrição da aquisição, informada pelo responsável;
- `observacoes` → Armazena as observações importantes notadas pelo responsável no momento da aquisição;
- `tempoAquisicao` → Registra o tempo que foi necessário para realizar a aquisição de dados em segundos;
- `intervaloTensaoInferior` → Armazena o menor resultado de tensão encontrado durante o escoamento;
- `intervaloTensaoSuperior` → Armazena o maior resultado de tensão encontrado durante o escoamento;
- `nomeImagem` → Registra o nome da imagem tirada no momento da aquisição.

As chaves estrangeiras (*Foreign Keys*) contidas nesta relação, Figura 5, compõem os relacionamentos que serão descritos abaixo:

- `aparatoFK`: o atributo “`aparatoExperimental`” referencia a relação “`aparatoExperimental`”. Essa referencia possui multiplicidade de 1 para n. Isso quer dizer que cada aquisição pode possuir/referenciar apenas um aparato experimental e um aparato experimental pode estar presente em várias aquisições;
- `estacaoFK`: o atributo “`estacao`” referencia a relação “`estacao`”. Possui multiplicidade de 1 para n. Sendo que cada aquisição pode possuir/referenciar apenas uma estação e uma estação pode estar presente em várias aquisições;
- `faseGasFK`: o atributo “`faseGas`” referencia a relação “`faseGas`”. Possui multiplicidade de 1 para n. Sendo que cada aquisição pode possuir/referenciar apenas uma fase gás e uma fase gás pode estar presente em várias aquisições;

- faseLiquidaFK: o atributo “faseLiquida” referencia a relação “faseLiquida”. Possui multiplicidade de 1 para n. Sendo que cada aquisição pode possuir/referenciar apenas uma fase líquida e uma fase líquida pode estar presente em várias aquisições;
- padraoFK: o atributo “padrao” referencia a relação “padrao”. Possui multiplicidade de 1 para n. Sendo que cada aquisição pode possuir/referenciar apenas um padrão e um padrão pode estar presente em várias aquisições;
- responsavelFK: o atributo “responsavel” referencia a relação “responsavel”. Possui multiplicidade de 1 para n. Sendo que cada aquisição pode ser registrada por apenas um responsável e um responsável pode estar presente em várias aquisições.

aquisicao	
idAquisicao	INT(11)
padrao	INT(11)
aparatoExperimental	INT(11)
diametroInterno	FLOAT
inclinacao	FLOAT
faseLiquida	INT(11)
faseGas	INT(11)
data	DATE
responsavel	INT(11)
frequencia	FLOAT
estacao	INT(11)
nomeArquivo	VARCHAR(255)
JL	FLOAT
JG	FLOAT
descricao	VARCHAR(255)
observacoes	VARCHAR(255)
tempoAquisicao	INT(11)
intervaloTensaoInferior	FLOAT
intervaloTensaoSuperior	FLOAT
nomeImagem	VARCHAR(255)
Indexes	
PRIMARY	
aparatoFK	
estacaoFK	
faseGasFK	
faseLiqFK	
padraoFK	
responsavelFK	

Figura 5 – Relação Aquisição.

A Figura 6 apresenta a modelagem da relação “padrao”. Observa-se a existência de apenas um atributo que caracteriza sua estrutura. Através desta relação é possível recuperar o tipo de padrão identificado no momento da aquisição.

O atributo é descrito abaixo:

- tipoPadrao → Representa o tipo de padrão encontrado no momento da aquisição. Os padrões mais comuns são: Anular, Semi-Anular, Agitado, Golfadas, Capa Esférica e Bolhas respectivamente conforme a Figura 1 para o escoamento bifásico vertical, e Estratificado, Ondular, Pistonado, Pistonado Aerado, Bolhas Dispersas, Anular e Disperso conforme a Figura 2.

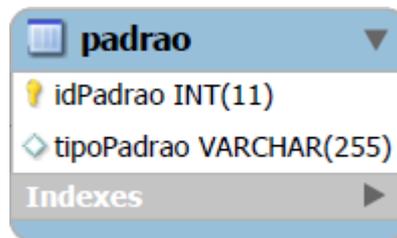


Figura 6 - Relação Padrão.

A modelagem da relação “aparatoExperimental” é apresentada na Figura 7. Também se observa a existência de apenas um atributo que caracteriza sua estrutura. Através desta relação é possível recuperar o tipo de aparato que foi utilizado para adquirir as tensões da aquisição.

O atributo é descrito abaixo:

- nomeAparato → Armazena o nome do aparato que foi utilizado para adquirir as tensões da aquisição. O aparato pode variar por modelos de sensores.

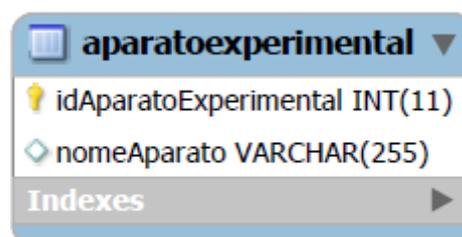


Figura 7 - Relação Aparato Experimental.

A Figura 8 expõe a modelagem da relação “faseLiquida”. Nota-se que a mesma é uma relação de apenas um atributo. Através desta relação é possível recuperar o tipo de líquido identificado no momento do escoamento bifásico.

O atributo é descrito abaixo:

- faseLiquida → Armazena o nome da fase líquida do escoamento bifásico. Exemplos de fase líquida podem ser água e petróleo.

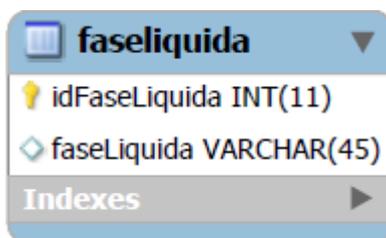


Figura 8 - Relação Fase Líquida.

A apresentação da modelagem da relação “faseGas” é feita na Figura 9. Observa-se a existência de apenas um atributo que caracteriza sua estrutura. Através desta relação é possível recuperar o tipo de gás identificado no momento do escoamento bifásico.

O atributo é descrito abaixo:

- faseGas → Armazena o nome da fase gás do escoamento bifásico. Exemplos da fase gás podem ser ar e metano.

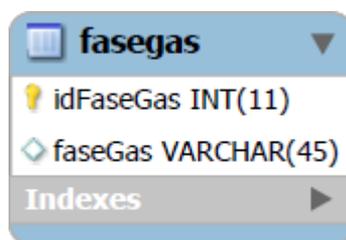


Figura 9 - Relação Fase Gás.

A Figura 10 apresenta a modelagem da relação “responsavel”. Ela também é uma relação de apenas um atributo. Através desta relação é possível recuperar o nome do responsável pela aquisição.

O atributo é descrito abaixo:

- nomeResponsavel → Armazena o nome do responsável pela realização da aquisição no escoamento bifásico.

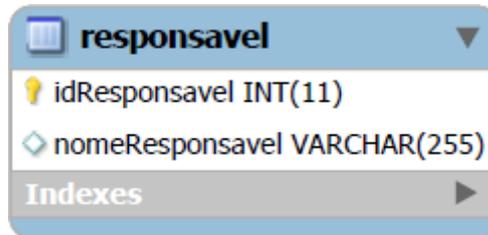


Figura 10 - Relação Responsável.

Na Figura 11 é descrita a relação “estacao”. Observa-se a existência de apenas um atributo que caracteriza sua estrutura. Através desta relação é possível recuperar o nome da estação em que foi realizada a aquisição.

O atributo é descrito abaixo:

- nomeEstacao → Armazena o nome da estação em que são realizadas as aquisições.

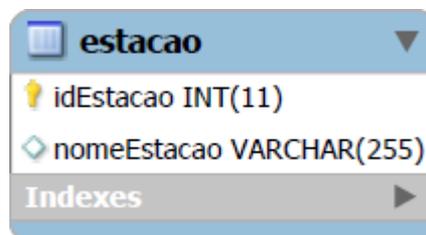


Figura 11 - Relação Estação.

A Figura 12 apresenta a modelagem da relação “dado”. Observa-se a existência de dois atributos e uma chave estrangeira que caracterizam sua estrutura. Através desta relação é possível recuperar as tensões de uma determinada aquisição.

Os atributos são descritos abaixo:

- idAquisicao → Através deste atributo é possível saber qual aquisição determinado dado pertence;
- x → Armazena a ordem em que o dado foi adquirido pelo aparato experimental. Exemplo: em uma aquisição de 180 mil pontos o atributo x varia de 1 a 180.000;
- tensao → Armazena a tensão adquirida pelo aparato experimental. De acordo com as tensões que são possíveis identificar os padrões.

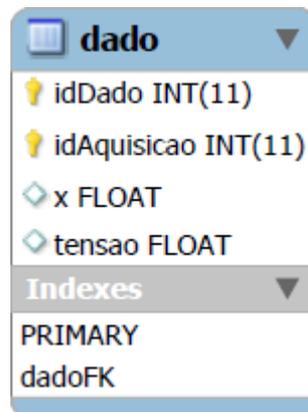


Figura 12 - Relação Dado.

A chave estrangeira (*Foreign Key*) contida nesta relação, Figura 12, compõe o relacionamento que será descrito abaixo:

- dadoFK: o atributo “idAquisicao” referencia a relação “aquisicao”. Possui multiplicidade de n para 1. Sendo que cada dado pode estar presente/referenciar apenas uma aquisição e uma aquisição pode possuir vários dados.

A relação “momentosEstatisticos” é exposta na Figura 13. Sua estrutura é composta por sete atributos. Através desta relação é possível recuperar os dados estatísticos de uma determinada aquisição.

Os atributos são descritos abaixo:

- idAquisicao → Através deste atributo é possível saber qual aquisição os momentos estatísticos pertencem;
- maximo → Armazena o máximo das tensões adquiridas na aquisição;
- minimo → Armazena a mínima tensão adquirida na aquisição;
- assimetria → Armazena a assimetria das tensões adquiridas na aquisição;
- curtose → Armazena a curtose das tensões adquiridas na aquisição;
- media → Armazena a média das tensões adquiridas na aquisição;
- desvioPadrao → Armazena o desvio padrão das tensões adquiridas na aquisição.

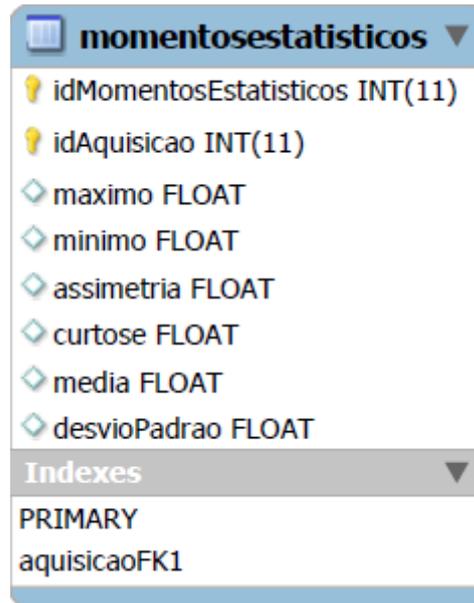


Figura 13 - Relação Momentos Estatísticos.

A chave estrangeira (*Foreign Key*) contida nesta relação, Figura 13, compõe o relacionamento que será descrito abaixo:

- aquisicaoFK1: o atributo “idAquisicao” referencia a relação “aquisicao”. Possui multiplicidade de 1 para 1. Sendo que cada “momentoestatisticos” pode referenciar apenas uma “aquisicao” e uma “aquisicao” pode possuir apenas uma referencia à “momentosestatisticos”.

A Figura 14 apresenta a modelagem da relação “usuario”. Observa-se a existência de quatro atributos que caracterizam sua estrutura. Através desta relação é possível recuperar todas as informações do usuário do sistema.

Os atributos são descritos abaixo:

- nome → Armazena o nome do usuário do sistema;
- login → Armazena o *login* do usuário que por sua vez deve ser único;
- senha → Armazena a senha do usuário;
- tipo → Armazena o tipo de usuário, ou seja, a permissão que o usuário possui sobre o sistema, administrador ou comum.

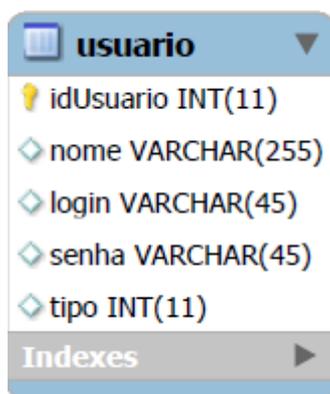


Figura 14 - Relação Usuário.

Conforme já apresentado, o BD do Sistema *SupportSystemFlow* possui dez relações, entre elas nove se relacionam, Figura 4. A Tabela 1 exemplifica todas as multiplicidades dos relacionamentos citados anteriormente. A ausência da relação "usuario" é justificada pela mesma não possuir referência com as demais.

Tabela 1 – Tabela de Multiplicidade dos Relacionamentos.

	Aquisição	Fase líq.	Fase gás	Dado	Padrão	Responsável	Estação	Ap. Exp.	Mom. Est.
Aquisição	-	1	1	N	1	1	1	1	1
Fase líq.	N	-	-	-	-	-	-	-	-
Fase gás	N	-	-	-	-	-	-	-	-
Dado	1	-	-	-	-	-	-	-	-
Padrão	N	-	-	-	-	-	-	-	-
Responsável	N	-	-	-	-	-	-	-	-
Estação	N	-	-	-	-	-	-	-	-
Ap. Exp.	N	-	-	-	-	-	-	-	-
Mom. Est.	1	-	-	-	-	-	-	-	-

Na Tabela 1, Mom. Est. significa momentos estatísticos, Ap. Exp. significa aparato experimental e Fase líq. significa fase líquida.

3.2.2. Tratamento dos Sinais - SSF

Todos os dados adquiridos por sensores de escoamento são armazenados em arquivos, estruturados de maneira simples para facilitar e agilizar a aquisição dos mesmos. Porém não é instintivo inseri-los no BD, ainda necessitam ser “tratados”.

O *SupportSystemFlow* oferece um módulo de redução de dados que garante a

eficiência e a qualidade das reduções. Sabe-se que todo sinal deste trabalho é obtido por meio de uma sonda capacitiva, um aparato experimental, instalado em ponto específico do duto de escoamento. A Figura 15 mostra o sinal medido pela sonda adquirido a uma frequência de 3KHz durante 60s, gerando um total de 180000 pontos que representam o sinal.

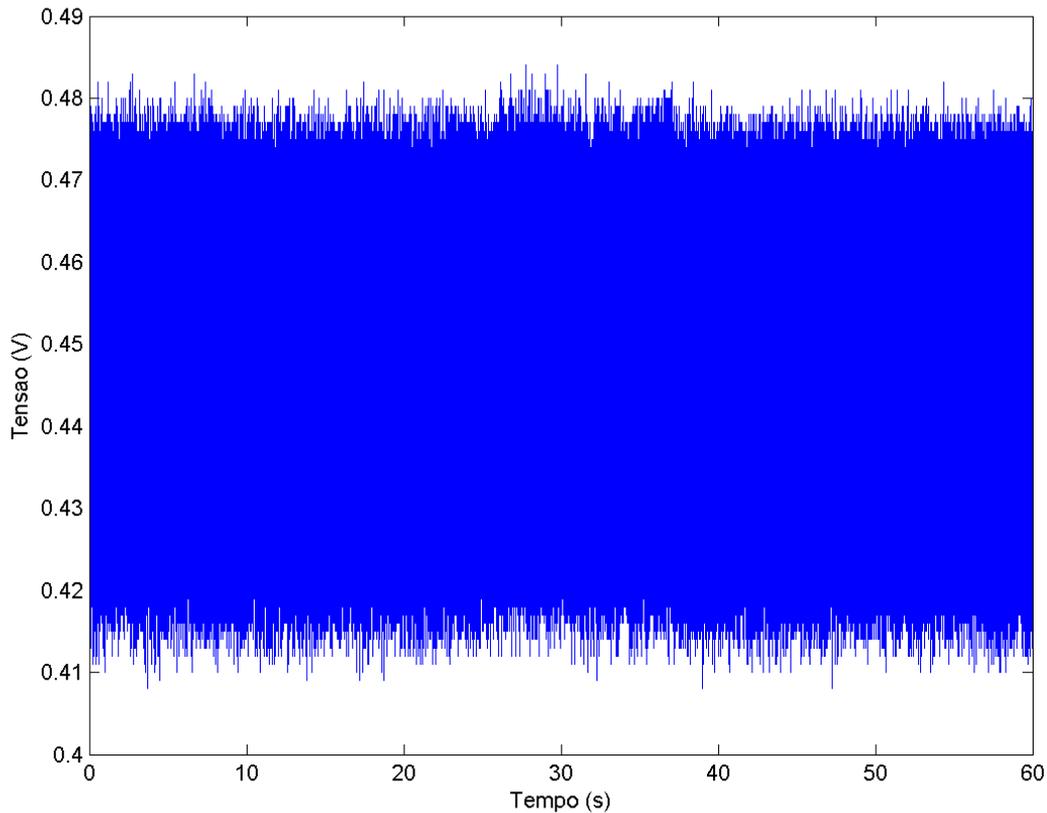


Figura 15 – Sinal Original Escoamento Bifásico.

Para obter informações relevantes destes dados é necessário aplicar funções matemáticas, porém estas funções podem consumir um alto tempo de processamento inviabilizando a utilização do sistema em tempo real (*online*).

Uma alternativa para este problema é realizar o procedimento estatístico de redução de dimensionalidade denominado amostragem sistemática discutida e detalhado por (Bolfarine; Bussab, 2005), (Cochran, 1977) e (Thompson, 1992). Este procedimento visa reduzir a dimensão do conjunto original mantendo características originais inalteradas.

A seguir serão descritos os passos para a realização da amostragem sistemática:

- Define-se inicialmente o número de pontos (R) que irão compor o subconjunto de dimensão reduzida;
- Deriva-se o intervalo de coleta de pontos através da divisão do número de pontos (N) do conjunto original pelo número de pontos (R) do conjunto reduzido, obtendo o intervalo (k);
- Finalmente, a partir do ponto inicial, do conjunto original, retira-se um elemento a cada k intervalo para o conjunto reduzido.

As Figuras 16 e 17 ilustram o procedimento de redução de dimensionalidade aplicado nos padrões de escoamento bifásico.

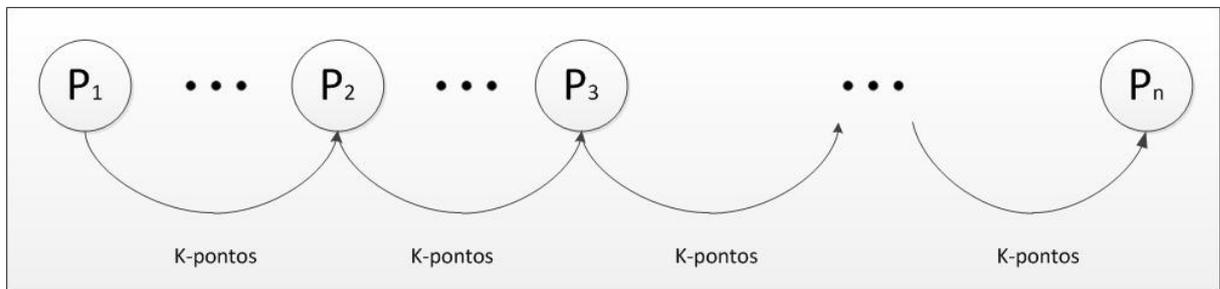


Figura 16 – Redução de dimensionalidade I.



