



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Diego Henrique Emygdio Lázaro

Aplicação de Máquinas de Vetores de Suporte na Identificação de
Perfis de Alunos de acordo com Características da Teoria das
Inteligências Múltiplas

São José do Rio Preto

2016

Diego Henrique Emygdio Lázaro

Aplicação de Máquinas de Vetores de Suporte na Identificação de Perfis de Alunos de acordo com Características da Teoria das Inteligências Múltiplas

Dissertação elaborada junto ao Programa de Pós-Graduação em Ciência da Computação – Área de Concentração em Sistemas Distribuídos e Arquitetura de Computadores, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Norian Marranghello

São José do Rio Preto

2016

Lázaro, Diego Henrique Emygdio.

Aplicação de máquinas de vetores de suporte na identificação de perfis de alunos de acordo com características da teoria das inteligências múltiplas / Diego Henrique Emygdio Lázaro. -- São José do Rio Preto, 2016
73 f. : il., tabs.

Orientador: Norian Marranghello

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Computação - Matemática. 2. Máquinas de vetores de suporte.
3. Ensino a distância - Ensino auxiliado por computador. 4. Agrupamento por aptidão (Educação) 5. Inteligências múltiplas. 6. Aprendizado do computador.
I. Marranghello, Norian. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas.
III. Título.

CDU – 518.721

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Diego Henrique Emygdio Lázaro

Aplicação de Máquinas de Vetores de Suporte na Identificação de Perfis de Alunos de acordo com Características da Teoria das Inteligências Múltiplas

Dissertação elaborada junto ao Programa de Pós-Graduação em Ciência da Computação – Área de Concentração em Sistemas Distribuídos e Arquitetura de Computadores, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Comissão Examinadora

Prof. Dr. Norian Marranghello

Unesp- São José do Rio Preto

Orientador

Prof. Dr.^a Renata Spolon Lobato

Unesp- São José do Rio Preto

Prof. Dr. Henrique Dezani

Fatec- Rio Preto

São José do Rio Preto

31 de maio de 2016

Lista de Figuras

| | |
|---|-----------|
| Figura 1: Indução de classificador em aprendizado supervisionado (LORENA, 2007) | 16 |
| Figura 2: O hiperplano ótimo separando os dados com a máxima margem ρ, os vetores-suporte (support vectors) e uma distribuição dos dados no R^2 (SEMOLINI, 2002)..... | 20 |
| Figura 3: Análise da região de separação de um Hiperplano ótimo (SEMOLINI, 2002)... | 21 |
| Figura 4: Exemplos de valores e situações da variável de folga ξ (SEMOLINI, 2002) | 22 |
| Figura 5: Métodos Multiclasses | 24 |
| Figura 6: Diagramas de Treino (a) e Teste (b) | 27 |
| Figura 7: Algoritmo de Determinação do hiperplano ótimo no espaço de características, (Verdi, 2001)..... | 31 |
| Figura 8: Decomposição em problemas biclasse | 32 |
| Figura 9: Algoritmo de reconstrução Votação-distribuída (FÜRNKRANZ, 2002)..... | 36 |
| Figura 10: Alteração do algoritmo de reconstrução de Votação-distribuída (FÜRNKRANZ, 2002)..... | 37 |
| Figura 11: Matriz Confusão | 44 |
| Figura 12: Valores de Exatidão das Informações do Classificador após Treinamento | 44 |
| Figura 13: Principais fases do processo de conhecimento (Fayadd,1996)..... | 50 |
| Figura 14: Algoritmo K-means (BUSSAB, 1990) | 52 |
| Figura 15: Exemplificação Algoritmo K-means (BUSSAB, 1990) | 53 |
| Figura 16: Representação da classificação da Máquina Validação das Inteligências Múltiplas | 54 |
| Figura 17: Representação SVM para definição das inteligências predominantes em um indivíduo..... | 55 |

Lista de Tabelas

| | |
|--|-----------|
| Tabela 1: Valores críticos associados ao grau de confiança na amostra | 42 |
| Tabela 2: Leitura da Matriz Confusão na Horizontal | 44 |
| Tabela 3: Leitura da Matriz Confusão na Vertical..... | 44 |
| Tabela 4: Tipo de Respostas | 47 |
| Tabela 5: Coleta de Amostras..... | 57 |
| Tabela 6: Informações sobre espaço amostral | 58 |
| Tabela 7: Percentual de respondentes por curso | 59 |
| Tabela 8: Percentuais Inteligências por Curso | 60 |

Sumário

| | |
|---|-----------|
| Capítulo 1- Introdução..... | 10 |
| 1.1 Motivação..... | 10 |
| 1.2 Objetivos | 12 |
| 1.3 Estado da Arte | 12 |
| 1.4 Organização do Trabalho | 14 |
| Capítulo 2- Fundamentação Teórica | 15 |
| 2.1 Conceitos Básicos de Aprendizado de Máquina | 15 |
| 2.2 Máquinas de Vetores de Suporte..... | 17 |
| 2.2.1 Hiperplano ótimo para classes linearmente separáveis | 18 |
| 2.2.2 Hiperplano ótimo para classes não linearmente separáveis | 21 |
| 2.5 Métodos de Classificação Binária | 23 |
| 2.5.1 Problemas de Classificação de Dados | 25 |
| 2.6 Métodos de Votação Direta..... | 27 |
| 2.6.1 Método de Votação-direta com probabilidades | 27 |
| 2.7 Método de Votação-distribuída | 29 |
| 2.7.1 Método de Votação-distribuída com probabilidades..... | 29 |
| 2.8 Treinamento de SVMs Não Lineares | 30 |
| 2.9 SVMs para Várias Classes..... | 31 |
| 2.10 Balanceamento de Classes | 34 |
| 2.11 Método de Agrupamento | 36 |
| 2.12 Sistema de Recomendação..... | 39 |
| 2.13 Teoria das Inteligências Múltiplas | 41 |
| 2.14 Amostragem Aleatória Simples | 42 |
| 2.15 Matriz Confusão | 43 |
| 2.15.1 Exatidão das informações para classificação | 45 |
| Capítulo 3- Metodologia | 47 |
| 3.1 Aplicação da Teoria das Inteligências Múltiplas para identificação de Perfis..... | 47 |
| 3.2 Definição da Ferramenta de Desenvolvimento para o Mecanismo de Classificação | 48 |
| 3.2.1 Técnicas Utilizadas pelo WEKA | 48 |

| | |
|--|----|
| 3.2.2 Biblioteca de Implementação - LibSVM | 49 |
| 3.3 Definição do Modelo Classificador para Sistemas de Recomendação..... | 49 |
| 3.4 Filtro Resample aplicado em Balanceamento de Classes | 50 |
| 3.5 Aplicação Algoritmo SimpleKMeans para Agrupamento de Classes | 51 |
| 3.6 Aplicação de Métodos de Decomposição de problemas multiclasse..... | 53 |
| Capítulo 4- Experimentos..... | 56 |
| 4.1 Descrição do experimento..... | 56 |
| 4.1.1 Análise Estatística..... | 58 |
| 4.2 Análise dos Resultados | 59 |
| Conclusão | 61 |
| Referências..... | 63 |
| APÊNDICE A – Questionário | 67 |
| APÊNDICE B – Questionário Gardner | 71 |

Resumo

Nesta dissertação foi desenvolvido um mecanismo de classificação capaz de identificar o perfil de um aluno de acordo com características da teoria das inteligências múltiplas, baseado em *Support Vector Machines* (SVMs, sigla em inglês para Máquinas de Vetores de Suporte), métodos de agrupamento e balanceamento de classes. O objetivo dessa classificação consiste em permitir que os tutores responsáveis por gerar o material para aulas em ferramentas de apoio ao ensino à distância possam utilizar este método de classificação para direcionar o conteúdo ao aluno de forma a explorar sua inteligência múltipla predominante. Para realização dos experimentos, duas SVMs foram criadas, utilizando o método de classificação baseado em k problemas binários, que reduzem o problema de múltiplas classes a um conjunto de problemas binários. Os resultados obtidos durante as fases de treino e teste das SVMs foram apresentados em percentuais por meio de um algoritmo de agrupamento particionado. Esses percentuais ajudam a interpretar a classificação do perfil de acordo com as inteligências predominantes. Além disso, com o uso de métodos de balanceamento de classes, obteve-se melhora no desempenho do classificador, assim, aumentando a eficácia do mecanismo, pois, suas taxas de incorreções foram baixas.