Figura 17 – Redução de dimensionalidade II.

A Figura 18 apresenta o mesmo sinal mostrado na Figura 15 com 99% reduzida, ou seja, apenas 1800 pontos do conjunto original. Desta forma utilizou-se um intervalo de k igual a 100, o que significa uma coleta a cada 100 pontos.

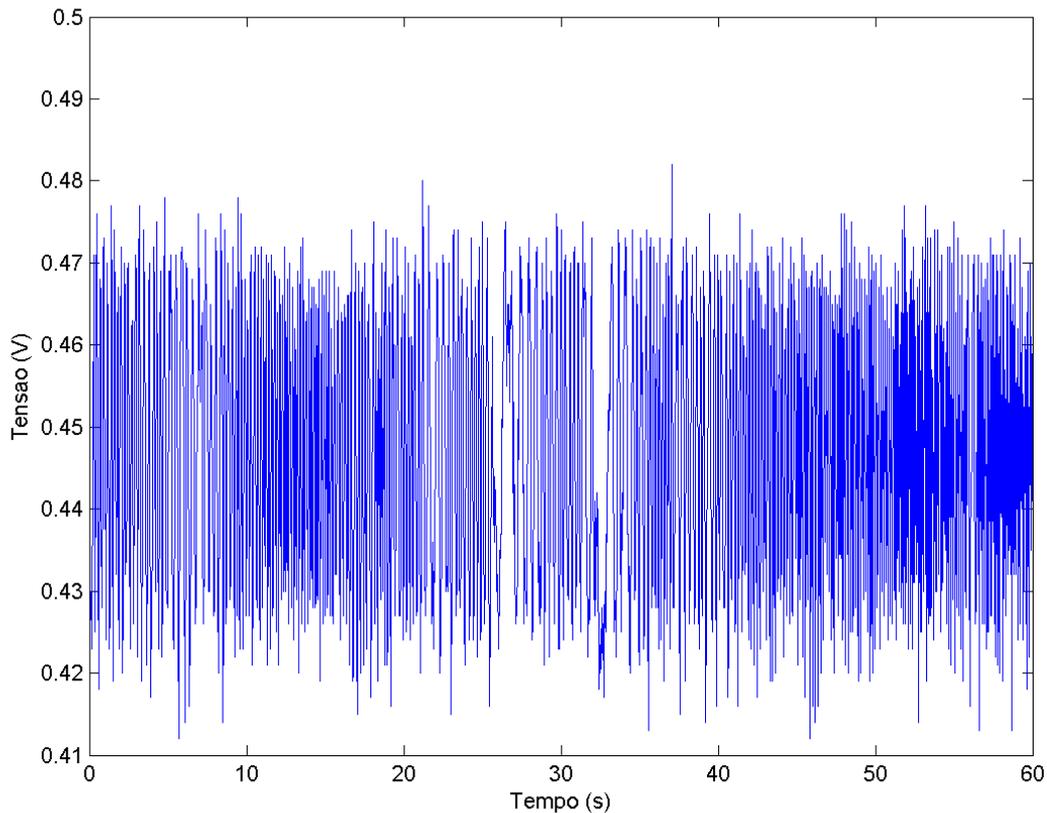


Figura 18 – Sinal Reduzido Escoamento Bifásico.

Para obter-se um melhor desempenho, foram geradas reduções a partir da base de dados completa. Estas reduções reproduzem com a devida precisão suas bases nos momentos estatísticos (média, desvio padrão, assimetria e curtos), não tendo perdas na confiabilidade dos testes. Isso possibilitou um grande ganho na velocidade de consulta, ou seja, no desempenho. Inicialmente, era necessário consultar cento e oitenta mil tuplas para poder gerar um gráfico de linhas e com trinta, vinte e dez por cento dos dados os gráficos possuíram o mesmo comportamento.

Após a redução, é necessário validar a representatividade do subconjunto. Esta validação é calculada através da função densidade probabilidade (*Probability Density Function* - PDFs) para ambos os conjuntos, original e reduzido.

A Figura 19 mostra as PDFs geradas a partir dos conjuntos. Nota-se que ambas possuem similaridades em seus comportamentos, mostrando que é possível utilizar um sinal de menor dimensão com a mesma capacidade de representação.

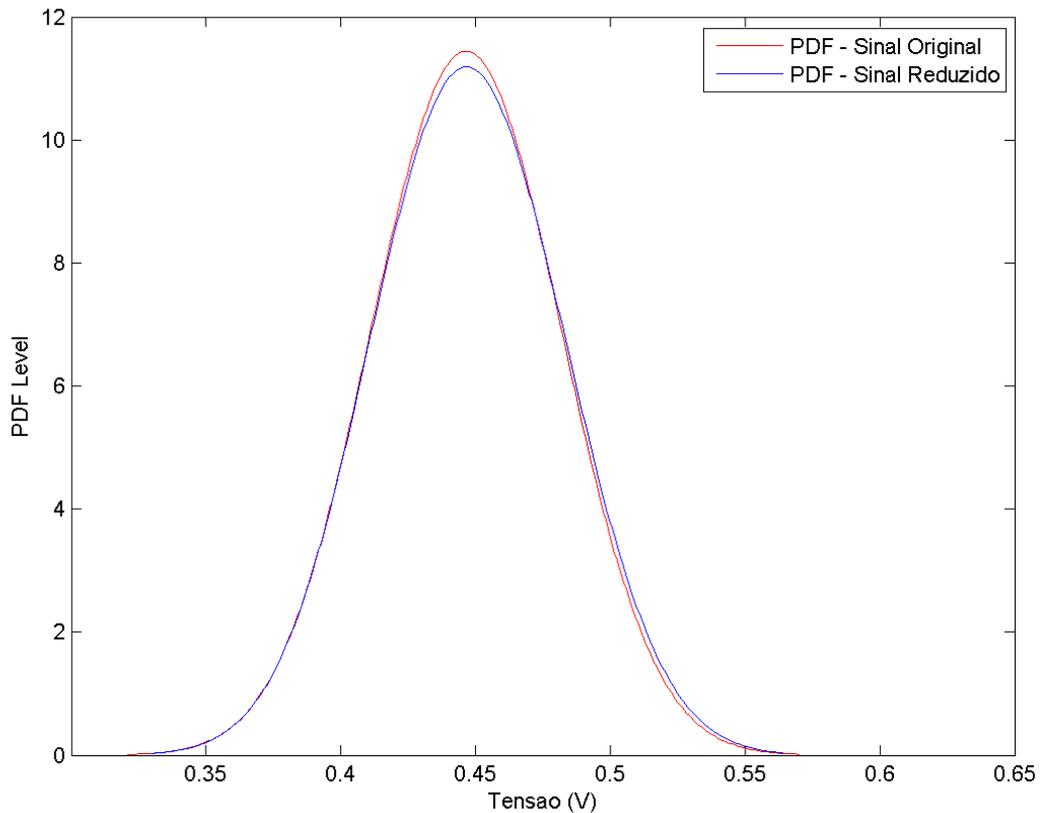


Figura 19 – PDFs dos Sinais Original e Reduzido.

Os momentos estatísticos das aquisições foram escolhidos para serem utilizados como entrada nos modelos inteligentes, tanto na fase de treinamento quanto na fase de classificação. Ao utilizar os momentos estatísticos os modelos criados tendem a replicar o comportamento do padrão observado. Neste trabalho foram utilizadas quatro variáveis estatísticas, sendo elas:

- Média;
- Desvio Padrão;
- Assimetria;
- Curtose.

A escolha destes quatro momentos é justificada pelo fato de apenas estes quatro momentos é possível representar cada padrão, não necessitando de mais momentos que também representariam, porém não apresentaria mudanças na representação. No trabalho realizado por (Rosa; Salgado; Ohishi; Mastelari, 2010) isso é comprovado.

3.3. Interface/Fronteira com o Usuário

Segundo (Salgado, 2009) “*Geralmente, a análise de resultados é um procedimento artesanal, que demanda um grande esforço por parte do analista, que muitas vezes necessita de utilizar a interpretação dos resultados gerados pelos modelos*”. Frente a isto, o Sistema *SupportSystemFlow* possui componentes gráficos a fim de auxiliarem a leitura dos dados apresentados aos usuários. “*Quando o projeto de interface gráfica é bem definido, o ambiente de controle do software se torna compreensível, agradável e intuitivo, dando aos usuários uma sensação de satisfação e aumento de interesse pelo sistema*” (SALGADO, 2009).

A interface do *SupportSystemFlow* foi projetada com o princípio de orientação a objetos. A construção da mesma foi baseada no modelo de interface presente na tese de doutorado de (Salgado, 2009), Figura 20.

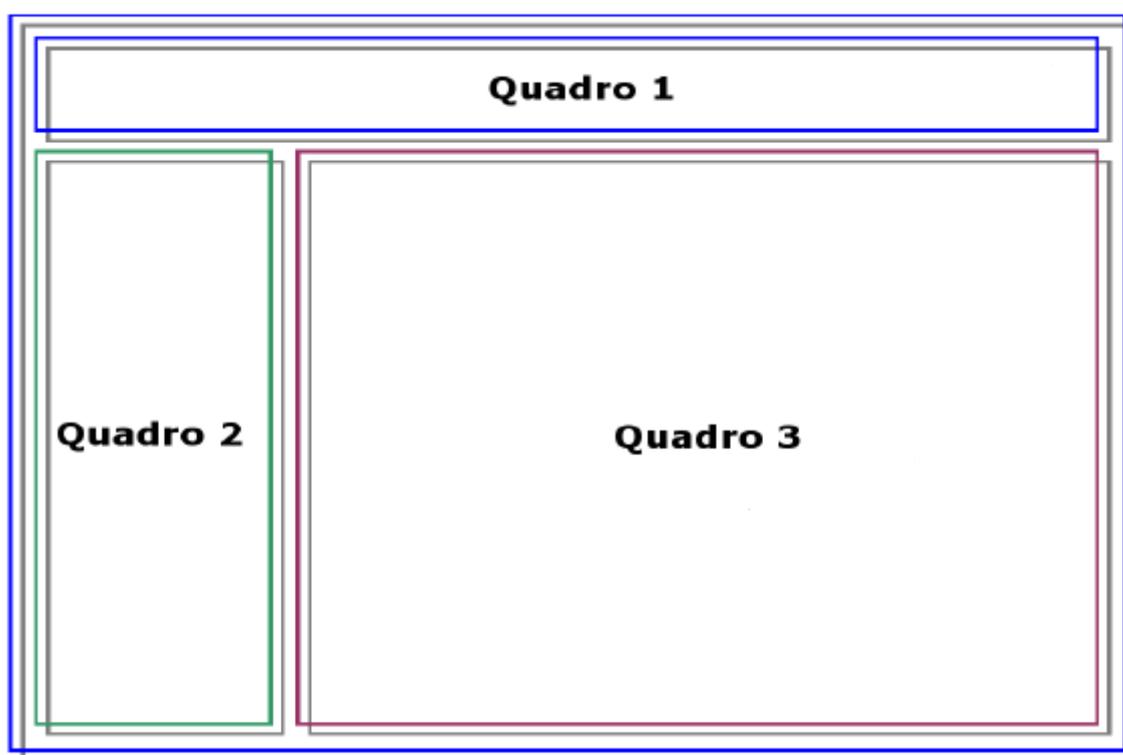


Figura 20 - Layout da Interface Gráfica com Usuário.

Fonte: (Cicogna, 2003).

No Quadro 1, Figura 20, localizam-se a barra de ferramentas e os menus de acesso aos componentes do Sistema. As listas de conteúdos, informações tratadas pelo Sistema, são apresentadas no Quadro 2. Finalmente, no Quadro 3, quadro central, são apresentados detalhes sobre os itens presentes no segundo quadro.

Os próximos itens apresentam interfaces gráficas dos componentes do *SupportSystemFlow*, que viabilizam a interatividade com o usuário de forma intuitiva.

3.3.1. Tela Principal – SSF

A tela principal, como já descreve em seu nome, é o principal componente gráfico do Sistema. É através dela que o usuário deve ter acesso a todas as outras funcionalidades. A mesma foi baseada no modelo proposto na Figura 20.

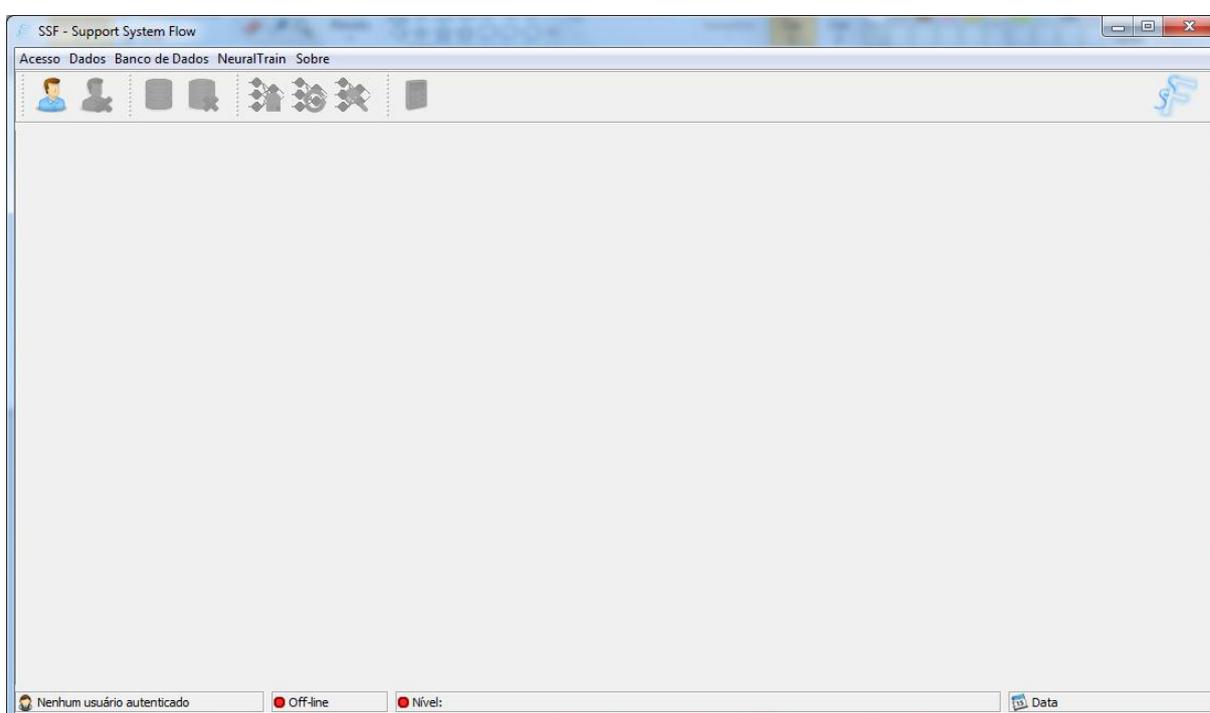


Figura 21 – Tela Principal *SupportSystemFlow*.

Na Figura 21 é apresentada a tela principal, primeira visão após a inicialização do Sistema SSF, Figura 22.



Figura 22 – Inicialização *SupportSystemFlow*.

A Figura 22 tem o intuito de chamar a atenção do usuário, alertando a ele que as funcionalidades do Sistema estão prestes a serem apresentadas.

3.3.2. Acesso ao SSF

A segurança de controle de acesso foi um item bem projetado para o SSF. Através do controle de usuários, é possível permitir visões e acessos a componentes diferentes. No SSF existem duas categorias de usuários. A primeira é a categoria de Administrador do Sistema, que possui acesso total. A segunda é o Usuário Comum, onde o referido possui acesso limitado, sendo proibido de funções como deletar o BD, controlar usuários entre outras.

Na Figura 23 é possível notar que o acesso é controlado por um *Login* e uma senha. Onde o SSF conta com um gerenciador de acessos através desta autenticação requerida.

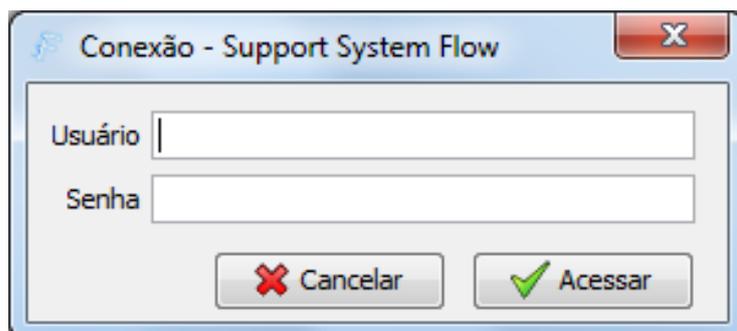


Figura 23 – Tela Login *SupportSystemFlow*.

Após a autenticação do usuário são liberados os componentes da Tela Principal, Figura 24.

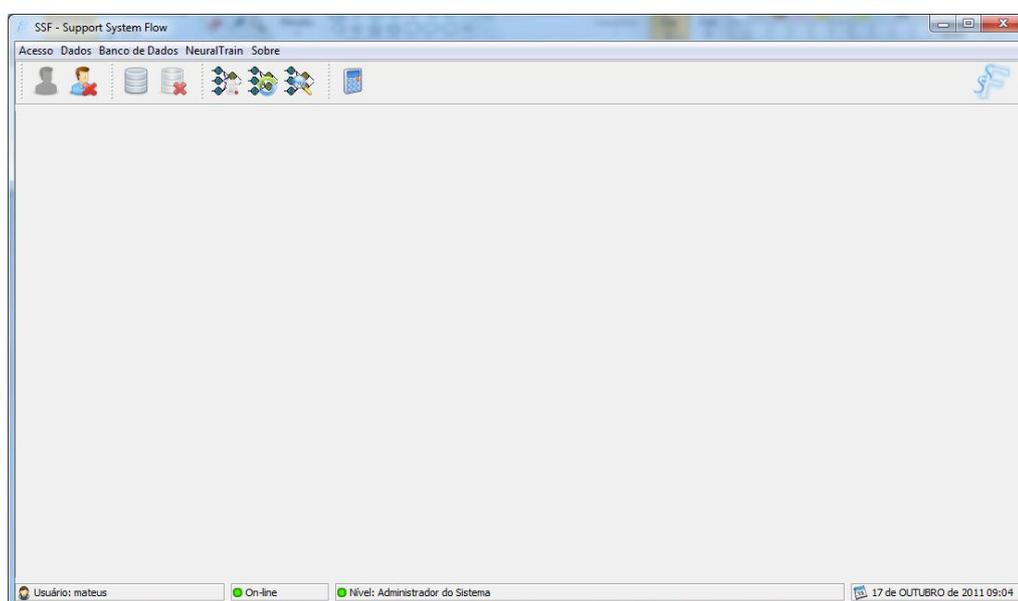


Figura 24 – Tela Principal Liberada *SupportSystemFlow*.

3.3.3. Dados – SSF

Para o funcionamento de um Sistema sempre serão necessários dados ou informações. Esta afirmação é importante, pois ressalta que sem os mesmos o Sistema não tem sentido. A justificativa de um Sistema sempre está relacionada com dados a serem tratados. Frente a isto, neste trabalho, foram feitas pesquisas para encontrar maneiras de apresentar grandes quantias de dados de forma intuitiva e simples. A Figura 25 apresenta um exemplo de apresentação dos dados na Tela Principal.

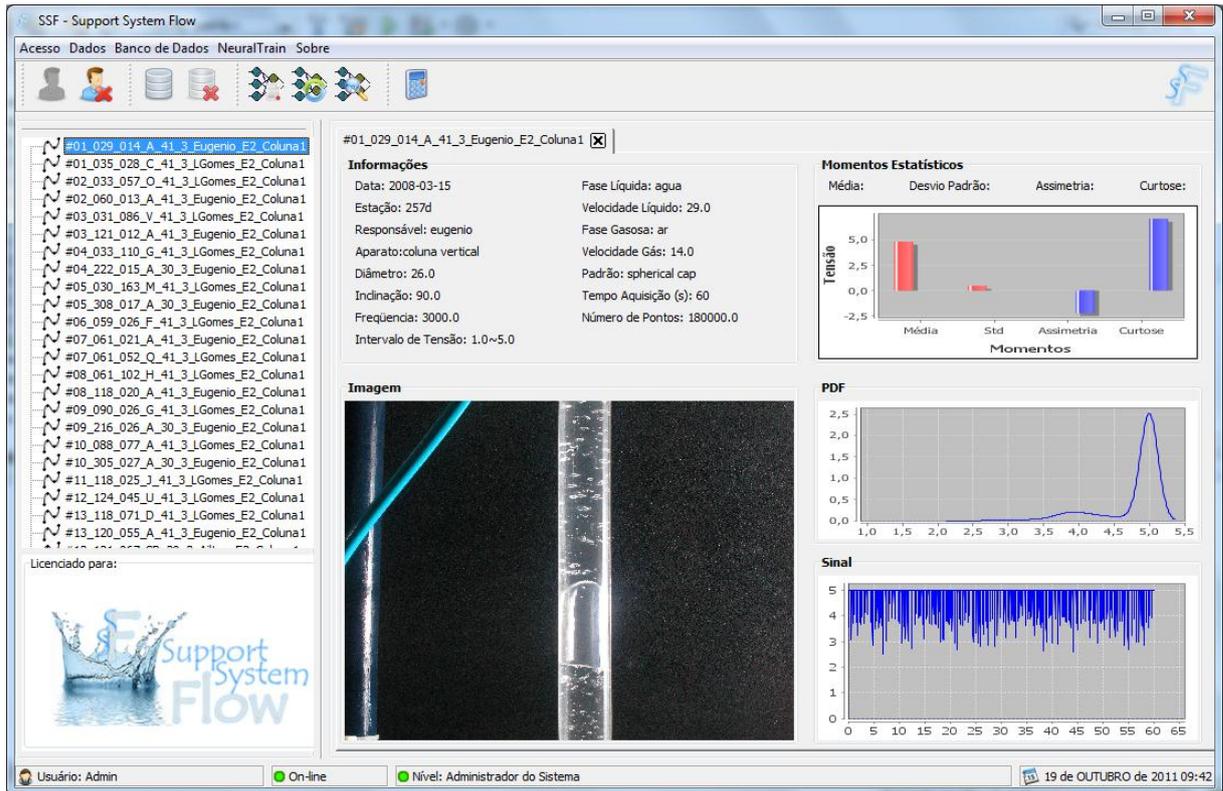


Figura 25 – Apresentação dos Dados *SupportSystemFlow*.

Não somente a manipulação dos dados é importante, quanto também a importação dos mesmos. Pois através da importação é possível compor a base de dados de um SSD. Para o Sistema SSF foi construído um componente de importação que oferece vários modos de importação dos dados. A partir deste componente, Figura 26, o usuário manipula o BD de maneira imperceptível.

Support System Flow - Importador Aquisições

Importador: Aquisições

Marcar Todos
 Limpar

#01_029_014_A_41_3_Eugen
 #01_035_028_C_41_3_LGome
 #02_033_057_O_41_3_LGome
 #02_060_013_A_41_3_Eugen
 #03_031_086_V_41_3_LGome
 #03_121_012_A_41_3_Eugen
 #04_033_110_G_41_3_LGome
 #04_222_015_A_30_3_Eugen
 #05_030_163_M_41_3_LGome
 #05_308_017_A_30_3_Eugen
 #06_059_026_F_41_3_LGome
 #07_061_021_A_41_3_Eugen
 #07_061_052_Q_41_3_LGome
 #08_061_102_H_41_3_LGome
 #08_118_020_A_41_3_Eugen
 #09_090_026_G_41_3_LGome
 #09_216_026_A_30_3_Eugen
 #10_088_077_A_41_3_LGome
 #10_305_027_A_30_3_Eugen
 #11_118_025_J_41_3_LGome
 #12_124_045_U_41_3_LGome
 #13_118_071_D_41_3_LGome
 #13_120_055_A_41_3_Eugen
 #13_121_067_SB_30_3_Aitor
 #14_029_023_C_41_3_LGome

Para seleccionar mais de um item pressione o Ctrl.

Nome Arquivo: #02_060_C Fase Líquida: Agua
 Data: 2008-03-18 Velocidade Líquido: (m/s) 060
 Estação: 257D Fase Gasosa: AR
 Responsável: Eugenio Velocidade Gás: (m/s) 013
 Aparato: Coluna ver Padrão: Spherical C
 Diâmetro: 26 Tempo Aquisição (s): 60
 Inclinação: 90 Número de Pontos: 180000
 Frequência: 3000 Intervalo de Tensão: 1~5
 Série: A
 Descrição: Dados utilizados nos experimen
 Observações: Dados coletados por Aiton (OUT

Redefinir Nivel Tensão
 Mínimo: 0 Máximo: 5
 Todas as Colunas
 Colunas: 2-3;7-9;13
 ex: 1-4; 5-10; 13

Figura 26 – Componente de Importação *SupportSystemFlow*.

A Figura 27 apresenta o processo de importação de dados em andamento.

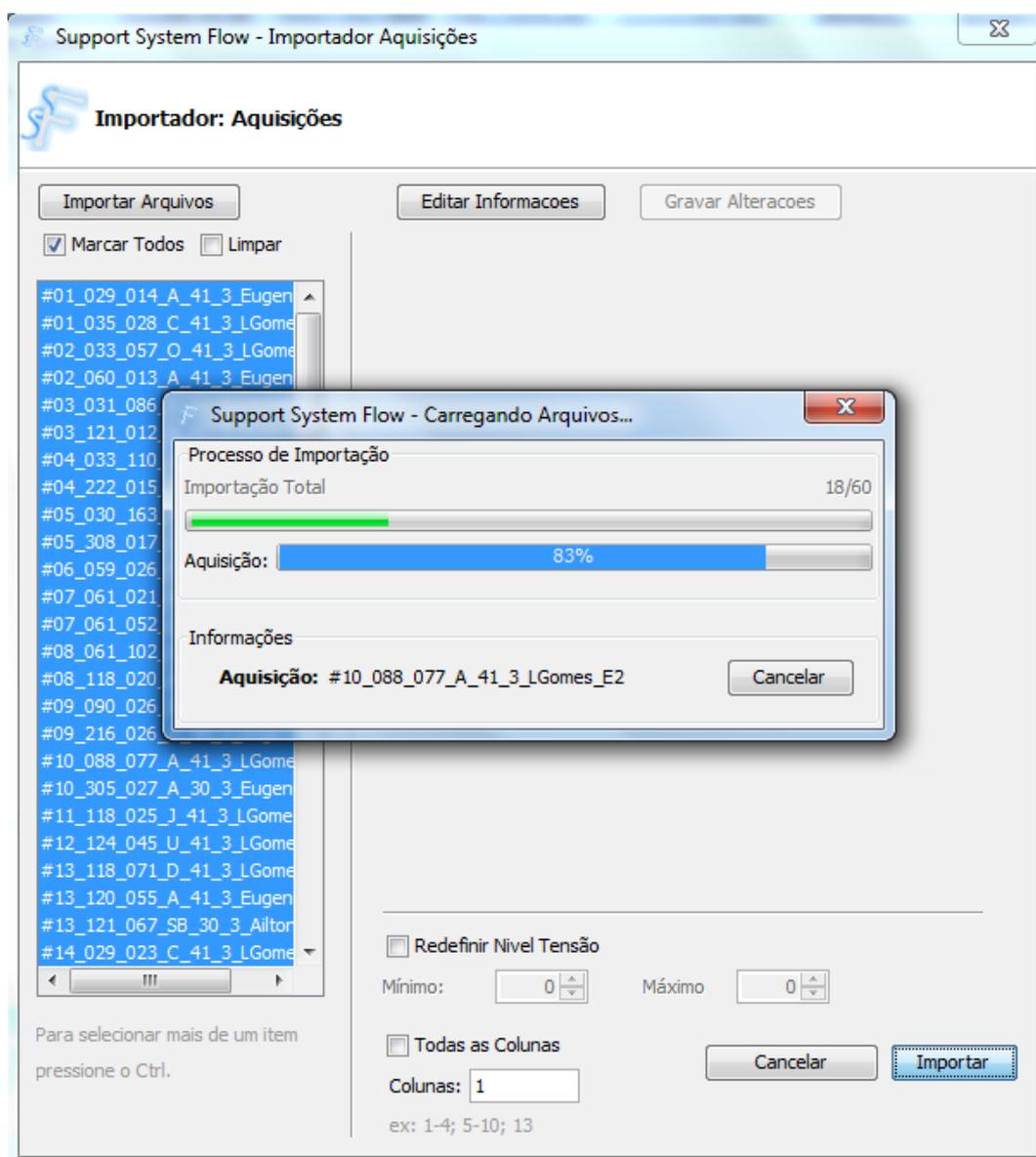


Figura 27 – Processo de Importação de Dados *SupportSystemFlow*.

3.3.4. Banco de Dados – SSF

O BD, como já foi citado anteriormente, para um SSD deve ser robusto e eficiente. Não deve apresentar lentidão, mesmo tendo de suportar grandes quantias de dados, e oferecer um controle rigoroso de usuários com o objetivo de assegurar a consistência dos dados e a proteção contra roubos de informações. A manipulação do BD é um requisito que nem sempre é encontrado nos DSSs. Com limitações que este requisito apresenta, foram projetados para o Sistema SSF componentes que possibilitam essa manipulação, tanto de conexão, criação e exclusão do BD. A Figura 28 apresenta o módulo de configuração de conexão do BD do SSF.

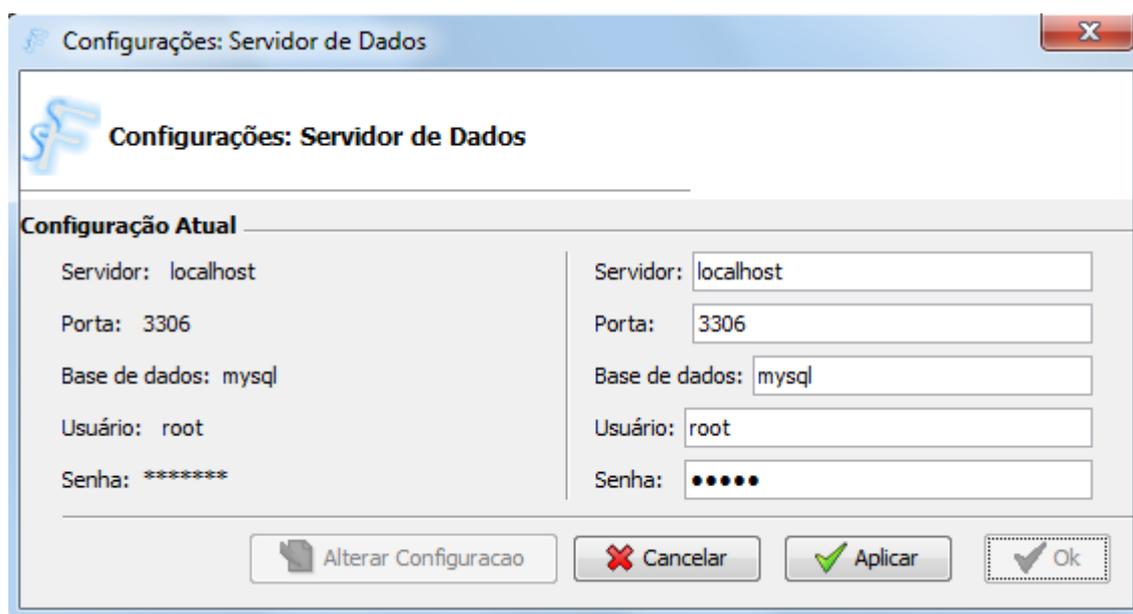


Figura 28 – Módulo de configuração de Conexão de BD *SupportSystemFlow*.

O SSF também dispõe de um módulo que possibilita a criação e a exclusão do BD, Figura 29. Antes da exclusão, o usuário é indagado se realmente deseja executar tal operação.

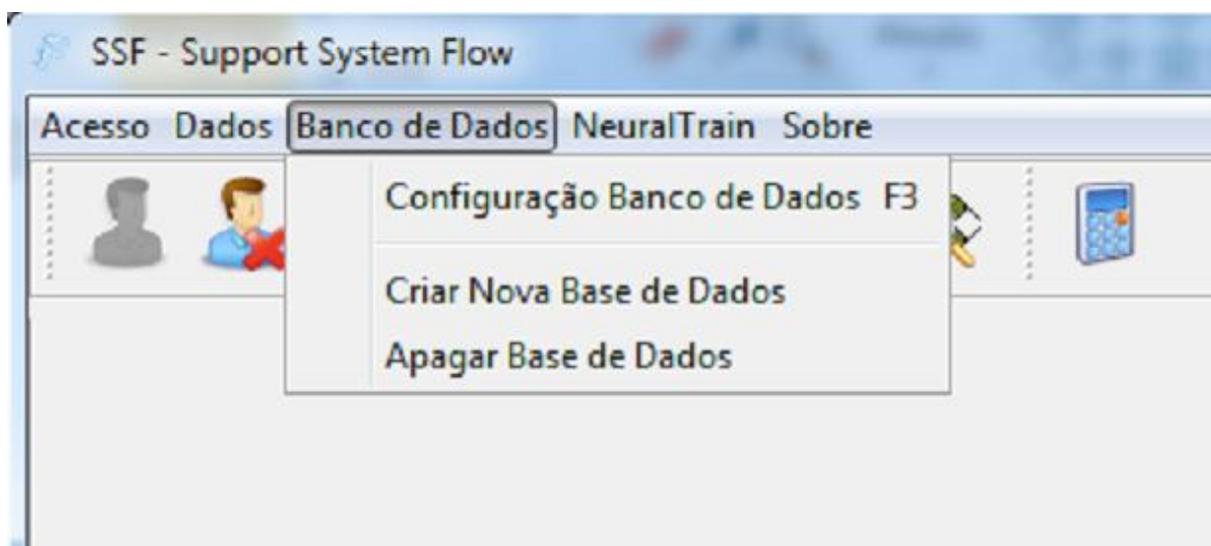


Figura 29 – Menu de Gerenciamento do BD *SupportSystemFlow*.

Vale ressaltar que a opção de apagar o BD do Sistema, Figura 29, só é possível através da categoria de usuário administrador. Conseqüentemente, se o usuário estiver “logado” como usuário comum estas funções não estão disponíveis.

3.3.5. Componentes Inteligentes – SSF

Os componentes inteligentes compõem o módulo principal do SSF. Sem os mesmos as tomadas de decisões são inativas. Um modelo inteligente pode funcionar sem um sistema de telas de interfaces, porém um DSS só deixa de ser um conjunto de telas e interfaces a partir de sua junção com modelos inteligentes robustos e “calibrados” para o problema a ser tratado.

Módulos inteligentes foram construídos para auxiliarem os usuários a tomarem suas decisões com base em análises de gráficos, classificação de padrões e momentos estatísticos. Para estas análises, o usuário pode configurar um novo estudo, Figura 30 e 31, podendo compor os conjuntos de treinamento, validação e teste manualmente, Figura 32, ou de forma aleatória, Figura 33.

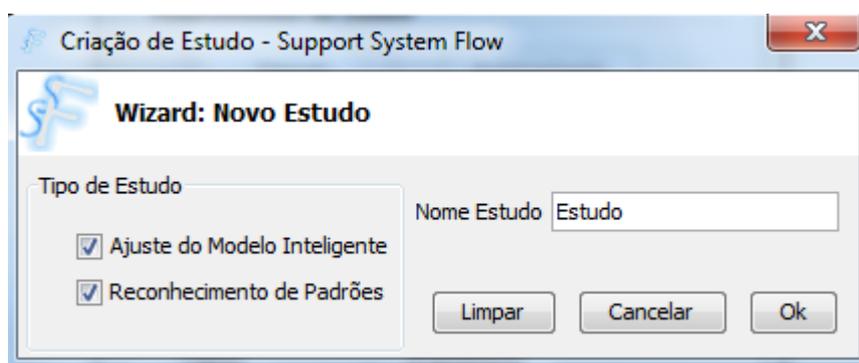


Figura 30 – Nomeando Estudo Inteligente – *SupportSystemFlow*.

Vale ressaltar que para os componentes apresentados a seguir apresenta-se o campo “*termo momentum*”, porém o mesmo não se apresenta em todos os componentes inteligentes utilizados neste trabalho, apenas presente nos métodos *Back Propagation* com *termo momentum* (BPM) e Método *Gradient* (GRAD), que serão apresentados no próximo Capítulo. Para os componentes que não possuem o termo momentum, este campo é ocultado em todos os módulos do Sistema.