Palavras-Chave: Máquinas de Vetores de Suporte. Métodos de Classificação. Método de Agrupamento Particionado. Balanceamento de Classes. Aprendizado de Máquina.

Abstract

In this work, it was developed a mechanism in order to classify students' profiles according to the Theory of Multiple Intelligences, based on Support Vector Machines (SVMs), cluster methods and classes balancing. By using these classifications, tutors, who prepare materials for classes in specific tools for distance education purposes, are able to suggest contents for students so that they are able to explore their predominant multiple intelligence. To perform these experiments, SVMs were created by using classification methods based on binary problems that reduce multiple classes problems into a set of binary problems. The results generated during the training and the SVM test stages were presented in percentages by using partitioning clustering algorithm. These percentages are helpful for analysis of profiles classifications according to multiple intelligences. Besides that, by using classes balancing methods, it was possible to obtain improvements on the classifier performance and, consequently, the mechanism efficiency was increased as well, considering the fact that inaccuracy rates were low.

Keywords: Support Vector Machines. Classification Methods. Partitioning Clustering Method. Classes Balancing. Machine Learning.

Capítulo 1- Introdução

Ao analisarmos a situação atual da prática educativa em nossas escolas, identificaremos problemas como: a grande ênfase dada à memorização, pouca preocupação com o desenvolvimento de habilidades para reflexão crítica e a autocrítica do conhecimento que o aluno aprende. As ações educativas centradas nos professores que determinam o que ensinar e como deve ser aprendido interferem diretamente na formação do aluno e atrapalha seu crescimento profissional e pessoal.

O mercado de trabalho necessita de profissionais qualificados para atender à demanda, gerando, assim, vagas de emprego que, em alguns casos, não são preenchidas devido à queda na qualidade de ensino.

Um dos principais fatores que influenciam na qualidade da formação profissional é a dificuldade em frequentar cursos de média e longa duração. Muitas pessoas não conseguem conciliar o trabalho, a família e entre outros compromissos para buscar uma capacitação profissional.

Diante disso, alternativas como cursos à distância têm se tornado opções para a qualificação profissional. Assim, ferramentas de apoio ao ensino que permitem ao aluno poder estudar em horários que se adequem às suas rotinas, possibilitam a continuidade de sua formação profissional.

No entanto, mesmo com o uso das ferramentas de apoio ao ensino, podem surgir outros motivos para reduzir o interesse do aluno, por exemplo, a dificuldade em aprender sem a presença física do professor para direcionar as aulas de forma a explorar as características que favoreçam o processo ensino aprendizagem. Dessa forma, o aluno precisará se adaptar à metodologia de ensino da ferramenta. Essa adaptação pode não acontecer, levando à evasão do curso.

1.1 Motivação

A área de reconhecimento de padrões consiste na classificação de diversos exemplos existentes em uma determinada base de dados como pertencentes a um tipo específico de padrão

dentre os possíveis padrões que essa base possa representar. Vários métodos são empregados nessa área, no entanto, nos últimos anos, um método vem se destacando entre os demais: *Support Vector Machine* (SVM, sigla em inglês). As SVMs são máquinas de aprendizagem que se baseiam na Teoria da Aprendizagem Estatística, treinadas por meio de um algoritmo supervisionado. Elas foram propostas em 1992. Desde então são empregadas em diversos setores, como por exemplo, medicina, biomedicina, entre outros.

A proposta deste trabalho é o desenvolvimento de um mecanismo de classificação que pode ser aplicado em um sistema de recomendação computacional [33] e, assim, ser utilizado por ferramentas de apoio ao ensino à distância como método de classificação de alunos para direcioná-los a uma forma de estudo que possa favorecer o ensino individualizado, explorando as principais características de inteligência de cada aluno.

O Ambiente de apoio ao ensino, chamado *Classroom* criado por Costa Neto (2012), dá ênfase na Teoria das Inteligências Múltiplas (TIM) para fornecer ferramentas e guias de criação para aulas virtuais, facilitando a composição e a exposição de complementos para aulas presenciais [6].

Thays Santos, *et al.* (2015) com o trabalho intitulado de “Perspectivas da Teoria das Inteligências Múltiplas para a Educação a Distância” aborda uma pesquisa com base na TIM e as implicações de sua trajetória no desenvolvimento aplicado à Educação à Distância, tendo como objetivo entender como as competências intelectuais estão sendo estimuladas e combinadas no contexto de EaD, produzindo, assim, outras formas de ensino aprendizagem capazes de fomentar diferentes habilidades [8].

Outro trabalho correlato é o de Roberta Barbosa (2012) desenvolveu um método computacional capaz de auxiliar os médicos dermatologistas no diagnóstico de lesões de pele por meio de imagens digitais. Com este método pretende-se auxiliar o dermatologista a classificar as lesões de pele utilizando a regra ABCD (Assimetria, Borda, Cor e Diâmetro) e análise de textura, bem como identificar as lesões: nevos melanocíticos, ceratose seborréica e melanoma [30].

Rafaela Andreola, *et al.* (2010), com o trabalho intitulado de “Classificação de Imagens Hiperespectrais Empregando *Support Vector Machines*”, fala de um estudo que investiga o desempenho do classificador SVM na classificação de imagens em alta dimensionalidade. Como SVM opera em um par de classes a cada vez, propõe-se aqui a sua implementação em uma estrutura em forma de árvore binária, onde somente duas classes são tratadas em cada nó. A precisão da imagem temática produzida por este esquema de classificação é avaliada para duas funções *kernel* distintas e em função do valor para dimensionalidade dos dados [2].

Assim, como nesses exemplos, pôde-se notar o uso de SVM para problemas de classificação em áreas distintas como diagnósticos médicos, reconhecimento de padrões, problemas de classificação e recuperação de imagem.

1.2 Objetivos

Esse trabalho tem como objetivo:

- Utilizar SVMs para desenvolver um mecanismo de classificação de perfis de acordo com a Teoria das Inteligências Múltiplas, que aplique métodos de agrupamento e balanceamento de classes.

1.3 Estado da Arte

A teoria das Inteligências Múltiplas foi desenvolvida por Howard Gardner, psicólogo da Universidade de Harvard na década de 70. Após anos de estudos, publicou seu primeiro livro, “Estruturas da Mente”, no ano de 1983, no qual gerou um grande impacto na área da educação, pois, acreditava-se que, por meio da Teoria das Inteligências Múltiplas, muitos problemas no campo educacional poderiam ser resolvidos.

A teoria de Gardner [14, 15] afirma que todos os indivíduos normais dispõem de oito inteligências sendo elas: linguística, lógico-matemática, espacial, corporal-cenestésica, musical, interpessoal, intrapessoal e naturalista, e que cada indivíduo tem maior habilidade para desenvolver uma determinada inteligência. Essa teoria possibilita trabalhar com estímulos específicos e com isso desenvolver adequadamente determinada habilidade de cada um.

Segundo Armstrong [1], Gardner ofereceu um meio de mapear a ampla gama de capacidades dos seres humanos em oito categorias ou “inteligências”. Porém, inicialmente elas foram mapeadas como apenas sete e, segundo Armstrong [1], a Teoria das Inteligências Múltiplas abre as portas para ampla variedade de estratégias de ensino que podem ser implementadas na sala de aula. Há décadas essas estratégias já têm sido utilizadas por alguns professores, pois as mesmas oferecem a oportunidade de desenvolver estratégias de ensino inovadoras.

A Teoria das Inteligências Múltiplas alerta (cuidado com afirmações categóricas) que não existe um conjunto de estratégias de ensino que funciona melhor para todos os alunos [1]. Cada

indivíduo tem inclinações diferentes para as oito inteligências de modo que qualquer estratégia específica provavelmente será muito bem-sucedida com um grupo de alunos e não tão bem-sucedida com outros grupos. Por exemplo, professores que usam figuras e imagens no ensino atingirão alunos com uma orientação mais espacial, mas terão efeitos diferentes sobre aqueles com inclinações mais físicas ou verbais.