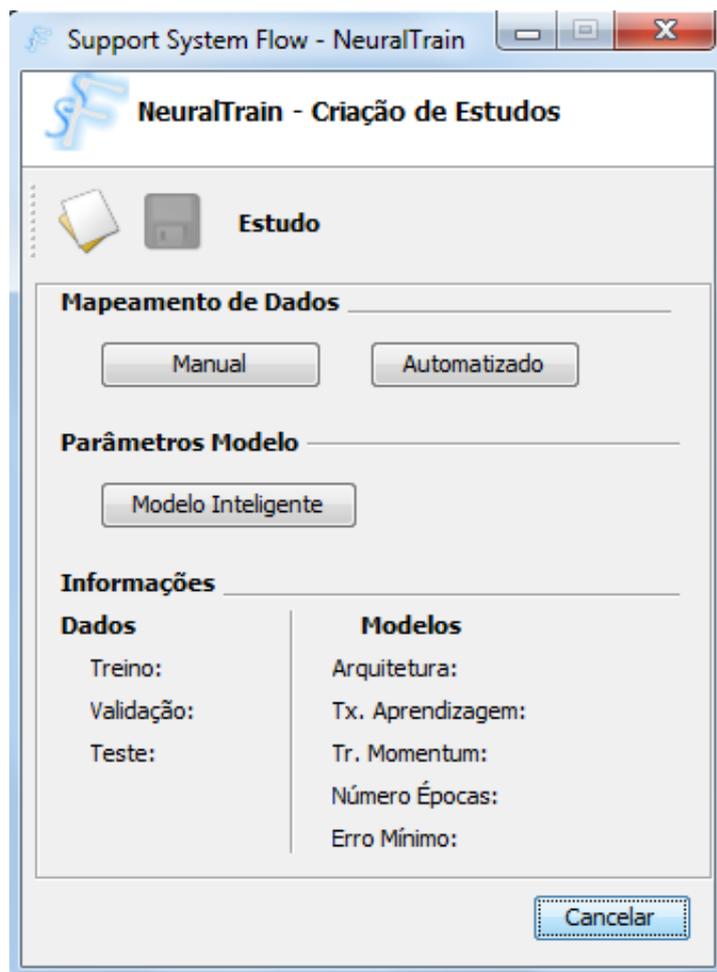


Figura 31 – Módulo Criação Estudo Inteligente – *SupportSystemFlow*.

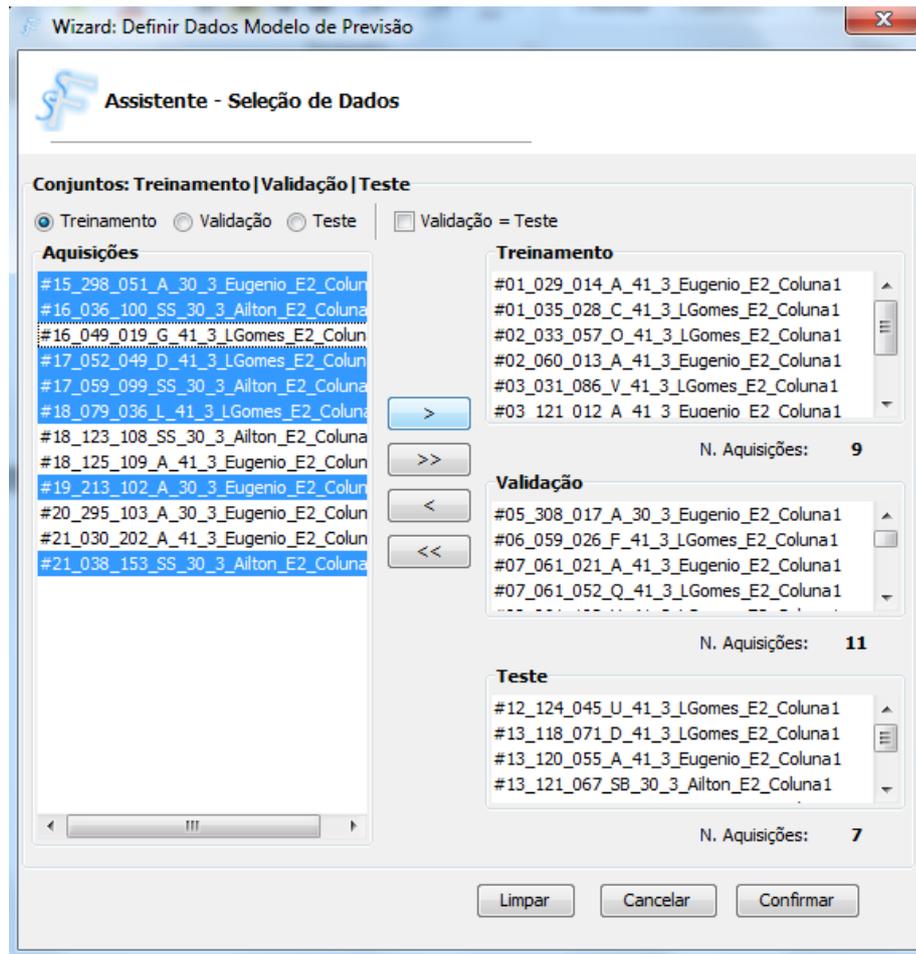


Figura 32 – Configuração Conjuntos Manualmente – *SupportSystemFlow*.

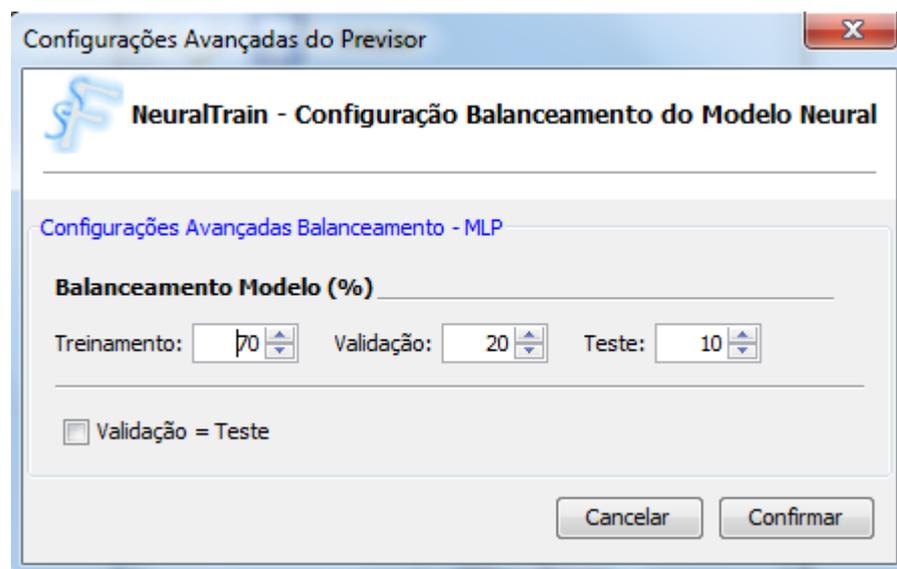


Figura 33 – Configuração Conjuntos Aleatoriamente – *SupportSystemFlow*.

Nota-se que na Figura 32 e 33 é apresentada a opção do conjunto de validação ser igual ao conjunto de teste. Isso se justifica a utilização quando o conjunto total em questão for

relativamente pequeno, ou seja, apresente poucos padrões não suportando compor os dois conjuntos.

Além de criar um novo estudo, o usuário do SSF tem a opção de treinar um estudo criado anteriormente, Figura 34. Também é apresentada a opção de normalização das entradas e saídas, com o objetivo de encontrar melhores resultados.

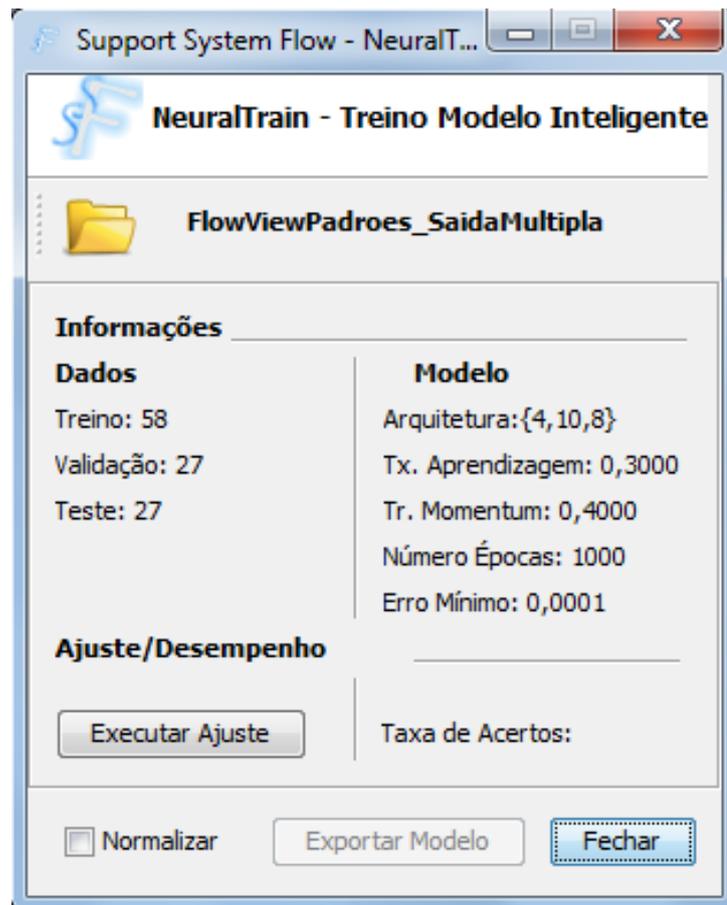


Figura 34 – Treinamento Estudo – *SupportSystemFlow*.

Após executar o ajuste de um estudo é possível exportar o mesmo com o intuito de executar classificações posteriores, Figura 35.

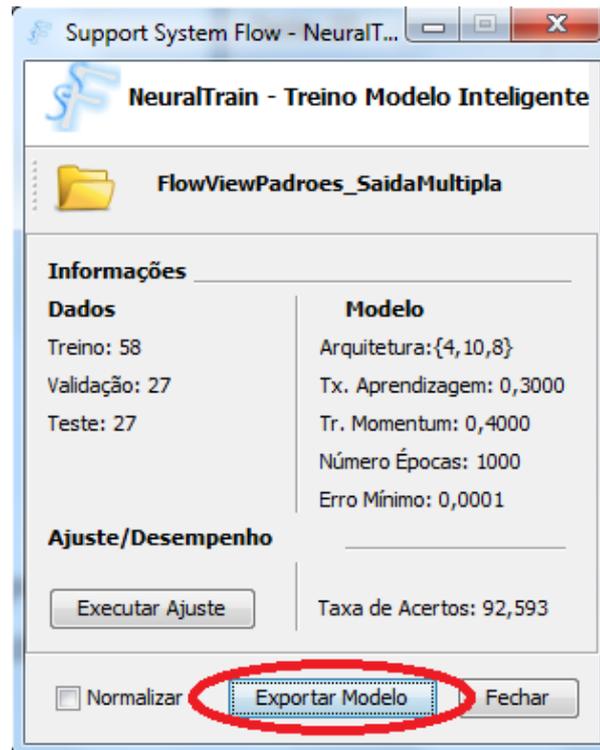


Figura 35 – Exportando Treinamento – *SupportSystemFlow*.

Finalmente, o último componente inteligente a ser apresentado é o de busca exaustiva, Figura 36. Através do mesmo, o usuário do SSF consegue encontrar os melhores parâmetros para o estudo desejado. É possível notar, em azul, uma opção de salvar. Esta opção serve para salvar as saídas, os pesos das redes e todos outros parâmetros encontrados na busca.

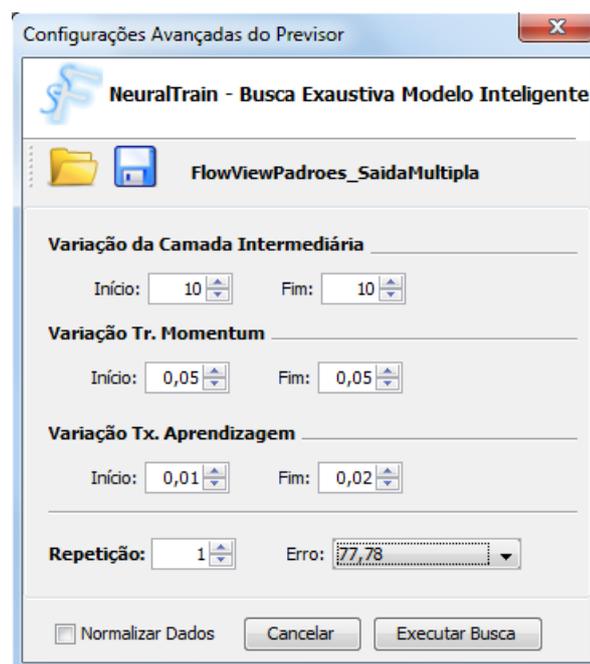


Figura 36 – Busca Exaustiva – *SupportSystemFlow*.

4

Modelos Inteligentes

“Todo modelo é errado, mas alguns são úteis” (BOX, 1979). O projetista é responsável por escolher os modelos que mais se adequam ao problema tratado.

Este Capítulo abordará sobre os modelos inteligentes usados para compor a abordagem híbrida para identificação de padrões proposta neste trabalho. No item 4.1. serão apresentadas as Redes Neurais Artificiais e suas variações. No item seguinte, 4.2., será explicado o Algoritmo K-Vizinhos, que é baseado em Distância Euclidiana. No item 4.3. encontra-se o Algoritmo *Support Vector Machine* (SVM), um modelo robusto que possui taxas altas de acertos. Finalmente, o item 4.4. apresenta o *Ensemble*, que através da combinação de todos os modelos pode propor classificações com maiores taxas de acertos.

4.1. Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são um ramo da Inteligência Artificial (IA), fazendo parte da IA Conexionista. "*Uma rede neural é formada por um conjunto de unidades de processamento simples que se comunicam enviando sinais uma para a outra através de conexões ponderadas*" (BILOBROVEC; MARÇAL; PILATTI, 2004).

IA Conexionista basicamente visa à representação e reprodução do comportamento de raciocínio cerebral por meio de algoritmos chamados de neurônios artificiais. Podendo trabalhar de modo *off-line* ou *online*, onde o treinamento da rede é feito antes das previsões (*offline*) e o treinamento é realizado constantemente junto com as previsões (*online*).

RNAs necessitam de entradas, pesos e camadas de neurônios para classificarem padrões, ou seja, gerarem saídas. Os pesos são individuais de cada neurônio são ajustados no treinamento do modelo. Alguns modelos inteligentes utilizam o atributo *bias* com objetivo de auxiliar o ajuste e a polarização dos pesos.

Na Figura 37, pode-se observar que apresentadas entradas, geram-se saídas correspondentes.

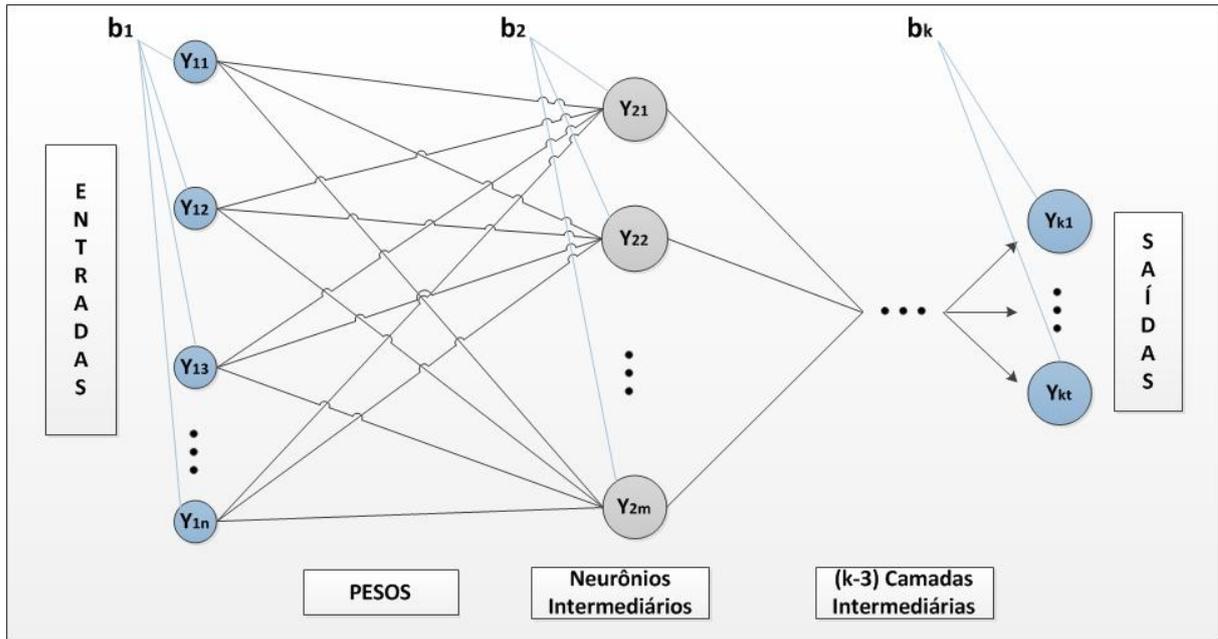


Figura 37 – Rede Neural Artificial.

Considerando Y um vetor de j neurônios (totais), W de j pesos e B de j *bias*, cada saída (S_k), Figura 38, pode ser calculada pela Fórmula (1):

$$S_k = \sum_{i=1}^j (Y_i \times W_i) + B_i \quad (1)$$

As RNAs utilizadas são do tipo MLP e possuem os algoritmos de treinamento propostos no relatório técnico de (Castro;Zuben,1998). São elas:

- *Back Propagation* com termo *momentum* (BPM);
- Método *Scaled Conjugate Gradient* (SCG);
- Método *Fletcher & Reeves* (FR);
- Método *Polak-Ribière* (PR);
- Método *One-Step Secant* (OSS);
- Método *Davidon-Fletcher-Powell* (DFP);
- Método *Broyden-Fletcher-Goldfarb-Shanno* (BFGS);
- Método *Gradient* (GRAD).

A Figura 38 apresenta um diagrama das diferentes estratégias de treinamento para os métodos citados. O objetivo do treinamento é encontrar erros mínimos (chamados mínimos globais). Das estratégias de treinamento são usados métodos de primeira e segunda ordem, onde os métodos de segunda ordem são melhores para encontrar o mínimo global do que o método de primeira ordem.

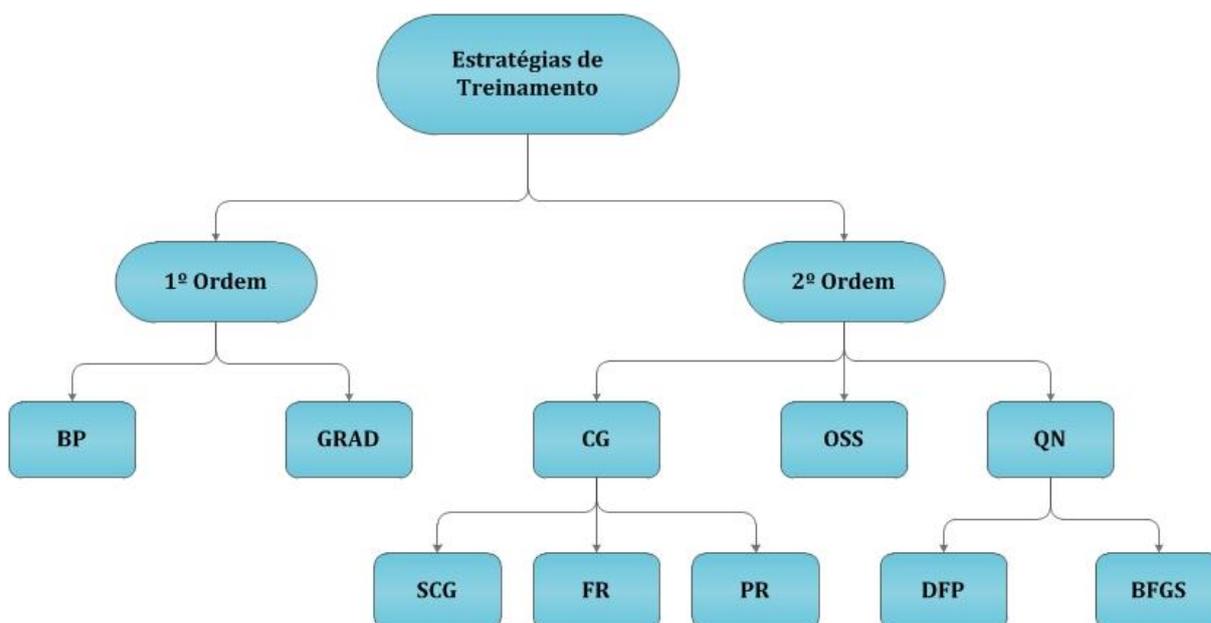


Figura 38 – Estratégia de Treinamento Rede Neural.

Fonte: (Castro & Zuben,1998).

Na Figura 37, observa-se que o elemento “CG” usado para os modelos SCG, FR e PR significa *Conjugate Gradient* ou seja, Gradiente Conjugado. Já o elemento “QN” usado nos modelos DFP e BFGS, significa *quasi-Newton* ou seja, semi *Newton*.

Além destes métodos citados, as RNAs contam com as funções de ativação. Estas funções são utilizadas para intensificar, reduzir ou apenas propagar o sinal de cada neurônio da rede neural. Existem vários tipos de funções de ativação, podendo citar a linear, tangente hiperbólica, sigmóide logística, gaussiana e outras. Para o bom funcionamento da rede neural, os dados que estas funções tratam precisam estar dentro de seus intervalos específicos. Para isso, geralmente são realizadas normalizações em todo o conjunto antes das classificações e depois da classificação é realizado o processo inverso. Se os dados de entrada para a função não tiverem em seu intervalo determinado, não significa que a rede não funcione, mas se estiverem, a probabilidade de acerto será maior.

4.2. Algoritmo K-Vizinhos

O Algoritmo K-Vizinhos pode ser encontrado também como Algoritmo da Distância Euclidiana ou também como *k-Nearest Neighbor* (KNN).

Segundo (Weinberger; Blitzer; Saul, 2006) “A regra do Algoritmo K-Vizinhos é classificar cada padrão pelo rótulo entre os seus *k* vizinhos mais próximos, no conjunto de treinamento, ou seja, cada entrada é combinada com um valor do treino correspondente. Seu desempenho, portanto, depende fundamentalmente da distância métrica usada para identificar vizinhos mais próximos”.

A Fórmula (2), Distância Euclidiana, apresenta o cálculo necessário para a identificação do padrão mais próximo no conjunto de treino. Sendo $P = (x_1, \dots, x_n)$ e $Q = (y_1, \dots, y_n)$:

$$D(P, Q) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

A proposta de classificação do KNN é a mais simples de todos os modelos deste trabalho, porém isso não significa que sua taxa de acerto esta entre as piores. Pelo contrário, o mesmo apresenta-se como um bom classificador. Isso é importante, pois garante a variabilidade que o *Ensemble* requer para garantir uma boa combinação.

4.3. Algoritmo Support Vector Machines (SVM)

A técnica de classificação do *Support Vector Machines* é conhecida como *Support Vector Classification* com capacidade tanto de classificar padrões binários quanto padrões do tipo multi-classe, Figura 39. Para realizar a classificação de uma amostra a técnica utiliza o conceito de margem, ou seja, a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado para a separação dessas classes (LORENA et al., 2007).

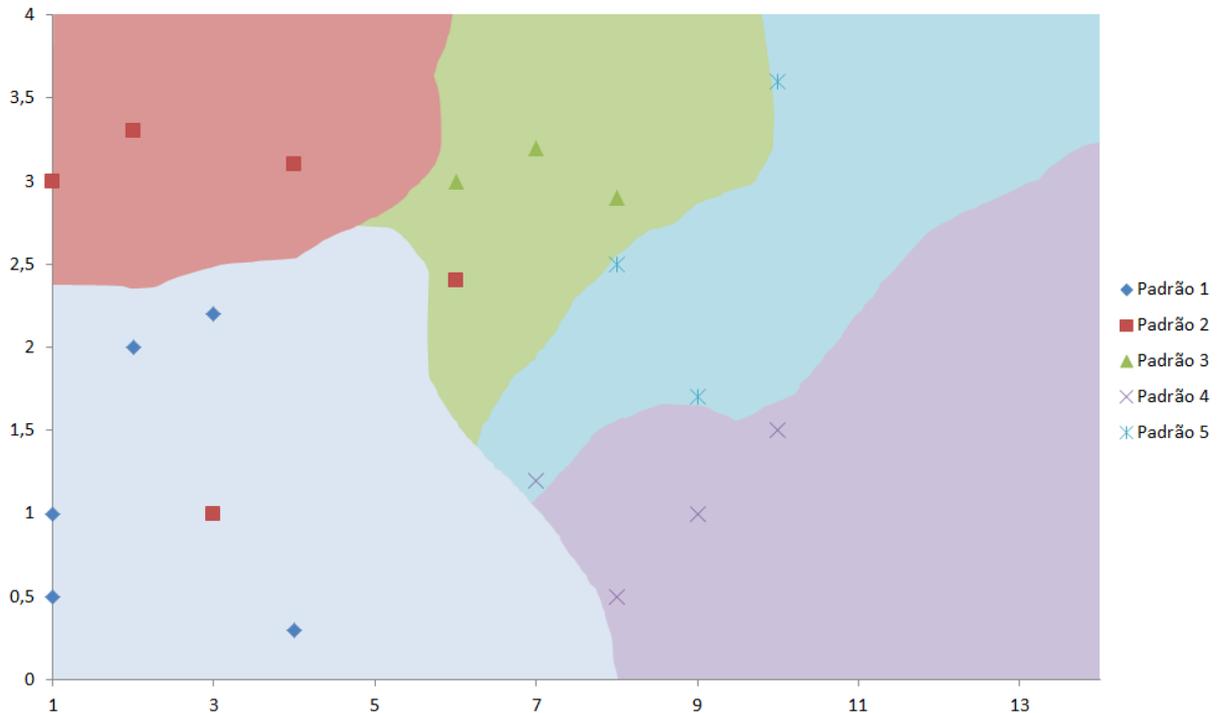


Figura 39 – Classificação SVM.

Para compor a margem de separação entre duas classes o SVC utiliza os vetores suporte. Esses vetores são pontos que delimitam a margem da classificação. Ou seja, o ponto limite mais próximo da outra classe. Ao destacar estes pontos o SVC não utiliza os outros pontos da amostra e eles são considerados irrelevantes, facilitando assim o processamento (Smola et al, 2000). Esses pontos limites podem ser traçados utilizando reta (lineares) ou através de outra função, conhecida como função *Kernel*.

Entretanto o problema pode não ser linearmente separável e nesse caso é empregado o conceito de erro para que o método da separação de hiperplanos possa ser generalizado. Mas ao empregar o conceito de erro a complexidade da técnica cresce drasticamente, por exemplo, para regularizar um dos parâmetros empregados, o C (*cost*), na maioria das vezes é necessário escolhê-lo através de buscas dinâmicas (Cortes, 1995). O treinamento da técnica pode ser resumida por (Semolini, 2002):

“O treinamento da SVM consiste em um problema de otimização quadrático que é atrativo pela garantia da convergência para um mínimo global da superfície de erro (exceto quando algum problema de precisão numérica está presente), onde o erro refere-se à diferença entre a resposta desejada e a saída da SVM”.

Tanto para o treinamento quanto para a classificação dos padrões de escoamento bifásico proposto neste trabalho foi utilizada a biblioteca *LibSVM* que é uma implementação das SVM para computadores (Chang; Lin, 2001) possuindo um grande prestígio na literatura.

4.4. Ensemble

O *Ensemble* ou combinador de algoritmos tem o objetivo de combinar os modelos inteligentes com o intuito de atingir melhores resultados. Existem vários tipos de abordagens para se configurar um *Ensemble*. Porém será explicada a mais simples, podendo estar entre as mais eficientes, dependendo da composição de seus dados.

A Figura 40 exemplifica o funcionamento de um *Ensemble*. Vale notar que os modelos inteligentes são treinados em paralelo e em seguida cada um propõe suas saídas. O módulo de combinação seleciona os padrões que ocorrem com maior frequência entre os modelos. Então, o *Ensemble* apresenta sua saída combinada. Se os componentes possuírem variabilidade, as chances de aumentarem as taxas de acertos são altas.

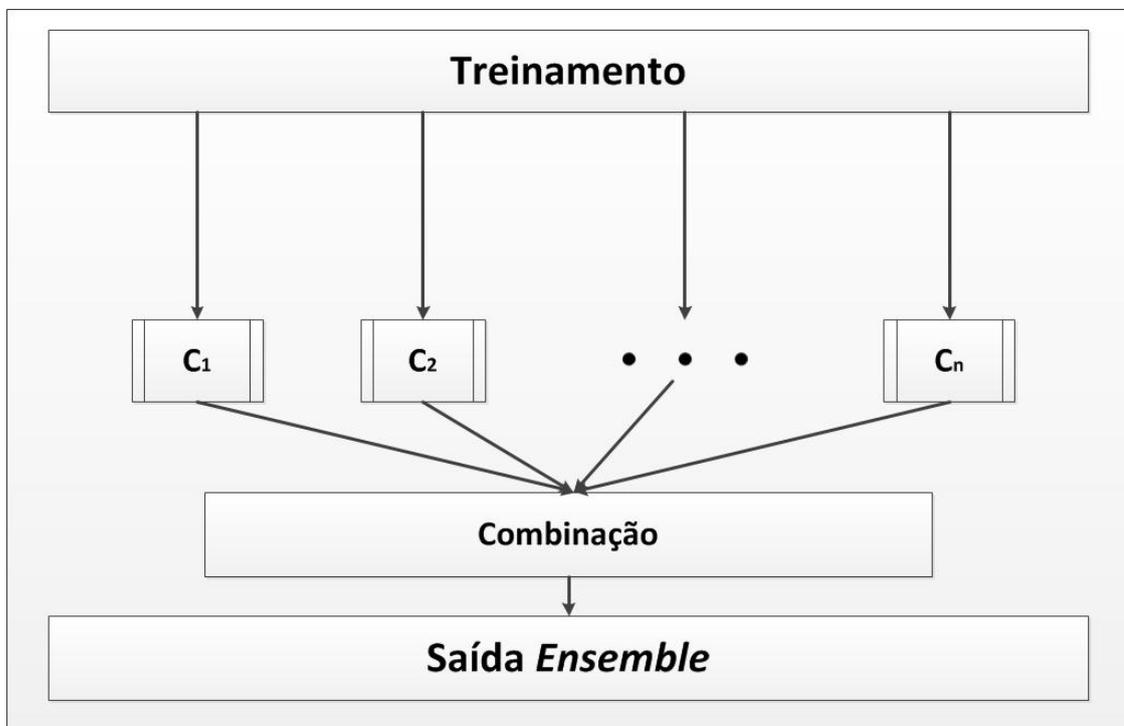


Figura 40 – Ensemble.

Observa-se, Figura 40, que o *Ensemble* pode possuir n componentes inteligentes (C_n), onde cada componente não depende do tipo dos outros, podendo representar qualquer modelo (RNAs, SVM, K-Vizinhos e outros).

Segundo (Hansen; Salamon, 1990):

“O uso de vários componentes inteligentes, com um senso, desempenha-se muito melhor do que usar um único componente. Componentes de treino com taxas ligeiramente inferiores são um subproduto, é desejável a utilização de tais componentes, mesmo quando seu desempenho é significativamente pior do que o melhor desempenho encontrado. O melhor desempenho, do Ensemble, ainda pode ser alcançado através de um planejamento cuidadoso para uma classificação do mesmo usando os melhores parâmetros disponíveis em conjunto com diferentes componentes”.

Um fato interessante, citado anteriormente, é que mesmo com os piores componentes encontrados nas buscas e garantindo a presença de alguns melhores, ainda assim o *Ensemble* consegue atingir altas taxas de acertos. Isso ocorre devido ao fator de variabilidade da composição do mesmo, onde cada componente possui uma taxa específica de acerto, e estes componentes nem sempre erram juntos, ou seja, não cometem o mesmo erro específico; Com isto, no momento da combinação, estes erros são ignorados por não se repetirem em todos os modelos, e há uma maior probabilidade dos acertos se repetirem mais que os erros, pois os componentes foram treinados e selecionados; Assim o *Ensemble* escolhe a opção certa, atingindo uma taxa de acerto superior ou próxima da melhor componente que o compõe.

O *Ensemble* comete erros quando os componentes que o compõe erram as mesmas classificações e optam pelo mesmo erro. Isso pode ocorrer por aleatoriedade ou então por problemas de modelagem com o padrão específico.

Quando ocorre um empate no número de ocorrências de padrões no momento da escolha do padrão do *Ensemble*, deve-se adotar um critério de desempate, mesmo que seja aleatório. Porém é aconselhável que se criem regras de desempates, para que as probabilidades de escolha da opção correta se tornem maiores. Um exemplo de regra de desempate, pode ser a escolha da opção fornecida pelo componente que participa do empate e possui a maior taxa de acerto até o momento de execução do ocorrido (taxa de acerto parcial).

Estes fatos citados anteriormente podem ser observados na Tabela 2. Supondo a existência de 4 componentes inteligentes, 10 padrões (P(1),...,P(10)) no conjunto total e 7 saídas a serem classificadas por cada componente.

Tabela 2 – Funcionamento *Ensemble*.

Saída Esperada	Componente(1)	Componente(2)	Componente(4)	Componente(4)	<i>Ensemble</i>
P(1)	P(1)	P(2)	P(1)	P(7)	P(1)
P(2)	P(2)	P(2)	P(2)	P(6)	P(2)
P(9)	P(9)	P(5)	P(5)	P(9)	P(5)
P(7)	P(7)	P(9)	P(7)	P(1)	P(7)
P(10)	P(10)	P(10)	P(10)	P(10)	P(10)
P(1)	P(1)	P(3)	P(5)	P(8)	P(1)
P(4)	P(7)	P(4)	P(4)	P(4)	P(4)
Taxa de Acerto(%)	85,72	42,86	71,43	42,86	85,72

O fato de acerto do *Ensemble*, anteriormente comentado, pode ser observado em vários padrões da Tabela 2. Já o caso de erro do *Ensemble* é observado na saída esperada que possui o padrão P(9), onde o *Ensemble* cometeu um erro, por escolher o padrão P(5). Neste mesmo caso, ocorreu um caso de empate, porém o critério que foi utilizado não é eficiente para ser utilizado em inícios de execuções; No caso foi utilizado o critério de maior taxa de acerto parcial, porém mesmo assim não foi possível distinguir as escolhas devido a outro empate, então foi feita uma escolha aleatória.

Um caso de sucesso de desempate pode ser observado no penúltimo padrão da saída esperada; O *Ensemble* utilizou o critério de desempate de maior taxa de acerto parcial e acertou.

5

Metodologia Utilizada

Este Capítulo apresenta a metodologia utilizada para a realização deste trabalho. A metodologia utilizada pode ser considerada metodologia de experimentação, que teve o objetivo de aprimorar algoritmos de classificação para o problema do escoamento bifásico utilizando-se de um ensemble.

No item 5.1., apresenta-se a descrição dos métodos abordados frente ao problema de classificação da aquisição de escoamento bifásico. Na sequência, item 5.2., possui a descrição da criação dos modelos inteligentes citados no Capítulo 4. O terceiro item, 5.3., aborda sobre a construção do combinador de algoritmos inteligentes, o *Ensemble*. Finalmente, o último item, o item 5.4. encontram-se as execuções dos passos que levaram a construção do Sistema SSF.

5.1. Classificação de Aquisições

A classificação das aquisições de escoamento pode ser realizada de várias maneiras. Segundo (Corrêa, 2009b):

“A indicação dos padrões de escoamento é realizada a partir das medidas de certas propriedades do escoamento. Um parâmetro muito utilizado é obtido pela medida da fração do gás sobre uma linha, área ou volume em relação ao total de uma seção restrita da tubulação. Esta propriedade é denominada fração de vazio e pode ser obtida por várias técnicas: atenuação de radiação, raios X ou radiação gama; variação de pressão; fechamento rápido de válvulas; variação da impedância.” (CORRÊA, 2009b).

As aquisições deste trabalho foram obtidas por variação da impedância. Essa técnica de aquisição foi escolhida pelo Laboratório de Fenômenos Multifásicos da Faculdade de

Os sinais de tensão obtidos nas aquisições de escoamento foram transformados em momentos estatísticos e atrelados ao padrão correspondente, trabalho feito por um especialista em escoamento bifásico.

Neste trabalho, foi utilizado um conjunto de 174 aquisições, possibilitando a confecção de 5 conjuntos, 4 conjuntos com 35 aquisições e 1 conjunto com 34 aquisições. Esta proposta é feita para ser possível a variação de composição dos conjuntos de teste, validação e treino a fim de garantir que os modelos inteligentes selecionados para realizarem a classificação de padrões realmente generalizaram os padrões e não decoraram (*overfitting*). Todos os conjuntos são balanceados, ou seja, distribuições dos padrões de forma igual entre os conjuntos. Para a composição destes conjuntos foram executados os seguintes passos:

- Formou-se uma lista circular de 5 conjuntos vazios, preparados para receberem os padrões;
- Dividiu-se o conjunto total em 14 subconjuntos, separados por padrões;
- Para cada subconjunto padrão, enquanto o mesmo não ficou vazio;
 - Escolheu-se um elemento aleatoriamente;
 - Inseriu o escolhido na lista de conjuntos;
 - Removeu-se o elemento escolhido do subconjunto padrão;
 - Avançou o apontador da lista circular de conjuntos para o próximo elemento;

Após a divisão do conjunto total de padrões em 5 subconjuntos balanceados, sorteou-se um deles para fixá-lo como conjunto de Teste. Reservou-se os 4 conjuntos restantes para realizar a composição do conjunto de Validação (1 conjunto) e Treino (junção de 3 conjuntos) de maneira aleatória durante as buscas exaustivas para encontrar as melhores configurações de cada modelo.

Foram feitas várias etapas de buscas exaustivas e a cada etapa construiu-se todo o conjunto aleatoriamente, com os mesmos passos descritos anteriormente. Para cada modelo inteligente foram feitas várias buscas exaustivas com base em comparações no conjunto de Validação. Ao final de todas as buscas, escolheram-se os 200 elementos melhores entre todos os modelos para compor o *Ensemble*. Então todos os modelos inteligentes foram testados sobre o mesmo conjunto de Teste e geraram as saídas propostas para o combinador de algoritmos. Os resultados destas etapas encontram-se no Capítulo 6.