Devido a essas peculiaridades entre alunos, os professores devem usar ampla variedade de estratégias de ensino. Na medida em que o professor mudar a inteligência enfatizada, sempre haverá um momento em que o aluno fará uso de sua inteligência mais desenvolvida. Portanto, entende-se que para cada inteligência que se pretende avaliar ou desenvolver, uma estratégia diferente deverá ser utilizada, abrindo margem para mais uma linha de estudos que consistem em entender essas técnicas e aplicá-las aos alunos analisando os resultados.

Neste contexto, está inserida a proposta desta dissertação (deste trabalho), criando um mecanismo que seja capaz de identificar as inteligências e traçar o perfil de um aluno, assim, fornecendo aos professores subsídios que favoreçam no processo de ensino aprendizagem.

Para isso, optou-se por utilizar SVM como técnica de classificação devido a um conjunto de características:

- *Boa capacidade de generalização*: os classificadores gerados por uma SVM em geral alcançam bons resultados de generalização. A capacidade de generalização de um classificador é medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado em seu treinamento. Na geração de prognosticadores por SVMs, portanto, é evitado o superajustamento (*overfitting*), situação na qual o prognosticador se torna muito especializado no conjunto de treinamento, obtendo baixo desempenho quando confrontado com novos padrões [40].
- *Robustez em grandes dimensões*: as SVMs são robustas diante de objetos de grandes dimensões, como, por exemplo, imagens. Comumente há a ocorrência de *overfitting* nos classificadores gerados por outros métodos inteligentes sobre esses tipos de dados [40].
- *Convexidade da função objetivo*: a aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo global. Esta é uma vantagem sobre, por exemplo, as Redes Neurais Artificiais, em que há a presença de mínimos locais na função objetivo a ser minimizada [40].
- *Teoria bem definida*: as SVMs possuem uma base teórica bem estabelecida dentro da Matemática e Estatística [40].

Entre as características citadas, o destaque das SVMs está em sua capacidade de generalização. Estes resultados foram apresentados por Vapnik e Chervonenkis (1998), através da Teoria de Aprendizado Estatístico, proposta por estes autores na década de 60 e 70 [44].

1.4 Organização do Trabalho

Esta dissertação está dividida em quatro capítulos distribuídos da seguinte maneira:

Capítulo 1: Neste capítulo é apresentado o estado da arte, bem como, objetivos e motivação do trabalho.

Capítulo 2: Neste capítulo são abordados os conceitos utilizados para desenvolvimento do projeto, entre eles SVMs e sua teoria, Métodos de Classificação, Aprendizado de Máquina, Métodos de Votação, Balanceamento e Agrupamento de Classes, entre outros.

Capítulo 3: Neste capítulo são detalhados o uso dos conceitos do capítulo 2, de maneira que possa se desenvolver o mecanismo de classificação.

Capítulo 4: Neste capítulo são apresentados os experimentos e os detalhes de sua realização. Além disso, são apresentadas análises dos resultados.

Por fim, é apresentada a conclusão do projeto, bem como, sugestões para projetos futuros.

Capítulo 2- Fundamentação Teórica

Neste capítulo são abordados os conceitos utilizados para desenvolvimento do projeto, entre eles SVMs e sua teoria, Métodos de Classificação, Aprendizado de Máquina, Métodos de Votação, Balanceamento e Agrupamento de Classes, entre outros.

2.1 Conceitos Básicos de Aprendizado de Máquina

As técnicas de Aprendizado de Máquina (AM) utilizam princípio de inferência denominado indução. A partir disso, obtêm-se conclusões genéricas de um conjunto particular de exemplos. Por se tratar de um aprendizado indutivo, pode ser dividido em dois tipos principais: supervisionado e não supervisionado [21] e [26].

No aprendizado supervisionado tem-se a figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada e saída desejada. O algoritmo de AM extrai a representação do conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente [28].

No aprendizado não-supervisionado não há a presença de um professor, ou seja, não existem exemplos rotulados. O algoritmo de AM aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade. Essas técnicas são utilizadas principalmente quando o objetivo for encontrar padrões ou tendências que auxiliem no entendimento dos dados [41].

O tipo de aprendizado abordado neste trabalho é o supervisionado. Neste caso, dado conjunto de exemplos rotulados na fórmula (x_i, y_i) em que x_i representa um exemplo em y_i , denotando o seu rótulo. Deve-se produzir um classificador, também denominado modelo, prognosticador ou hipótese, capaz de prever precisamente o rótulo dos novos dados. Esse processo de indução de um classificador a partir de uma amostra de dados é denominado treinamento [21] e [26]. O classificador obtido também pode ser visto como uma função f , a qual recebe um dado x e fornece uma predição y .

Os rótulos ou classes representam o fenômeno de interesse sobre o qual se deseja fazer previsões. Nesta dissertação, considera-se o caso em que os rótulos assumem valores discretos $1, \dots, k$. Tem-se, então, um problema de classificação. Um problema de classificação no qual $k = 2$ é denominado binário. Para $k > 2$, configura-se, portanto, um problema multiclases.

Cada exemplo, também referenciado por dado ou caso, é tipicamente representado por um vetor de características. Cada característica, também denominada atributo, expressa um determinado aspecto do exemplo [27]. Normalmente, há dois tipos básicos de atributos: nominal e contínuo. Um atributo é definido como nominal (ou categórico) quando não existe uma ordem entre os valores que ele pode assumir (por exemplo, entre cores). No caso de atributos contínuos, é possível definir uma ordem linear nos valores assumidos.

Um requisito importante para as técnicas de AM é que elas sejam capazes de lidar com dados imperfeitos, denominados ruídos [26]. Muitos conjuntos de dados apresentam esse tipo de caso, sendo alguns erros comuns a presença de dados com rótulos e/ou atributos incorretos. A técnica de AM deve idealmente ser robusta a ruídos presentes nos dados, procurando não fixar a obtenção dos classificadores sobre esse tipo de caso [27]. Deve-se também minimizar a influência de ruídos no processo de indução. A existência de ruídos implica, tipicamente, em prejuízos à interpretação dos resultados dos testes estatísticos aplicados às amostras.

Os conceitos referentes à geração de um classificador a partir do aprendizado supervisionado são representados de forma simplificada na Figura 1. Tem-se nessa figura um conjunto com n dados. Cada dado x_i possui m atributos, ou seja, $x_i = (x_{i1}, \dots, x_{im})$. As variáveis y_i representam as classes. A partir dos exemplos e das suas respectivas classes, o algoritmo de AM extrai um classificador. Pode-se considerar que o modelo gerado fornece uma descrição compacta dos dados fornecidos [7].

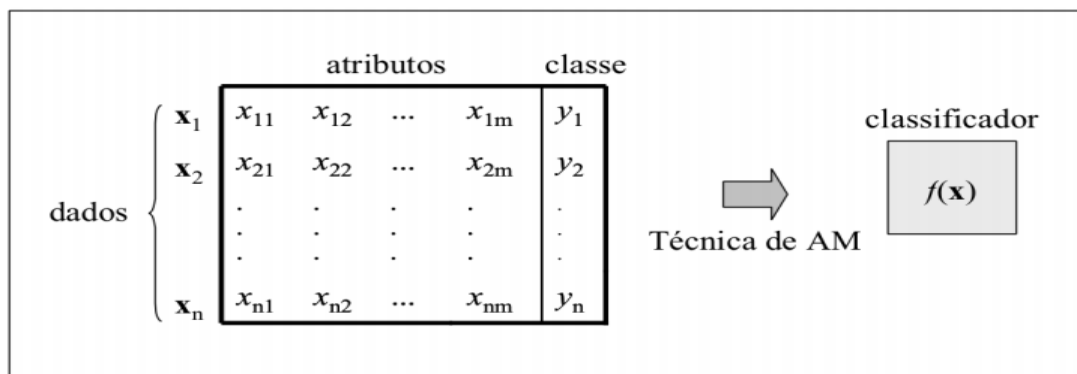


Figura 1: Indução de classificador em aprendizado supervisionado (LORENA, 2007)

A obtenção de um classificador por um algoritmo de AM a partir de uma amostra de dados também pode ser considerada um processo de busca [26]. Procura-se, entre todas as hipóteses que o algoritmo seja capaz de gerar a partir dos dados, aquela com melhor capacidade de descrever o domínio em que ocorre o aprendizado.

Para estimar a taxa de predições corretas ou incorretas (também denominadas taxa de acerto e taxa de erro, respectivamente) obtidas por um classificador sobre novos dados, o conjunto de exemplos é, em geral, dividido em dois subconjuntos disjuntos: de treinamento e de teste [26]. O subconjunto de treinamento é utilizado no aprendizado do conceito e o subconjunto de teste é utilizado para medir o grau de efetividade do conceito aprendido na predição da classe de novos dados.

Um conceito comumente empregado em AM é o de generalização de um classificador, definida como a sua capacidade de prever corretamente a classe de novos dados. No caso em que o modelo se especializa nos dados utilizados em seu treinamento, apresentando uma baixa taxa de acerto quando confrontado com novos dados, tem-se a ocorrência de um superajustamento (*overfitting*) [26] e [7]. É também possível induzir hipóteses que apresentem uma baixa taxa de acerto mesmo no subconjunto de treinamento, configurando uma condição de subajustamento (*underfitting*). Essa situação pode ocorrer, por exemplo, quando os exemplos de treinamento disponíveis são pouco representativos ou quando o modelo obtido é muito simples [7].

2.2 Máquinas de Vetores de Suporte

As SVMs utilizam métodos de classificação supervisionada baseados em particionamento, cujo objetivo é construir hiperplanos para separar pontos de duas classes.

Essa técnica já foi utilizada com sucesso em diversas aplicações de reconhecimento de padrões, como por exemplo: categorização de textos, categorização de SPAM, reconhecimento de caracteres manuscritos, reconhecimento de textura, análise de expressões de genes, entre outras.

O funcionamento das SVMs é definido da seguinte maneira: dadas duas classes e um conjunto de pontos que pertencem a essas classes, as SVMs determinam o hiperplano que os separa, de forma a colocar maior quantidade possível de pontos da mesma classe do mesmo lado ao mesmo tempo que a distância de cada classe a esse hiperplano é maximizada.

Normalmente, as máquinas de vetor suporte operam em um espaço de dimensão maior que a dimensão dos dados originais, no chamado espaço de características. É nesse espaço que se consegue a maximização desejada para obter uma melhor capacidade de generalização de classificação.

Para conseguir essa maximização, é necessária uma formulação para resolver problemas de otimização quadrática com restrições.

A classificação tem como objetivo, a partir de um banco de dados contendo objetos pré-classificados (objetos cuja classe é conhecida), construir um modelo que seja capaz de classificar automaticamente novos objetos (cuja classe é desconhecida) em função de suas características. O modelo criado é chamado de modelo classificador.

O uso de SVMs é capaz de resolver problemas de classificação de dados, gerando classificadores que apresentam bons resultados. Porém, esses classificadores possuem uma limitação de interpretação, não sendo possível compreender a saída obtida. Por isso, nesse trabalho, este modelo utiliza técnicas de agrupamento e balanceamento de dados durante a classificação realizada pela SVM.

2.2.1 Hiperplano ótimo para classes linearmente separáveis

Quando o aprendizado supervisionado é aplicado ao problema de classificação, as amostras de treinamento são formadas pelo conjunto de dados de entrada associados às suas correspondentes respostas pré-classificadas (rótulos ou dados de saída). Após o treinamento, o objetivo é classificar novas amostras, ainda não rotuladas [45].

Considerando o seguinte conjunto de dados de treinamento:

$$(x_i, y_i) \quad 1 \leq i \leq N, x \in R^m, y \in \{+1, -1\} \quad (1)$$

Onde x_i , é o dado de entrada para a amostra i e y_i é a resposta desejada.

Classificações binárias são frequentemente realizadas pelo uso de funções $g: X \subseteq R^m \rightarrow R$ com a seguinte estratégia: as amostras são designadas para a classe positiva, se $\square(x) \geq 0$, e caso contrário, para classe negativa.

Será considerado nesta seção que as classes representadas pelos rótulos $y_i = +1$ e -1 são linearmente separáveis. A superfície de decisão será representada por um hiperplano na forma [45]:

$$\mathbf{g}(x) = (w^t x) + b = 0 \quad (2)$$

Onde $w \in R^m$ é o vetor de pesos, e $b \in R$ é o intercepto.

Assim podemos aplicar a seguinte estratégia de decisão:

$$\begin{aligned} (w^t x) + b &\geq 0 \text{ para } y = +1; \\ (w^t x) + b &< 0 \text{ para } y = -1. \end{aligned} \quad (3)$$

Para descrever o lugar geométrico dos hiperplanos separadores, será utilizada a seguinte forma canônica (onde o vetor w e o escalar b são novamente escalados de tal maneira a atender as desigualdades):

$$\begin{aligned} (w^t x) + b &\geq +1 \text{ para } y = +1; \\ (w^t x) + b &\leq -1 \text{ para } y_i = -1. \end{aligned} \quad (4)$$

A seguir, é apresentada a notação compacta para as desigualdades (4):

$$y[(w^t x) + b] \geq 1. \quad (5)$$

Para um dado vetor de pesos w e intercepto b , a separação entre o hiperplano $\square(x) = (w^t x) + b = 0$ e o dado de entrada mais perto é chamada de *margem de separação* denotada por ρ . Sempre que for possível obter um $\rho > 0$, existirão infinitos hiperplanos, dentre os quais se busca um hiperplano particular em que a margem de separação ρ é maximizada. De acordo com essa condição, a superfície de decisão é dita ser o hiperplano ótimo e a técnica de aprendizado de máquina utilizado para a determinação desse hiperplano é denominada *Support Vector Machines* (SVM), sendo que os dados de treinamento que se encontram à distância ρ do hiperplano são chamados vetores-suporte [45].

Na Figura 2, é apresentada uma visão geométrica da construção do hiperplano ótimo para um espaço bidimensional, além da interpretação dos vetores-suporte.

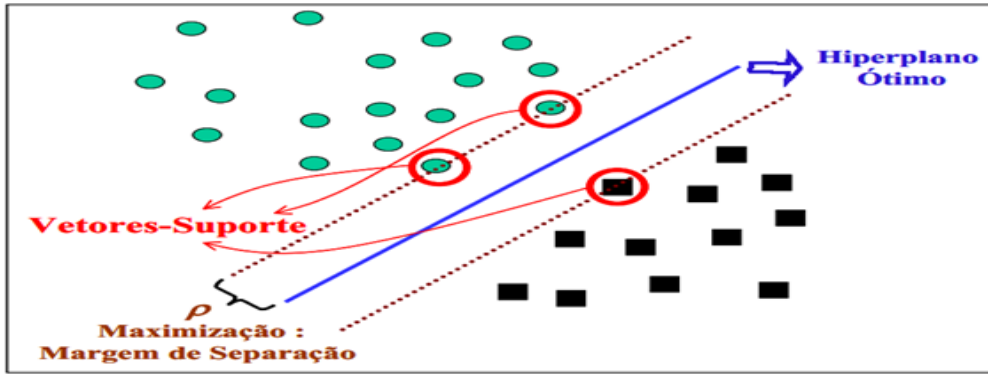


Figura 2: O hiperplano ótimo separando os dados com a máxima margem ρ , os vetores-suporte (support vectors) e uma distribuição dos dados no R^2 (SEMOLINI, 2002).

Os dados para os quais o resultado da equação (5) é igual a 1 são os vetores-suporte, pois são aqueles que se encontram à distância ρ do hiperplano ótimo.

Os vetores-suporte exercem um papel importante nas operações desse tipo de aprendizagem de máquina. Em termos conceituais, eles são os pontos que se encontram mais perto da superfície de decisão e, portanto, são os de classificação mais difícil. Como tal, eles têm uma relação direta com a localização da superfície de decisão.