A metodologia utilizada para a construção dos métodos de classificação, presentes no Capítulo 4, é detalhada no item seguinte.

5.2. Busca Exaustiva

Busca exaustiva é o processo de buscar um objetivo variando-se as formas e combinando todas as possibilidades de se atingir este objetivo. Para a realização da busca exaustiva neste trabalho, cada modelo inteligente utilizou um número e intervalo de variáveis diferentes a serem combinadas. O objetivo da busca foi encontrar os melhores componentes de acordo com a comparação da maior taxa de acerto sobre o conjunto de validação.

Para todas as buscas foram realizadas a composição aleatória dos conjuntos, apenas fixando o conjunto Teste. Isso foi feito para garantir que os modelos inteligentes não decoraram os padrões (*overfitting*) e realizar as comparações de taxas de acertos (fixou-se o mesmo conjunto de Teste). Porém para gerar os resultados deste trabalho, foram executadas várias vezes a execução completa de todos os modelos e a cada execução o conjunto de teste variou, possibilitando gerar as taxas de acertos de cada modelo pela média das taxas de acertos das execuções.

Para o modelo inteligente K-Vizinhos a busca exaustiva foi feita apenas variando o conjunto de treinamento, que é o único atributo que pode variar neste modelo.

Vale notar que para realizar o ajuste de todos os parâmetros dos componentes, foram feitas execuções em máquinas separadas a fim de distribuir o processamento. O tempo utilizado para a realização total desta etapa foi de aproximadamente 30 horas em dois computadores com características como: processador *Intel Core 2 Quad*, 2,8 GHz, 8 GB de memória RAM e sistema operacional *Windows 7* de 64 Bits.

A seguir serão apresentadas as buscas exaustivas realizadas nos modelos RNAs e SVM.

5.2.1. Busca Exaustiva - RNAs

A busca exaustiva para as RNAs foi realizada utilizando a variação dos atributos presente na Tabela 5. Todos os modelos RNAs utilizaram o erro mínimo e número máximo de épocas com valor de 0.001 e 3000 respectivamente. Todos estes valores apresentados se referenciam aos atributos de treino do modelo.

Tabela 5 – Busca Exaustiva - RNAs.

Modelo	Tx. Ap.	Nº Neu. C.I.	Tr. Mom.
BPM	[0.001 – 0.5]	[3 - 15]	[0.001 – 0.5]
SCG	-	[3 - 15]	-
FR	-	[3 - 15]	-
PR	-	[3 - 15]	-
OSS	-	[3 - 15]	-
DFP	-	[3 - 15]	-
BFGS	-	[3 - 15]	-
GRAD	-	[3 - 15]	[0.001 – 0.5]

Onde Tx. Ap. Significa taxa de aprendizagem, Nº Neu. C.I. significa número de neurônios na camada intermediária e Tr. Mom. significa termo *momentum*.

5.2.2. Busca Exaustiva - SVM

A busca exaustiva para os modelo SVM foi realizada utilizando a variação dos atributos presente na Tabela 6. Todos estes valores apresentados se referenciam aos atributos de treino do modelo.

Tabela 6 – Busca Exaustiva - SVM.

Atributo	Intervalo
Função de Erro	[C-SVC]
Parâmetro c (custo)	[1 – 10000]
Parâmetro g (<i>gamma</i>)	[0.0001 – 500]
Parâmetro r (coef0)	[0 – 500]
Parâmetro d (<i>degree</i>)	[1 – 3]
Parâmetro t (tipo do <i>kernel</i>)	[0 – 2]

5.3. Modelos Inteligentes

Os modelos inteligentes propostos neste trabalho tem o objetivo de classificar de modo eficiente padrões de escoamento bifásico. O Sistema SSF oferece uma facilidade na utilização dos modelos inteligentes, através de interfaces amigáveis e boa navegabilidade. A classificação de padrões só é viabilizada por conta destes componentes. A etapa de elaboração e implementação dos citados, foi uma etapa priorizada neste trabalho.

Grande porcentagem da confiabilidade de um Sistema de Suporte à Decisão ou *Decision Support System* (DSS) está presente em suas taxas de acertos. Não justifica possuir um DSS completo de interfaces bem projetadas, BD robusto e o mesmo falhar em suas decisões. Para resolver isto, a proposta deste trabalho foi implementar vários ramos que classificam padrões de escoamento bifásico e suportassem a junção dos dois regimes, horizontal e vertical.

O problema tratado pelos modelos inteligentes foi à classificação de padrões de escoamento bifásico nos regimes vertical e horizontal. As entradas foram às mesmas para todos os modelos, excluindo-se o Ensemble que realizou a combinação dos resultados; As entradas foram os seguintes momentos estatísticos da aquisição dos valores de tensão do escoamento:

- Média;
- Desvio Padrão;
- Assimetria;
- Curtose.

5.3.1. Redes Neurais Artificiais

Para utilizar as RNAs neste trabalho, foram feitos os seguintes passos:

- Para cada algoritmo de treinamento proposto (BPM, SCG, FR, PR, OSS, DFP, BFGS e GRAD);
 - Foi executada a busca exaustiva referente ao algoritmo;
 - Selecionou-se as 20 melhores configurações encontradas;
- Após ter selecionado as melhores configurações de cada algoritmo, estes foram acoplados aos melhores componentes inteligentes que compuseram o *Ensemble*.

A Figura 41 e 42 mostram as modelagens feitas para as RNAs deste trabalho, com saída única e múltipla respectivamente.

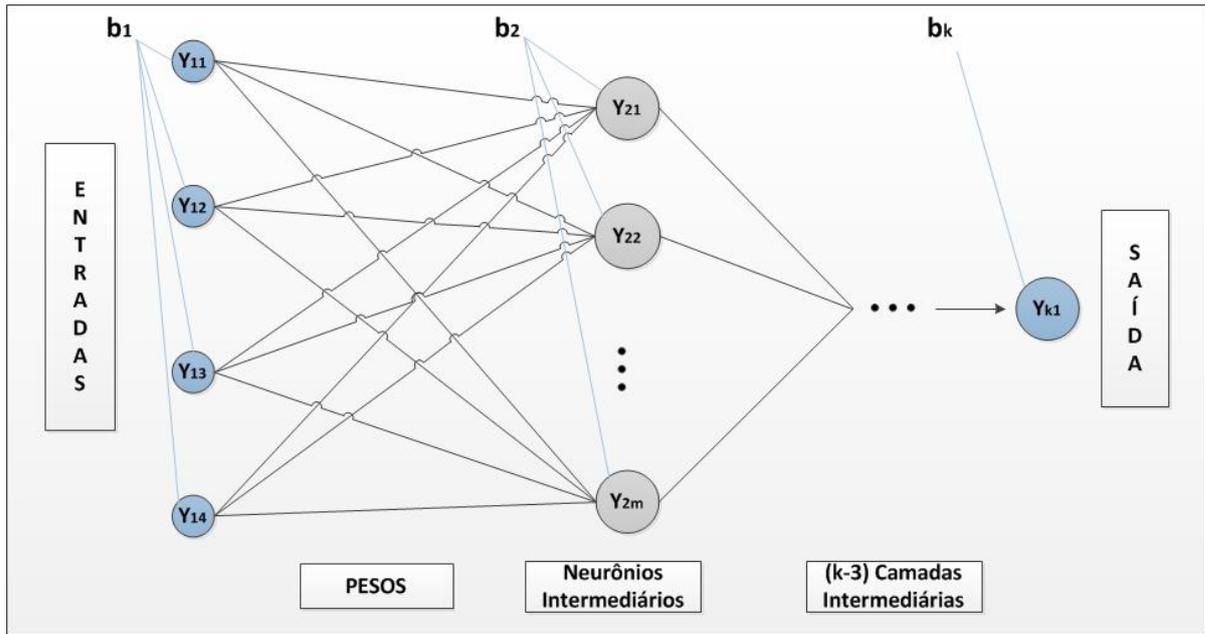


Figura 41 – Rede Neural Artificial – Saída Única.

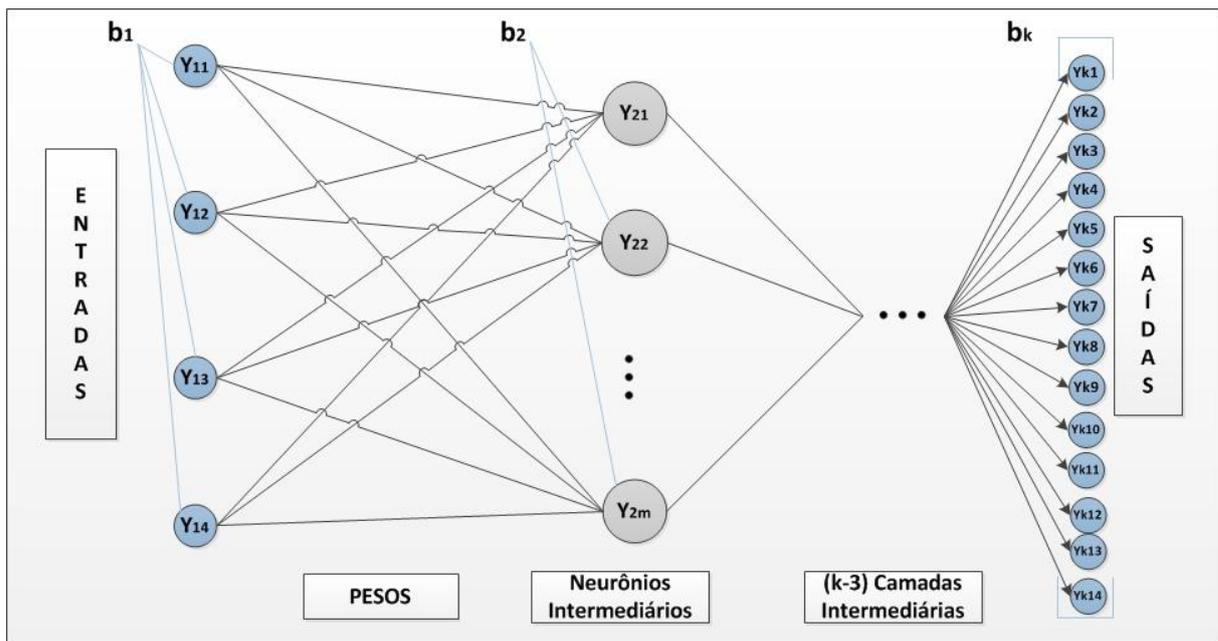


Figura 42 – Rede Neural Artificial – Saída Múltipla.

5.3.2. Algoritmo K-Vizinhos

O Algoritmo K-Vizinhos utilizado neste trabalho seguiu os seguintes passos:

- A cada entrada inserida no algoritmo a ser classificada, percorreu-se todo o conjunto de treinamento e para cada elemento do conjunto de treino;

- Calculou-se a Distância Euclidiana, Fórmula 2 apresentada no Capítulo 4;
- Armazenou-se a menor distância da entrada com um elemento do conjunto treino;
- Após encontrar a menor distância entre a entrada e um elemento do conjunto de Treinamento, atribui-se o padrão do elemento do conjunto de treino como saída equivalente para a entrada informada.

A Figura 43 ilustra a classificação feita pelo Algoritmo K-Vizinhos com base no conjunto de treinamento. Vale ressaltar que o valor de K utilizado foi igual ao número de entradas dos modelos inteligentes, ou seja, quatro entradas.

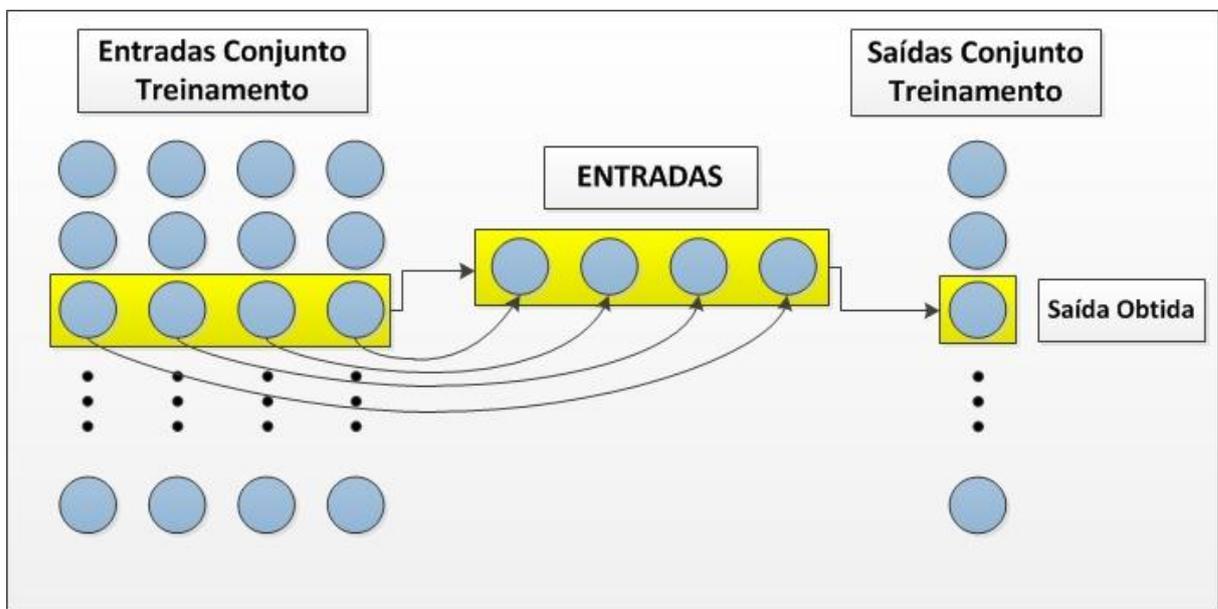


Figura 43 – Classificação K-Vizinhos.

5.3.3. Algoritmo *Support Vector Machines* (SVM)

Para a utilização da LIBSVM neste trabalho, foram executados os seguintes passos:

- Foi montado um vetor de configurações de *kernels* para a busca exaustiva;
- Foi feita a busca exaustiva e selecionados 50 melhores componentes;
- Estes 50 melhores componentes foram acoplados com os melhores componentes de outros modelos inteligentes para compor o *Ensemble*.

As taxas de acertos encontradas com este modelo inteligente foram ótimas. Uma análise dos resultados deste e de outros componentes será feita no Capítulo 6.

5.4. Ensemble

O *Ensemble* ou combinador de algoritmo é um componente fundamental para este trabalho. Nele se asseguram os melhores acertos do Sistema SSF. Seu objetivo central é escolher o melhor modelo inteligente que se adeque ao padrão apresentado; Essa escolha se realiza através de históricos de acertos de cada modelo ao padrão identificado e também ao número de saídas semelhantes entre todos os modelos. Sua construção é relativamente simples em relação a outros componentes. Porém, também foi atribuída grande importância na construção deste *Ensemble* por se tratar de um componente prioritário onde o mesmo se integra.

O *Ensemble* proposto é composto pelos resultados dos modelos inteligentes, RNAs, K-Vizinhos e SVM. Para a construção do mesmo, foi feita uma estrutura que suporta análises de sistemas externos. Com isso, o mesmo assegura sua variabilidade de componentes inteligentes e uma maior generalização.

“Vale destacar que um dos requisitos fundamentais para obter sucesso ao utilizar as técnicas de ensembles, é que os componentes de previsão generalizem de forma diferente, pois não faz sentido combinar modelos que adotam os mesmos procedimentos e hipóteses para a solução de um problema” (LIMA, 2010).

A Figura 44 ilustra o funcionamento do *Ensemble* Híbrido proposto e construído neste trabalho.

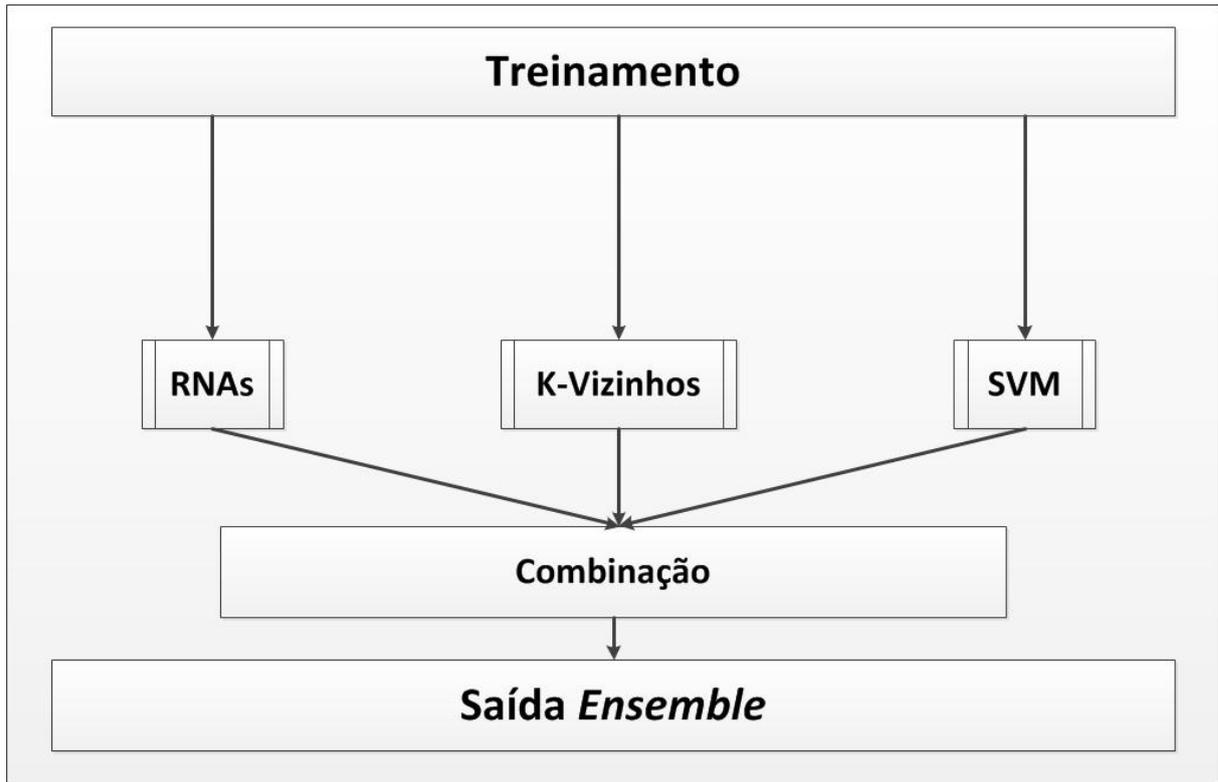


Figura 44 – *Ensemble* Híbrido.

Vale observar que os treinamentos são executados em paralelo, Figura 44, e que o combinador de algoritmo pode ser composto por vários modelos inteligentes, sem um número fixo definido. Para os resultados deste trabalho foram utilizados os 200 melhores componentes inteligentes, encontrados com as buscas exaustivas.