Considerando a forma canônica (4) com a seguinte modificação (hiperplanos com a margem igual a 1 são conhecidos como hiperplanos canônicos), temos [45]:

$$(w^T x^{sv+}) + b = +1, \text{ sendo } x^{sv+} \text{ um vetor-suporte pertencente à classe } y = +1; \quad (6)$$

$$(w^T x^{sv-}) + b = -1, \text{ sendo } x^{sv-} \text{ um vetor-suporte pertencente à classe } y = -1;$$

Para calcular a distância algébrica dos vetores-suporte para o hiperplano ótimo, ou seja, o valor da margem ρ , é preciso primeiro normalizar o vetor de pesos w , e usando a equação do hiperplano canônico (6), temos:

$$\rho = \frac{1}{2} \left[\left(\frac{w}{\|w\|} \right)^T x^{sv+} - \left(\frac{w}{\|w\|} \right)^T x^{sv-} \right] = \frac{1}{\|w\|} \quad (7)$$

A equação (7) mostra que maximizar a margem de separação entre as classes é equivalente a minimizar a norma euclidiana do vetor de pesos w .

Em resumo, o hiperplano ótimo definido pela equação (5) apresenta um vetor de pesos w que leva à máxima separação entre as amostras positivas e negativas. Essa condição ótima é alcançada minimizando a norma euclidiana do vetor de pesos w .

2.2.2 Hiperplano ótimo para classes não linearmente separáveis

Considerando o caso mais difícil de classificação, quando as classes não são linearmente separáveis e dadas as amostras de treinamento, não é possível construir um hiperplano separador sem encontrar erros de classificação. Todavia, é possível encontrar um hiperplano que minimiza a probabilidade do erro de classificação junto às amostras de treinamento, tendo esse contexto sido aplicado nesse trabalho.

Sendo assim, a margem de separação entre as classes é dita flexível (*soft*), pois existirão pontos $(x_i, y_i) 1 \leq i \leq N$ que violam a inequação 5.

Esta violação pode acontecer de três maneiras, conforme apresentadas na figura 3:

- O ponto (x_i, y_i) se encontra dentro da região de separação, porém, no lado correto da superfície de decisão, (Figura 3a). Nesse caso, as classes são linearmente separáveis, porém, houve uma escolha incorreta do hiperplano.
- O ponto (x_i, y_i) se encontra no lado incorreto da superfície de decisão, porém, dentro da região de separação (Figura 3b). Também nesse caso as classes são linearmente separáveis, porém houve uma escolha por um hiperplano de maior margem.
- O ponto (x_i, y_i) se encontra no lado incorreto da superfície de decisão e fora da região de separação (Figura 3c).

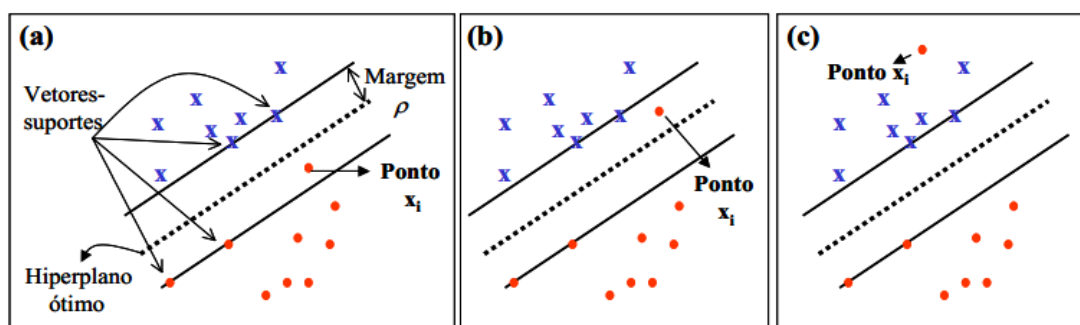


Figura 3: Análise da região de separação de um Hiperplano ótimo (SEMOLINI, 2002)

Para tratar do problema de classes não linearmente separáveis, introduziremos uma nova variável não-negativa, $\{\xi_i\} 1 \leq i \leq N$, na definição de hiperplano separador (superfície de decisão) apresentada a seguir [29]:

$$y_i[(w^T x_i) + b] \geq 1 - \xi_i, i = 1, \dots, N \quad (8)$$

Os escalares ξ_i são chamados de variáveis de folga, e medem os desvios dos pontos $(x_i, y_i) 1 \leq i \leq N$ para a condição ideal de separação das classes. Para $0 \leq \xi_i \leq 1$, o ponto se encontra dentro da região de separação, mas do lado correto da superfície de decisão. Para $\xi_i > 1$, o ponto se encontra no lado incorreto do hiperplano separador. Os vetores suportes são os pontos em que o resultado da equação (8) é igual a $1 - \xi_i$, mesmo que $\xi_i > 0$, conforme apresentado na figura 4.

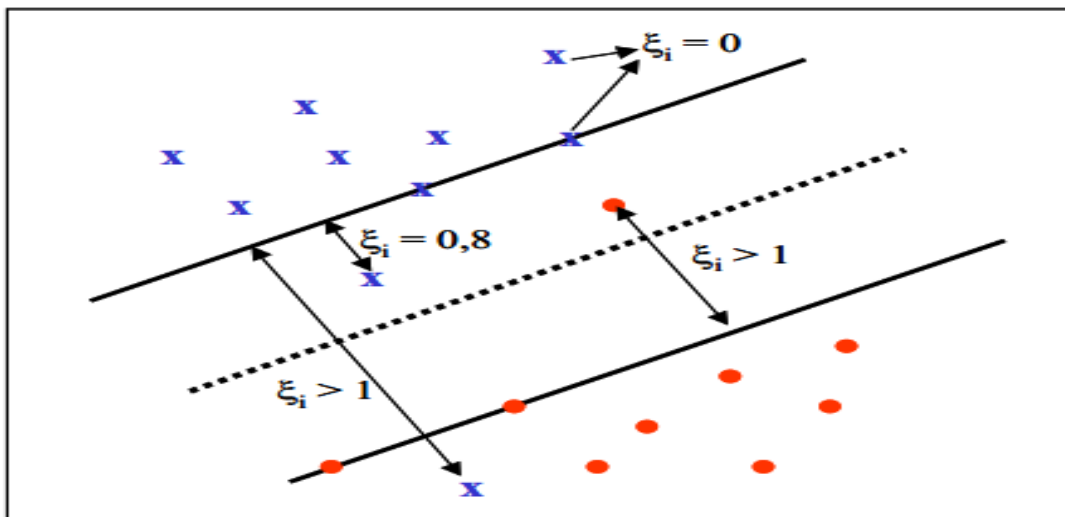


Figura 4: Exemplos de valores e situações da variável de folga ξ (SEMOLINI, 2002)

Note que, retirando do conjunto de treinamento, uma amostra em que $\xi_i > 0$, a superfície de decisão tem grande chance de mudar, mas retirando uma amostra em que $\xi_i = 0$ e o resultado da equação (8) seja maior do que 1, a superfície de decisão permanecerá sendo a mesma.

O objetivo é encontrar o hiperplano separador em que o erro de classificação incorreta, baseado no conjunto de treinamento, é minimizado. Podemos fazer isso minimizando a função

$$\theta(\xi) = \sum_{i=1}^N I(\xi_i - 1) \quad (9)$$

em relação ao vetor de pesos w , considerando a equação do hiperplano separador (8) e mais a seguinte restrição:

$$(w^T w) \leq \rho^{-1} = A_k \quad (10)$$

Essa restrição é a condição para que os parâmetros w e b , que definem o hiperplano, minimizem o número de erros no conjunto de treinamento sob a condição que eles pertençam ao subconjunto de elementos da estrutura $S_k = \{(w^T x) + b: (w^T w) \leq A_k\}$ determinados pela constante A_k [45].

A função indicadora $I(\xi_i - 1)$ é definida por [17]:

$$I(\xi_i - 1) = \begin{cases} 0 & \text{se } (\xi_i - 1) \leq 0; \\ 1 & \text{se } (\xi_i - 1) > 0. \end{cases} \quad (11)$$

A minimização de $\theta(\xi)$ em relação a w é um problema de otimização não convexo da classe NP-completo (não determinístico em tempo polinomial). Ou seja, um problema é dito "polinomial", ou pertencente à classe P, se existe um algoritmo conhecido capaz de solucionar o mesmo cuja ordem de complexidade (no pior caso) seja polinomial em relação ao "tamanho" do problema. Caso não se conheça tal algoritmo, o problema não pertence a essa classe.

Para fazer esse problema de otimização matematicamente tratável, aproximamos a função $\theta(\xi)$ por [17]:

$$\theta(\xi) = \sum_{i=1}^N \xi_i \quad (12)$$

Restrito ao hiperplano separador (8) e à restrição (10).

Chamamos o hiperplano construído com base na solução desse problema de *otimização de hiperplano ótimo generalizado* ou, por *simplificação, de hiperplano ótimo* [17].

2.5 Métodos de Classificação Binária

Classificação Binária: As SVMs não lineares projetam os dados de entrada em um espaço de dimensão maior na qual os dados passam a ser linearmente separados. Para tanto, utiliza-se o teorema de Cover [36], em que se afirmar que um espaço de entrada com padrões não linearmente separáveis pode ser transformado em um novo espaço de características em que os padrões são

linearmente separáveis, desde que sejam satisfeitas as seguintes condições: a transformação seja não linear e a dimensão do espaço seja suficientemente grande. Dessa forma, torna-se possível construir um hiperplano ótimo separando as classes nesse espaço de características.

Uma segunda formulação para o problema das SVMs não lineares surge da aplicação do teorema de Mercer [36] para encontrar uma função chamada Kernel, que minimiza o custo computacional da avaliação das SVMs. A Função Kernel de maneira geral faria referência à forma do hiperplano. Neste trabalho, será usado o seguinte polinomial, dado pela expressão [23,31,33]:

$$K(x_i, x_j) = ((x_i \cdot x_j) + a)^p$$

Onde a e p são parâmetros.

Classificação Multiclasses: As SVMs foram originalmente desenvolvidas para classificação binária. No caso de classificação em k classes, $k > 2$, existem duas abordagens básicas (Figura 5). A primeira é a redução do problema de múltiplas classes a um conjunto de problemas binários. Dois métodos usam essa abordagem: "decomposição um por classe" (*one-against-all*), separação das classes duas a duas (*one-against-one*). A segunda abordagem é a generalização de SVMs binárias para mais de duas classes. O método que utiliza essa abordagem é o método de Crammer e Singer.

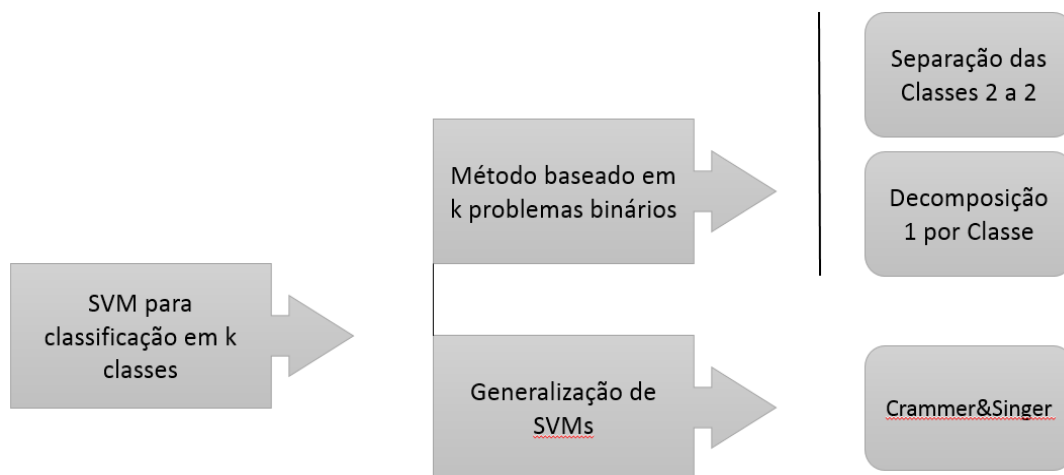


Figura 5: Métodos Multiclasses

Método Decomposição um por Classe: este método baseia-se na construção de k SVMs e classificação binária para separar uma classe de todas as outras. Em seguida, os resultados de todas as SVMs são agrupados, sendo feita a classificação desejada nas k classes.

Método Decomposição duas a duas: este método usa uma SVM binária para distinguir cada par de classes. Assim, são construídas $k(k - 1)/2$ SVMs. A classificação final é obtida a partir do resultado de todas as SVMs.

Método de Crammer e Singer: este método usa uma maneira mais natural de resolver o problema de classificação em k classes, $k > 2$, que é construir uma função de decisão considerando todas as classes de uma vez. Nesse método, todos os exemplos de treinamento são usados ao mesmo tempo.

Para este projeto, foram utilizados os métodos baseados em k problemas binários, ou seja, reduzindo o problema de múltiplas classes a um conjunto de problemas binários.

2.5.1 Problemas de Classificação de Dados

Nos últimos anos, os sistemas de informação têm evoluído exponencialmente. Cada vez mais são gerados e guardados dados em suporte digital e frequentemente de forma automática. A quantidade de informação residente nesses dados é de tal modo elevada que se torna cada vez mais difícil esses dados serem trabalhados manualmente e deles retirar informação útil, sobretudo em tempo hábil. A solução passa por utilizar esses mesmos sistemas de informação para trabalhar os dados e deles retirar informações [25].

Portanto, o anseio passa a ser em criar sistemas computacionais que sejam capazes de analisar fatos, de aprender descrições compactas e de compreender esses fatos. Em outras palavras, que sejam capazes de melhorar seu desempenho com base nos resultados passados.

A classificação de dados possui diversos objetos com várias características/atributos pertencendo cada um deles a um único grupo/classe. Assim, o objetivo é criar uma regra que seja capaz de prever a classe que pertence cada objeto, baseado em suas características.

Por exemplo:

- Dado um conjunto de respostas, determinar se uma inteligência é predominante ou não em indivíduo.

Considerando um conjunto de objetos definidos por:

\vec{a} = características ou atributos

$C = \{c_1, \dots, c_i\}$ = um conjunto de classes

Suponhamos que exista uma função f desconhecida e que associa cada objeto a um grupo

$$f: \vec{a} \rightarrow c_i$$

O problema de classificação consiste em:

- Dado um conjunto de treino definido por pares atributos-classes $\langle \vec{a}_i, c_i \rangle^n$,
- Encontrar uma função $\vec{f}: \vec{a} \rightarrow c_i$

Para se obter a classe a que pertence o exemplo com base nos atributos é utilizado um *modelo de decisão (ou classificador)*. Esse modelo é criado através de um processo de aprendizagem feito com base em informação histórica (exemplos para os quais são conhecidos os atributos e a respectiva classe). Esse histórico de exemplo define um *conjunto de treino* (Fig. 6(a)), o qual é aplicado a um algoritmo de aprendizagem que retorna um modelo de decisão.

Novos exemplos não classificados são passados ao (s) modelo (s) de decisão. O conjunto de novos exemplos denominam-se *conjunto de teste*. Este conjunto terá o mesmo tipo de atributos do conjunto de treino, mas a classe a que cada exemplo pertence é desconhecida, mas pertencente ao conjunto C.

A fase de teste consiste em obter predições da classe a que pertence cada exemplo do conjunto de teste.