6

Resultados

Este Capítulo apresenta os resultados obtidos nesta monografia. O item 6.1. expõe os resultados atingidos para a base de dados proposta. O segundo item, 6.2., apresenta os resultados obtidos com relação ao Sistema SSF. E finalmente, o item 6.3. relata sobre os resultados e taxas de acertos obtidas pelos componentes inteligentes apresentados no Capítulo 4.

6.1. Base de Dados

Os resultados, para a base de dados criada, encontrados foram satisfatórios. Com a utilização total dos Sinais de aquisições, o Sistema se mostra lento e inviável dependendo da situação (tomadas de decisões em tempo real), porém se mostra bem detalhado e eficiente se a situação for análises minuciosas de aquisições ocorridas (onde a preocupação com o tempo de processamento não é tão crítica). Em contra partida, o Sistema se mostra hábil a ser utilizado em tempo real através dos métodos de reduções e bases de dados reduzidas.

As Figuras 45 e 46 apresentam gráficos das consultas feitas no BD projetado para o SSF. Ambos os gráficos representam consultas realizadas na relação “dado”, Figura 4, seguindo a ordem de consultas repetidas (identificadores iguais) e consultas sorteadas (identificadores diferentes). Vale notar que as legendas VP e SP, significam respectivamente que as consultas foram realizadas em conjunto com vários processos diferentes executando em uma mesma máquina e as consultas foram realizadas isoladamente nesta mesma máquina, sem a existência de outros processos (*Single Process*).

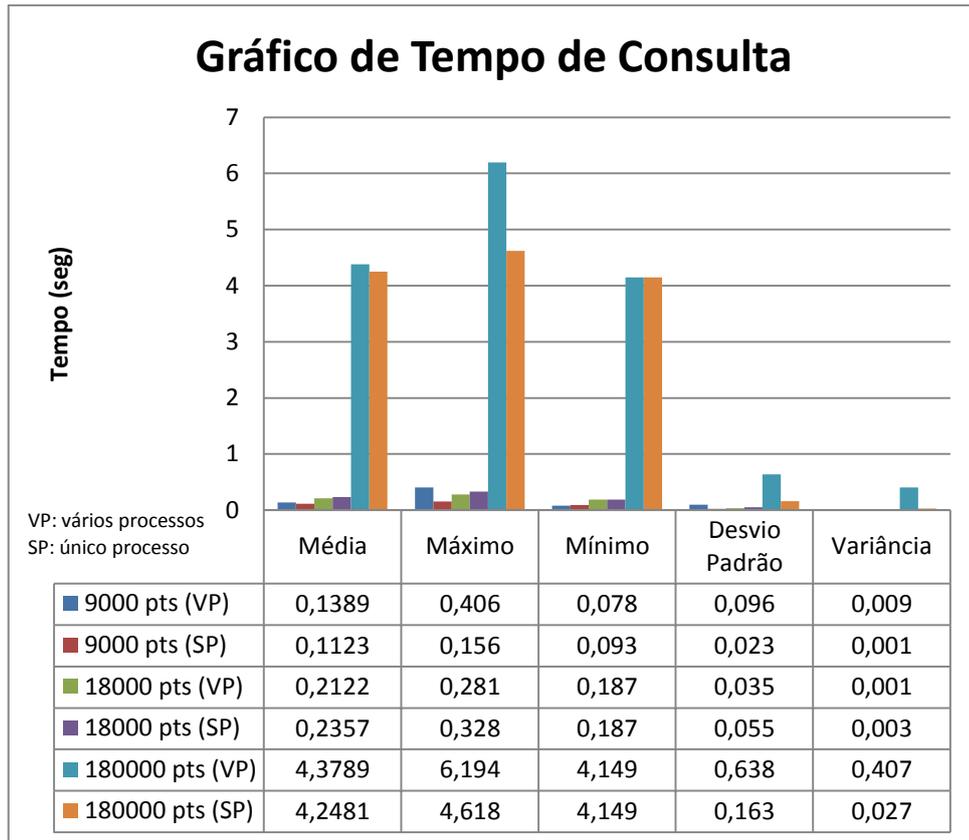


Figura 45 – Gráfico Tempo de Consulta I.

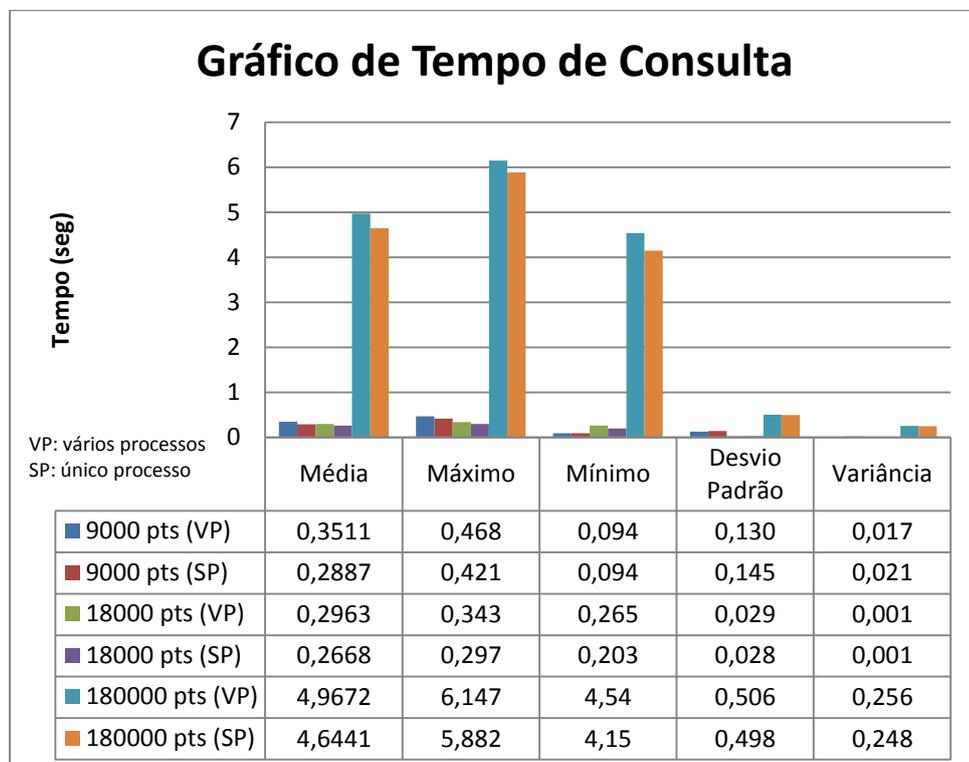


Figura 46 – Gráfico Tempo de Consulta II.

6.2. Componentes Inteligentes e *Ensemble*

Para obtenção dos resultados dos componentes inteligentes, foram feitas buscas exaustivas, com base no conjunto de validação, variando os parâmetros possíveis de cada modelo. As taxas de acertos obtidas, em relação ao conjunto de teste, foram significativamente boas.

Tratando os regimes de escoamento, vertical e horizontal, de maneiras isoladas (separadamente). As taxas de acertos para os dois regimes foram de 96% a 100%, tanto para os componentes inteligentes quanto para o *Ensemble*.

Porém o maior desafio era tratar os dois regimes nos mesmos conjuntos de treinamento, teste e validação, sem discriminá-los. Para realizar esta abordagem, foi utilizada uma técnica de normalização, que realiza a mudança dos intervalos dos dados sem apresentar perdas estatísticas. Sendo a Fórmula (3) a mudança do intervalo de tensão da aquisição para o intervalo [0 1].

$$DN(dO) = \frac{(1 - 0) \times (dO - valorMinAbs)}{(valorMaxAbs - valorMinAbs) + 0} \quad (3)$$

Observando que a variável “*DN*” significa o dado normalizado; “*dO*” significa o dado original em questão a ser normalizado; “*valorMinAbs*” o menor valor entre todos os dados do conjunto a ser normalizado e “*valorMaxAbs*” o máximo.

O número de padrões utilizados neste trabalho para a obtenção dos resultados de classificação foi: 85 padrões para o regime de escoamento vertical e 89 padrões para o regime de escoamento horizontal. Sendo que para a construção dos conjuntos de treinamento, validação e teste, foi respeitado o critério de balanceamento e aleatoriedade, citado no Capítulo 5. O conjunto de treinamento foi composto por três quintos do total dos padrões e os conjuntos de validação e teste com um quinto cada um.

A seguir serão apresentadas as duas formas utilizadas para obtenção dos resultados.

6.2.1. Normalização no intervalo [0 1]

Para a obtenção dos resultados deste item, todos os dados, os 174 padrões, foram normalizados conforme a Fórmula (3), ou seja, para o intervalo [0 1]. Esta normalização é

feita para que todos os dados de entrada fiquem no intervalo de qualquer função de ativação de uma RNA, assim podendo melhorar seu funcionamento.

A Tabela 7 apresenta as porcentagens de acertos obtidas, a partir de várias repetições, sobre o conjunto de teste, possuindo um desvio percentual de mais ou menos 2,5.

Tabela 7 – Tabela de Resultados Sobre o Conjunto de Teste [0 1].

Componente Inteligente	Taxa de Acerto (%)
BPM	57,14
SCG	54,28
FR	62,85
PR	51,42
OSS	48,57
DFP	45,71
BFGS	54,28
GRAD	57,14
K-Vizinhos	71,42
SVM	71,42
ENSEMBLE	65,71

Observa-se que o *Ensemble* não atingiu o melhor resultado possível, isso pode ocorrer devido à proporção de sua composição, ou seja, um percentual significativo de componentes pode ter errado o mesmo padrão levando o *Ensemble* ao erro devido à maioria.

A Tabela 8 apresenta a porcentagem dos erros de cada padrão comparando todos os componentes inteligentes com relação ao *Ensemble*.

Tabela 8 – Porcentagem Erro Padrões de Escoamento I.

Padrão de Escoamento	Componentes (%)	Ensemble (%)
Cheio	2,90	0,00
Bolhas	5,43	8,33
Capa Esférica	11,38	8,33
Golfadas	3,14	0,00
Agitado	9,00	8,33
Semi-Anular	6,10	8,33
Anular	6,22	0,00
Estratificado	12,19	16,67
Pistonado	14,55	25,01
Bolhas Dispersas	11,89	16,67
Disperso	10,97	8,33
Ondular	2,60	0,00
Pistonado Aerado	0,00	0,00
Vazio	3,63	0,00

Pode ser observado na Tabela 8, que o padrão Anular obteve 0,12% de erro superior ao padrão Semi-Anular em componentes inteligentes e o *Ensemble* não errou o padrão citado. Esta situação pode ocorrer, pois não significa que quando os componentes erram, optaram pelas mesmas escolhas, ou seja, o mesmo padrão; Em contra partida isso pode ter ocorrido com o padrão Semi-Anular, conseqüentemente, induzindo o *Ensemble* ao erro.

Também pode ser observado que os padrões que obtiveram maiores taxas de erros nos componentes inteligentes se refletiram no *Ensemble*, excluindo o padrão Anular que se encontra em um caso particular.

Nota-se que esta proposta, normalização para o intervalo [0 1] não atingiu boas taxas de acertos, um dos motivos que justificam estas baixas taxas é a proximidade dos dados. Isso quer dizer que o modelo acaba optando pelo padrão mais próximo ao valor calculado pelo componente e se as distâncias destes padrões forem muito pequenas, intervalos muito próximos, pequenas variações na classificação podem levar o componente inteligente a errar. Se isso ocorrer com todos os componentes, todas as taxas serão baixas e conseqüentemente a taxa de acerto do *Ensemble* também será baixa.

6.2.2. Sinal original - intervalos [0 1] e [2 5] – Caso 1

Para a obtenção dos resultados deste item, os dados dos regimes de escoamento foram utilizados nas tensões que os mesmos são adquiridos, [0 1] para o regime vertical e [2 5] para o regime horizontal. Mesmo os regimes sendo de tensões diferentes, eles continuaram compondo apenas um conjunto total.

A Tabela 9 apresenta as porcentagens de acertos obtidas sobre o conjunto de teste, possuindo um desvio percentual de mais ou menos 2,5.

Tabela 9 – Tabela de Resultados Sobre o Conjunto de Teste [0 1][2 5].

Componente Inteligente	Taxa de Acerto (%)
BPM	65,71
SCG	88,57
FR	82,85
PR	88,57
OSS	80,00
DFP	82,85
BFGS	74,28
GRAD	71,42
K-Vizinhos	82,85
SVM	88,57
<i>ENSEMBLE</i>	91,43

A Tabela 10 apresenta a porcentagem dos erros de cada padrão comparando todos os componentes inteligentes com relação ao *Ensemble*.

Tabela 10 – Porcentagem Erro Padrões de Escoamento II.

Padrão de Escoamento	Componentes (%)	Ensemble (%)
Cheio	6,56	0,00
Bolhas	7,87	0,00
Capa Esférica	18,74	33,33
Golfadas (Slug)	23,43	33,33
Agitado (UnsSlug)	11,72	0,00
Semi-Anular	5,53	0,00
Anular	19,02	33,33
Estratificado	2,62	0,00
Pistonado	0,19	0,00
Bolhas Dispersas	0,00	0,00
Disperso	0,38	0,00
Ondular	0,47	0,00
Pistonado Aerado	0,00	0,00
Vazio	3,47	0,00

Nota-se que os padrões que apresentaram uma maior porcentagem de erro no grupo dos componentes inteligentes, Tabela 10, foram os mesmos padrões que levaram o *Ensemble* ao erro de aproximadamente 8,57% sobre o conjunto de teste, Tabela 9.

Pode-se observar que esta proposta de utilização dos dados de maneira bruta se comportou muito melhor que a primeira proposta. Um dos motivos para este comportamento foi consequente da utilização dos dados originais, que deixaram uma maior distância entre os padrões e dentro de um mesmo conjunto de dados os modelos conseguiram distinguir dois tipos de conjunto de dados diferentes, o regime vertical e o regime horizontal.

6.2.3. Sinal original - intervalos [0 1] e [2 5] – Caso 2

Para o caso 2, utilização dos sinais de forma original no intervalo de [0 1] volts para o regime vertical e [2 5] volts para o regime horizontal, foram realizadas 100 repetições similares ao caso 1. Essas repetições foram feitas com o intuito de verificar a validade dos modelos, ou seja, a média dos resultados são os reais resultados de cada modelo.

Inicialmente foram gerados 100 grupos com os conjuntos de treino, teste e validação distintos, gerados aleatoriamente, apenas com similaridades de balanceamento. Este balanceamento garantiu a distribuição igualitária, entre os conjuntos, dos 174 padrões de aquisição utilizados neste trabalho que podem ser classificados por 14 padrões de

escoamentos. A cada iteração, foram feitas classificações em relação ao conjunto de validação com o intuito de selecionar os 100 melhores componentes inteligentes. Utilizou-se destes componentes escolhidos para realizar as classificações no conjunto de teste e combiná-las através do *Ensemble* para atingir melhores resultados.

A Figura 47 apresenta um gráfico com os momentos estatísticos de máximo, mínimo e média das classificações de cada componente inteligente para as 100 repetições, ou seja, para a classificação dos 100 conjuntos de testes gerados.

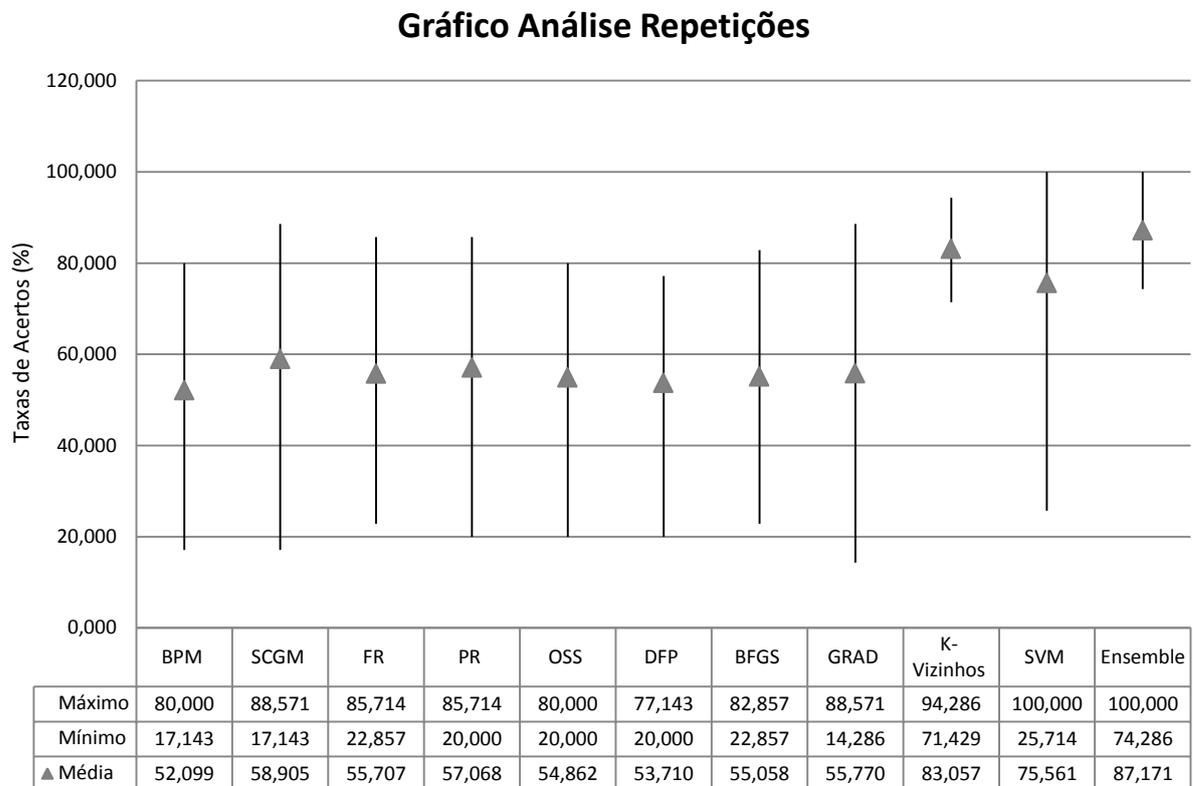


Figura 47 – Gráfico de Análise das Repetições.

Observa-se que a média do *Ensemble* não se apresentou muito alta, porém seu desvio padrão foi de 4,79%, consideravelmente razoável. Utilizando o desvio padrão, pode-se observar que a média de acerto no item anterior encontra-se no intervalo de acerto apresentado neste item, 91,43% para o caso 1 e o intervalo de 82,38% a 91,96% para o caso 2.

A Tabela 11 apresenta a média das repetições referente à porcentagem dos erros de cada padrão comparando todos os componentes inteligentes com relação ao *Ensemble*.

Tabela 11 – Médias de Erros dos Padrões de Escoamento.