Na figura 6 apresenta-se diagrama das fases de Treino e Teste de SVMs:

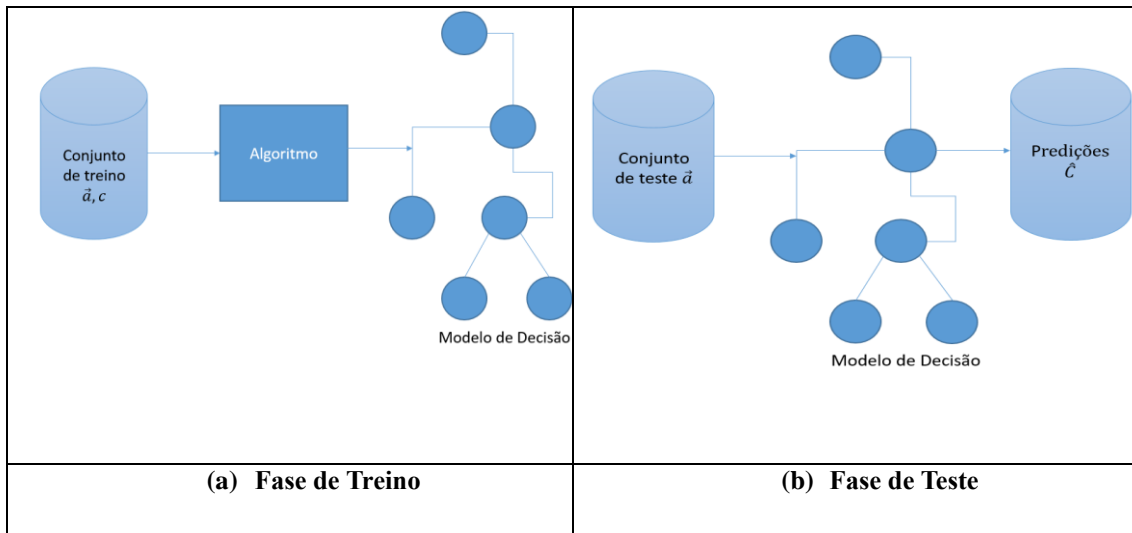


Figura 6: Diagramas de Treino (a) e Teste (b)

2.6 Métodos de Votação Direta

Após a etapa de decomposição e treino, obtém-se k' modelos de decisão. Na fase de reconstrução, cada exemplo é classificado por todos os modelos, dando origem a k' predições distintas [20, 21, 26, 28].

A predição final resulta da aplicação da função de reconstrução $r(a)$:

$$r(a) = \max[i \in \{1, \dots, k'\}, \sum_{j=1}^k (b_i, (a) = j)].$$

Basicamente, consiste na votação simples dos resultados dos diversos classificadores biclasse, sendo que a classe mais votada seria a da predição final. No caso de empate, a seleção é feita aleatoriamente entre as classes empatadas.

2.6.1 Método de Votação-direta com probabilidades

Esse método é uma variação do método apresentado na seção anterior. Esta variação é uso de probabilidades e requer que os classificadores usados retornem a probabilidade de um exemplo pertencer a cada classe [20].

Considerando:

$$p_{ij} = \text{Prob}(c_i | c_i \text{ ou } c_j), i \neq j$$

Devido a probabilidade de o (não contrair a preposição antes da oração infinitiva) exemplo ser da classe i , quando i é comparado com j , obviamente $p_{ij} = 1 - p_{ji}$. Cada modelo de decisão biclasse retorna, para além de uma classe c'_i , a probabilidade do exemplo ser dessa classe.

Por exemplo, no caso de 3 classes teríamos 3 modelos de decisão:

- $b_1: A - B$
- $b_2: A - C$
- $b_3: B - C$

Aplicando um exemplo a cada um dos modelos obteríamos, para cada modelo, uma de duas predições:

$$\begin{array}{llll}
 b_1(a) = A_{,P_{A-B}} & \Rightarrow P_{B-A}=1-P_{A-B} & V & b_1(a) = B_{,P_{B-A}} \Rightarrow P_{A-B}=1-P_{B-A} \\
 b_2(a) = A_{,P_{A-C}} & \Rightarrow P_{C-A}=1-P_{A-C} & V & b_2(a) = C_{,P_{C-A}} \Rightarrow P_{A-C}=1-P_{C-A} \\
 b_3(a) = B_{,P_{B-C}} & \Rightarrow P_{C-B}=1-P_{B-C} & V & b_3(a) = C_{,P_{C-B}} \Rightarrow P_{B-C}=1-P_{C-B}
 \end{array}$$

Com base nos diversos valores de p_{ij} obtém-se a matriz de probabilidades:

$$\begin{pmatrix} \cdot & P_{A-B} & P_{A-C} \\ P_{B-A} & \cdot & P_{B-C} \\ P_{C-A} & P_{C-B} & \cdot \end{pmatrix}$$

Define-se agora a probabilidade aproximada do exemplo ser da classe i como

$$p_i = \frac{2 \sum_{i \neq j} p_{ij}}{k(k-1)}$$

Finalmente, a predição seria a classe com maior probabilidade aproximada.

$$r(a) = \max(P_i)$$

2.7 Método de Votação-distribuída

Este método de reconstrução se aplica ao método de decomposição um contra todos e assume que cada classificador $b_i(a)$ seleciona uma classe $c_i, i = \{1, \dots, k\}$ ou uma (paralelismo) “outra”. Vale ressaltar que um problema multiclasse com k classes é decomposto em k' problemas de duas classes: uma classe *principal* contra as classes restantes, também conhecidas como *outra*, ou seja, classe única.

Para cada exemplo e , após a predição dos k' classificadores, a classe final predita é obtida através de uma votação.

Resumidamente, pode-se dizer que cada exemplo classificado k' vezes, obtendo-se uma classe c'_i por cada modelo de decisão. Para cada c'_i , caso a classe predita fosse diferente da classe “outra”, essa classe acumularia um ponto. Caso fosse “outra”, um ponto seria distribuído equitativamente por todas as classes restantes. No final, a escolha recai na classe que tiver acumulado mais pontos.

2.7.1 Método de Votação-distribuída com probabilidades

Com base no método de votação direta com probabilidades apresentado na seção 2.6, podemos aplicar as probabilidades ao método de reconstrução referido na seção 2.7, logo sendo também este associado à decomposição um contra todos, considerando:

$$P_{i-outra} = Prob(c_i | c_i \text{ ou } outra)$$

Como a probabilidade de o exemplo ser da classe c_i , quando c_i é comparada com as restantes classes.

Para cada uma das classes restantes a probabilidade será:

$$P_{outra-i} = \frac{1 - P_{i-outra}}{k' - 1}$$

2.8 Treinamento de SVMs Não Lineares

As SVMs lineares podem ser generalizadas de forma a realizar classificações não lineares. Para isso, é utilizado neste trabalho o artifício de levar os dados de treinamento para um espaço de maior dimensão, onde os dados neste espaço se tornam linearmente separáveis. Assim, pode-se aplicar as SVMs lineares sobre os dados mapeados neste espaço, determinando um hiperplano de margem maximizada. Para que isso ocorra, a SVM não linear utiliza funções de *kernel* (núcleo), que permitem o acesso a esses espaços complexos, sendo para este trabalho o núcleo polinomial, conforme a equação abaixo:

$$K(x_i, x_j) = ((x_i \cdot x_j) + a)^p$$

A vantagem de aplicar uma função de *kernel* é a simplicidade de cálculo e a capacidade de representar espaços muito abstratos. Portanto, essas funções devem pertencer a um domínio no qual seja possível o cálculo de produtos internos. Para isso, são utilizadas as condições estabelecidas pelo *Teorema de Mercer*, conforme essa definição, os *kernels* devem ser matrizes positivamente definidas, isto é, a matriz K , em que $K_{ij} = K(x_i, x_j)$ para todo $i, j = 1, \dots, n$, deve ter autovalores maiores que 0 [25, 35].

Vale ressaltar que, no caso do *kernel* polinomial utilizado neste trabalho, são funções com complexidade crescente na medida que o expoente p aumenta. Sabendo que p deve ser especificada pelo usuário.

A obtenção de um classificador por meio de SVMs envolve a escolha de uma função de *kernel*, além de parâmetros desta função e do algoritmo para determinação do hiperplano ótimo.

A seguir, é apresentado na figura 7 o algoritmo de Determinação do hiperplano ótimo no espaço de características [46].