Padrão de Escoamento	Componentes (%)	Ensemble (%)
Cheio	0,508	0,668
Bolhas	0,000	0,000
Capa Esférica	4,122	3,341
Golfadas (Slug)	0,253	0,223
Agitado (UnsSlug)	21,928	22,049
Semi-Anular	3,567	2,227
Anular	1,906	2,004
Estratificado	14,613	14,699
Pistonado	7,779	7,572
Bolhas Dispersas	9,233	9,131
Disperso	10,841	11,136
Ondular	7,978	8,909
Pistonado Aerado	15,928	16,481
Vazio	1,344	1,559

Pode-se observar que o *Ensemble* obteve uma média de erros similar as médias dos componentes inteligentes. Isso ocorreu devido à aleatoriedade dos conjuntos de treino, teste e validação a cada repetição, caso esperado com normalidade. Todas as análises realizadas, neste item de capítulo, comprovam que os componentes inteligentes e a combinação dos mesmos, através do *Ensemble*, apresentaram-se estáveis em suas taxas de acertos. Concluindo que a modelagem deste trabalho, mesclando os regimes verticais e horizontais é possível e sua utilização é segura.

7

Análises e Conclusões

Este Capítulo abordará sobre análises e conclusões deste trabalho como melhores componentes inteligentes encontrados, a relação do *Ensemble* com os componentes, objetivos cumpridos e sugestões para trabalhos futuros.

7.1. Melhores Componentes Inteligentes

Para cada tipo proposto de utilização dos conjuntos, foi possível encontrar alguns componentes com melhores rendimentos. Com a normalização do conjunto total para o intervalo [0 1] os melhores componentes foram o Algoritmo K-Vizinhos e o Algoritmo *Support Vector Machine* (SVM) com 71,42% de taxa de acerto sobre o conjunto Teste. Onde o erro das RNAs pode ser explicado devido aos pequenos intervalos entre os padrões. Pois, como as mesmas foram projetadas para generalizar e não decorar (*overfitting*), elas possuem um desvio padrão em suas classificações, esse desvio pode ser maior que o intervalo dos padrões, podendo levar a RNA a errar.

Para a utilização dos dados reais, com o conjunto total no intervalo [0 1] para o subconjunto do regime vertical e intervalo [2 5] para o subconjunto do regime horizontal; Os melhores componentes foram o Algoritmo *Support Vector Machine* (SVM), a RNA com o método Polak-Ribière (PR) e a RNA com o método Scaled Conjugate Gradient (SCG) todos com a mesma taxa de acerto, 88,57%, caso 1. E para o caso 2, o melhor componente nas repetições foi o K-Vizinhos com taxa de acerto de 83,05%.

Vale observar que o componente inteligente que mais se adequou para a modelagem dos dados deste trabalho foi o Algoritmo *Support Vector Machine* (SVM) sempre possuindo as melhores taxas. Podendo concluir que a proposta de utilização do mesmo combinado com outros componentes inteligentes, através do *Ensemble*, gera taxas superiores ou aproximadas ao melhor componente encontrado nas buscas exaustivas.

7.2. *Ensemble* x Componentes Inteligentes

Confrontando-se as taxas de acertos obtidas isoladamente de cada componente inteligente, com as taxas de acertos obtidas pelo *Ensemble*; Pode-se concluir que a utilização do *Ensemble* é viável; Porém apenas quando se tem certeza de que os componentes que o compõe possuem uma boa variabilidade e pelo menos uma porcentagem significativa dos componentes possuem boas taxas de acertos para a modelagem específica tratada.

Este comportamento pode ser observado para a segunda proposta de utilização deste trabalho, dados originais, intervalos [0 1] e [2 5]; O *Ensemble* atingiu sua taxa de acerto superior aos seus melhores componentes 91,43% contra 88,57% respectivamente, caso 1, e 87,17% contra 83,05% respectivamente, caso 2.

Porém o contrário pode ser observado na primeira proposta de normalização, intervalo [0 1]; O *Ensemble* teve sua taxa de acerto 5,71% inferior aos seus melhores componentes. Nota-se que foi respeitada a variabilidade de componentes, porém os mesmos não atingiram boas taxas de acertos com a modelagem proposta, conseqüentemente o *Ensemble* não conseguiu atingir boas taxas de acertos.

Pode-se concluir com as análises anteriores que o uso do *Ensemble* em relação aos componentes pode ser vantajoso, podendo trazer ganhos.

7.3. Conclusões

As metas propostas por este trabalho foram além das expectativas. Foi possível produzir, além de um DSS completo, uma modelagem híbrida, de regime de escoamento vertical e horizontal, para o mesmo. O SSF, sistema proposto neste trabalho, pode ser considerado um DSS específico para o escoamento bifásico nos regimes vertical e horizontal. Possuindo BD específico e projetado para suportar grandes volumes de dados, não perdendo eficiência e velocidade de consulta; Além de possuir um ótimo componente inteligente, que foi a proposta principal deste trabalho, a construção de uma abordagem híbrida para identificação de padrões de escoamento bifásico.

O *Ensemble* criado atingiu boas taxas de acertos, mesmo generalizando os regimes vertical e horizontal, ou seja, não os distinguindo para treinos e classificações. Porém vale ressaltar que a utilização dos dados reais, sem passar por normalizações, gera maiores taxas de acertos, deixando o Sistema mais confiável com relação à classificação de padrões.

A redução da base de dados realmente possui a mesma validade da base completa, não interferindo nos momentos estatísticos que servem como entradas para os modelos inteligentes.

Fundindo todos estes estudos além dos objetivos específicos, pode-se concluir que o objetivo geral também foi atingido com êxito. A criação do Sistema *SupportSystemFlow* (SSF) que é um Sistema de Suporte à Decisão com um combinador (*ensemble*) de algoritmos inteligentes, seguro, prático, amigável e que suporta manipulações em tempo real (*online*).

Vale observar que a proposta principal apresentada por este trabalho é uma proposta inédita. A proposta de classificar aquisições de regimes vertical e horizontal em apenas uma modelagem, de forma conjunta e em tempo real.

Foram feitas buscas na literatura e todos os modelos de classificação propostos distinguem entre classificação vertical, classificação horizontal ou um intermediário regime inclinado. Porém todos são tratados de formas separadas nos modelos inteligentes, diferente deste trabalho; Com taxas de acertos entre 91% utilizando o *Ensemble*, taxas de acertos significativas. Concluindo que, os modelos ajustados em conjunto com o combinador de algoritmos inteligentes são eficientes.

Se esta modelagem sugerida for colocada em prática (em campo), as mudanças necessárias no regime de escoamento são mínimas ou até desnecessárias. Se os aparatos estiverem desregulados ou estiverem no mesmo intervalo de tensão; Será necessário regular o intervalo de tensão dos aparatos experimentais, que fazem as aquisições do escoamento, para os regimes diferentes. Frente a isto, os aparatos no regime vertical seriam regulados para o intervalo de aquisição de 0 a 1 Volt; E os aparatos no regime horizontal seriam regulados para realizarem aquisições no intervalo de 2 a 5 Volts.

Concluindo, o setor de escoamento é um setor em que falhas não são permitidas. Pois se as mesmas ocorrerem podem gerar grandes prejuízos e perdas. Frente a isto, todo Sistema que se propõe trabalhar neste setor deve obedecer às normas de segurança. O Sistema SSF proposto nesta monografia é seguro quanto à confiabilidade dos dados, possui segurança a invasões e as taxas de acertos não fogem das margens apresentadas nos resultados. Pois foram feitas buscas a fim de encontrar todos os tipos possíveis de erros. A abordagem híbrida para identificação de padrões de escoamento bifásico se apresentou eficiente e satisfatória, devido a altas taxas de acertos atingidas. Frente a isto, o trabalho realizado nesta monografia é um trabalho confiável de ser referenciado, reproduzido e utilizado.

7.4. Trabalhos Futuros

Trabalhos futuros que serão realizados são: publicações em eventos científicos e revistas do setor de escoamento e energia.

Sugestões para trabalhos futuros que o autor deste trabalho propõe são:

- Migrar o SSF para um servidor *web* e adaptá-lo, com a proposta de utilizá-lo em qualquer estação de escoamento que possua acesso a internet;
- Ampliar a variação de componentes inteligentes, criando um trabalho com apenas este objetivo específico. Frente a isto, seria agregado um maior valor ao módulo inteligente deste trabalho;
- Criação de um módulo *broadcasting* para o SSF, com o objetivo de visualizar o escoamento em tempo real e remotamente.

REFERÊNCIAS

BILOBROVEC, M.; MARÇAL, R. F.; PILATTI, L. A.. *Aplicações de redes neurais artificiais em simulações*. XI SIMPEP. Nov. 2004. Disponível em: <<http://www.pg.utfpr.edu.br/ppgep/Ebook/ARTIGOS/9.pdf>>. Acessado em: 11 Nov. 2011.

BIN, S.; HONGJIAN, Z.; LU, C.; YUXIAO, Z.. *Flow regime identification of gas-liquid two-phase flow based on hht*. Chinese Journal Chemical Engineering, 2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1004954106600335>>. Acessado em: 11 Nov. 2011.

BISPO, C. A. F.; CAZARINI, E. W.. *A Evolução do Processo Decisório*. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP1998_ART094.pdf>. Acessado em: 08 Nov. 2011.

BOLFARINE, H.; BUSSAB, W. O.. *Elementos de Amostragem*. ABE - Projeto Fisher. São Paulo: Editora *Edgard Blücher*, 1ed., 2005.

BOX, G. E. P.. *Robustness in the strategy of scientific model building*. In LAUNER, R. L.; WILKINSON, G. N.. *Robustness in Statistics*, New York: Academic Press, 1ed, 1979. p. 201 – 236.

CAI, S.; TORAL, H.; QIU, J.; ARCHER, J. S.. *Neural network based objective flow regime identification in air-water two-phase flow*. Canadian Journal of Chemical Engineering. 1994. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/cjce.5450720308/abstract>>. Acessado em: 11 Nov. 2011.

CASTRO, L. N.; ZUBEN, F.J.V.. *Optimised Training Techniques For Feedforward Neural Networks*. Technical Report. Faculdade de Engenharia Elétrica, Unicamp, 1998. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.4551&rep=rep1&type=pdf>>. Acessado em: 11 Nov. 2011.

CHANG, C.; LIN, C.. *LIBSVM: A Library For Support Vector Machines*. 2001. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>. Acessado em: 09 Nov. 2011.

COCHRAN, W.. *Sampling Techniques*. New York: John Wiley, 3ed., 1977.

CORRÊA, F. C. *Sistema Embarcado Para Medidas De Escoamento Bifásico Gás-líquido*. Dissertação (Mestrado em Engenharia Mecânica), Unicamp, Faculdade de Engenharia Mecânica, Campinas, 2009.

CORRÊA, F. C.; MASTELARI, N.; SALGADO, R. M.; ROSA, E. S.; OHISHI, T.; *Sistema Inteligente Para Monitoramento e Detecção de Padrões de Escoamento Bifásico Gás-Líquido*. Instituto Brasileiro de Petróleo, Gás e Biocombustíveis: IBP, 2009b.

CORTES, C., VAPNIK, V., *Support Vector Networks*. Machine Learning 20. pp. 273-297. 1995.

DAILIANG, X.; HAIFENG, J.; ZHIYAO, H.; HAIQING, L.. *An Online Flow Pattern Identification System For Gas-oil Two-phase Flow Using Electrical Capacitance Tomography*. Instrumentation and Measurement - Technology Conference. 2004. Disponível em: <<http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F19%2F35958%2F01703936.pdf&authDecision=-203>>. Acessado em: 11 Nov. 2011.

DING, H.; HUANG, Z.; SONG, Z.; YAN, Y.. *Hilbert-huang transform based signal analysis for the characterization of gas-liquid two-phase flow*. Chinese Journal Chemical Engineering, 2007. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0955598607000040>>. Acessado em: 11 Nov. 2011.

EUCALYPT Systems, Inc. *Gas Flow Rate Calculation*. Disponível em: <<http://www.softscout.com/software/Energy-and-Natural-Resources/Oil-and-Gas/Gas-Flow-Rate-Calculation.html>>. Acessado em: 30 Mai. 2011.

FRANCO, E. F.. *Classificação de Padrões para o Escoamento Gás/Líquido via Modelos Inteligentes*. Trabalho de Conclusão de Curso (Bacharelado de Ciência da Computação), UNIFAL-MG, Alfenas-MG, 2010.

G1.*Vazamento de petróleo desafia a tecnologia no Golfo do México*. 2010. Disponível em: <<http://g1.globo.com/mundo/noticia/2010/05/vazamento-de-petroleo-desafia-tecnologia-no-golfo-do-mexico.html>>. Acessado em: 11 Nov. 2011.

HANSEN, L. K.; SALAMON, P.. *Neural Network Ensembles*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990. Disponível em: <<http://machine-learning.martinsewell.com/ensembles/HansenSalamon1990.pdf>>. Acessado em: 11 Nov. 2011.

HEUSER, C. A. *Projeto de Banco de Dados*. Editora Sagra Luzzatto: Instituto de Informática da UFRGS, 2001, 4ed., p.130-140.

HEWITT, G. F.. *Handbook of Multiphase Systems*. Hemisphere Publishing Corporation, 1982.

ISHII, M.. *Thermo-fluid dynamic theory of two-phase flow*. Collection de la Direction des Etudes et Recherches d'Electricite de France, 1975.

LEE, J. Y.; PARANJAPE, S. S.; KIM, N. S.; ISHII, M.. *Instantaneous and objective flow regime identification method*. In: 5th International Conference on Multiphase Flow, 204 p. Yokohama, Japan, 2004.

LIMA, D. B.. *Modelo Inteligente de Previsão de Vazões Afluentes para o Auxílio à Tomada de Decisão no Setor Elétrico*. Trabalho de Conclusão de Curso (Bacharelado de Ciência da Computação), UNIFAL-MG, Alfenas-MG, 2010.

LORENA, A. C.; CARVALHO, A. C. P. L. F.. *Uma Introdução às Support Vector Machines*, *Revista de Informática Teórica e Aplicada*, Vol. 14, No 2, 2007.

MATSUI, G.. *Identification of flow regimes in vertical gas-liquid two-phase flow using differential pressure fluctuations*. *International Journal Multiphase Flow*. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0301932284900077>>. Acessado em: 11. Nov. 2011.

MI, Y.; ISHII, M.; TSOUKALAS, L. H.. *Vertical two-phase flow identification using advanced instrumentation and neural networks*. *Nuclear Engineering and Design*, 1998. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S002954939800212X>>. Acessado em: 11 Nov. 2011.

MI, Y.; TSOUKALAS, L. H.; ISHII, M.; LI, M.; XIAO, Z.. *Hybrid fuzzy-neural flow identification methodology*. In: Proc. of the 5th IEEE International Conference on Fuzzy Systems. New Orleans, 1996. Disponível em: <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=552370>. Acessado em: 11 Nov. 2011.

MISHIMA, K.; ISHII, M.. *Flow regime transition criteria for upward two-phase flow in vertical tubes*. *International Journal Heat Mass Transfer*: v. 27, n.5, 1984, p.723-737.
ORACLE. *NetBeans*. Disponível em:<<http://netbeans.org/>>. Acessado em: 08 Nov. 2011.

POWER, D. J.. *A Brief History of Decision Support Systems*. DSSResources.COM. version 2.8. 2003. Disponível em: <<http://dssresources.com/history/dsshistoryv28.html>>. Acessado em: 09 Nov. 2011.

QUORUM - BUSINESS SOLUTIONS, INC. *PGAS Flow Measurement - Software Solution*. Disponível em: <<http://www.qbsol.com/software/pgas-flow-measurement-software/>>. Acessado em: 30 Mai. 2011.

ROSA, E.; SALGADO, R.; OHISHI, T.; MASTELARI, N.. *Performance Comparison of Artificial Neural Networks and Expert Systems Applied to Flow Pattern Identification in Vertical Ascendant Gas-Liquid Flows*. International Journal of Multiphase Flow. 2010. Disponível em: <<http://www.citeulike.org/article/7166526>>. Acessado em: 11 Nov. 2011.

SALGADO, R. M. *Sistema de Suporte à Decisão para Análise e Previsão de Carga por Barramento*. Tese (Doutorado em Engenharia Elétrica), Unicamp, Faculdade de Engenharia Elétrica e Computação, Campinas, 2009.

SALGADO, R. M.; OHISHI, T.; ROSA, E. S.; MASTELARI, N.; OLIVEIRA, H. C.. *Algoritmos de Agrupamento e Sistema Especialistas Aplicados no Reconhecimento de Padrões de Escoamentos Gás/Líquido*. Simpósio Brasileiro de Pesquisa Operacional. 2010. Disponível em: <http://www.sobrapo.org.br/sbpo2010/xliisbpo_pdf/72342.pdf>. Acessado em: 11 Nov. 2011.

SELLI, M. F.; SELEGHIM, P.. *On-line identification of horizontal two-phase flow regimes through gabor transform and neural network processing*. In: 10th Congresso Brasileiro de Ciências Térmicas e Engenharia - ENCIT. Rio de Janeiro, 2004. Disponível em: <<http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=ASMECP002006042630000813000001&idtype=cvips&gifs=yes&ref=no>>. Acessado em: 11 Nov. 2011.

SEMOLINI, R.. *Support Vector Machines, Inferência Transdutiva e o Problema de Classificação*. Dissertação (Mestrado em Engenharia Elétrica e de Computação), Unicamp, Faculdade de Engenharia Elétrica e Computação, Campinas, 2002.

SMOLA, A. J.; SCHÖLKOPF, B.. *Sparse greedy matrix approximation for machine learning*, Citeseer. 2000.

SUNRISE. *PIPENET*. Disponível em: <<http://www.sunrise-sys.com/>>. Acessado em: 30 Mai. 2011.

TAITEL, Y.; DUKLER, A. E.; BARNEA, D.. *Modeling Flow Pattern Transitions for Steady Upward Gás-Liquid Flow in Vertical Tubes*. AIChE Journal , v.26, 3ed., 1980, p.345-354.

THOMPSON, S.K. *Sampling*. New York: John Wiley, 1ed., 1992.

TSOUKALAS, L. H.; ISHII, M.; MI, Y. A.. *Neuro-fuzzy methodology for impedance-based multiphase flow identification*. Engineering Application of Artificial Intelligence, 1997. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0952197697000377>>. Acessado em: 11 Nov. 2011.

TUTU, N.. *Pressure fluctuations and flow pattern recognition in vertical two phase gas-liquid flows*. International Journal Multiphase Flow, 1982. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0301932284900181>>. Acessado em: 11 Nov. 2011.

WALLIS, G. B.. *One-dimensional Two-phase Flow*. New York : McGraw-Hill, 1ed.,1969, p. 336-346.

WEINBERGER, K. Q.; BLITZER, J.; SAUL, L. K.. *Distance Metric Learning for Large Margin Nearest Neighbor Classification*. In NIPS. MIT Press. 2006. Disponível em:<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.5831>>. Acessado em: 11 Nov. 2011.

WU, H.; ZHOU, F.; WU, Y.. *Intelligent Identification System Of Flow Regime Of Oil-gas-water Multiphase Flow*. International Journal of Multiphase Flow. 2001. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0301932200000227>>. Acessado em: 11 Nov. 2011.

YAMASHITA, Y.. *Time delay neural networks for the classification of flow regimes*. Computers in Chemical Engineering, 1997. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098135497875291>>. Acessado em: 11 Nov. 2011.