- 1: Para qualquer conjunto de treinamento $\square(S) = \{(\Phi(x_1), y_1), \dots, (\Phi(x_n), y_n)\}$
- 2: Seja $\alpha^* = (\alpha_1^*, \dots, \alpha_n^*)$ a solução do seguinte problema de otimização com restrições:
- 3: Maximizar: $\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \Phi(x_i) \cdot \Phi(x_j)$
- 4: Sob as restrições:
$$\begin{cases} \sum_{i=1}^n y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \dots, n \end{cases}$$
- 5: O par (w^*, b^*) apresentado a seguir define o hiperplano ótimo.
- 6: $w^* \leftarrow \sum_{i=1}^n \alpha_i^* y_i \Phi(x_i)$
- 7: $b^* \leftarrow -\frac{1}{2} \left[\min_{\{i/y_i=+1\}} (w^* \cdot \Phi(x_i)) + \max_{\{i/y_i=-1\}} (w^* \cdot \Phi(x_i)) \right]$

Figura 7: Algoritmo de Determinação do hiperplano ótimo no espaço de características, (Verdi, 2001).

Assim, o problema de classificação passa ser a avaliação da função:

$$g(x) = \text{sgn} \left(\sum_{x_i \in SV} \alpha_i^* y_i \Phi(x_i) \cdot \Phi(x) + b^* \right) = \begin{cases} +1 \text{ se } \sum_{x_i \in SV} \alpha_i^* y_i \Phi(x_i) \cdot \Phi(x) + b^* > 0 \\ -1 \text{ se } \sum_{x_i \in SV} \alpha_i^* y_i \Phi(x_i) \cdot \Phi(x) + b^* < 0 \end{cases}$$

Essa avaliação é resolvida com a aplicação de uma função de *kernel* que, neste caso, é a polinomial, pois, a única informação necessária sobre o mapeamento Φ é uma definição de como o produto interno $\Phi(x_i) \cdot \Phi(x_j)$ pode ser realizado para quaisquer x_i e x_j pertencentes ao espaço de entradas.

2.9 SVMs para Várias Classes

Originalmente, as SVMs são utilizadas para classificação dos dados em duas classes distintas. Estas podem ser denominadas positivas e negativas, respectivamente. Contudo, muitas aplicações envolvem o agrupamento em mais de duas classes, conforme este trabalho. No entanto, este fato não inviabiliza o uso das SVMs. Diversas técnicas são propostas para entender problemas deste tipo, entre elas a decomposição de problemas multiclasse.

Em um problema multiclases, o conjunto de treinamento é composto por pares (x_i, y_i) , tal que $y_i \in 1, \dots, k$.

A decomposição de problemas multiclases consiste em decompor um conjunto de dados com várias classes em vários conjuntos de duas classes (biclasse). A cada um desses conjuntos de dados biclasse é aplicado um algoritmo de classificação, obtendo-se vários modelos de decisão biclasse b_i (um por cada conjunto).

Posteriormente, na fase de teste, cada novo exemplo é classificado pelos vários modelos de decisão biclasse criados anteriormente. Para cada exemplo, tem-se um conjunto de predições. Esse conjunto de predições é aplicado à função de agregação com vistas para obtenção da predição final.

Para exemplificar o conceito, dado um problema de aprendizagem definido por um conjunto de exemplos D :

$$D = \langle \{\vec{a}, c\}^n \rangle, c \in \{c_1, c_2, \dots, c_k\}, k > 2$$

D é decomposto em vários conjuntos biclasse:

$$d(D) = B_1 \wedge B_2 \dots \wedge B_{k'}$$

Onde

$$B_i = \langle \{\vec{a}, c'_i\}^m \rangle, c'_i \in \{0,1\}, i \in \{1, \dots, k'\}$$

Esquemáticamente:

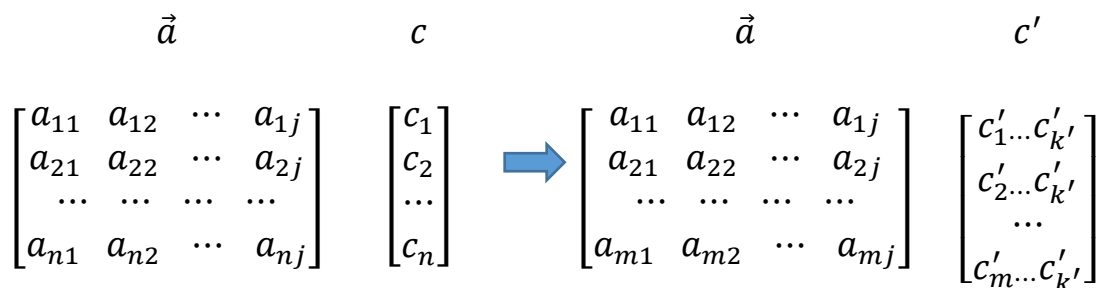


Figura 8: Decomposição em problemas biclasse

A cada problema B_i , é aplicado um algoritmo de aprendizagem, gerando o modelo de decisão b_i .

$$b_i = \text{Algoritmo}(B_i)$$

Posteriormente, na fase de classificação, cada novo exemplo é aplicado a todos os b_i obtendo-se um conjunto de predições F' :

$$F' = \{c'_1, c'_2, \dots, c'_k\}, c' \in \{0,1\}$$

Nesse conjunto é aplicada a função de reconstrução para obtenção da predição final:

$$f = r(F'), f \in \{c_1, c_2 \dots c_k\}$$

Onde f é uma classe do conjunto de dados multiclasse.

Portanto, na fase de decomposição, pretende-se decompor um problema multiclasse em vários problemas biclasse. Cada problema biclasse terá associado um modelo de decisão construído segundo um determinado algoritmo e com base num conjunto de dados com duas classes B_i .

Conforme seção 2.5, neste trabalho serão utilizados dois métodos de classificação, sendo decomposição um por classe e decomposição duas a duas.

No método Decomposição duas a duas, para cada uma das combinações é gerado um modelo de decisão. Um problema multiclasse com k classes é transformado em k' problemas biclasse, cada um com o seu modelo de decisão b , independente, onde:

$$k' = \frac{k(k-1)}{2}$$

Cada problema biclasse é definido por um conjunto de treino constituído pelos exemplos de apenas 2 classes. Por exemplo, para um problema original com 4 classes (A,B,C,D), seriam geradas 6 combinações e consequentemente 6 problemas biclasse: A-B,A-C,A-D,B-C,B-D,C-D. O conjunto de treino de A-B consistiria apenas nos exemplos pertencentes às classes A e B. Aplicando um algoritmo de classificação a este conjunto, obtém-se o classificador b_{AB} ; o conjunto de treino de B-D teria apenas os exemplos das classes B e D. Aplicando o mesmo algoritmo de classificação a este conjunto, obtém-se o classificador b_{DB} ; entre outros. Cada modelo de decisão b_i é construído, tendo como base um conjunto de treino de dimensão inferior ao conjunto de treino do problema multiclasse.

Já no método de decomposição um por classe, cada uma das k classes do problema original é comparada contra todas as outras.

Portanto, um problema multiclasse com k classes é decomposto em k' problemas de duas classes: uma classe principal contra as restantes classes (todas elas denominadas como uma classe única). Assim, têm-se:

$$k' = k$$

Assim, os vários conjuntos de treino biclasse têm a mesma dimensão do conjunto de treino multiclasse.

Após a decomposição e treino, obtém-se k' modelos de decisão. Na fase de reconstrução, cada exemplo é classificado por todos os modelos, dando origem a k' predições distintas.

Todas estas predições são biclasse formam o conjunto de predições $F' = \{c'_1, c'_2, \dots, c'_{k'}\}$. Na fase de reconstrução agregam-se essas diversas predições para obter a predição final $f = r(F'), f \in \{c_1, c_2, \dots, c_k\}$.

2.10 Balanceamento de Classes

Na fase de treinamento, etapa de decomposição, o Custo de Erros em alguns tipos de problemas de classificação, com diferentes erros de classificação, pode ter custos variados. No entanto, a maior parte dos métodos de classificação assume como objetivo a redução do número de erros independentemente do seu custo. Neste trabalho, é aplicado o balanceamento de classes por meio do método de amostragem que utiliza o filtro *Resample* detalhado nas próximas seções.

Uma base de dados é dita desbalanceada no domínio de classificação quando existem muito menos casos de algumas classes do que de outras [22]. As classes com pouca representação são chamadas de classes raras [48].

Em alguns classificadores o desbalanceamento de classes é indesejável, pois as classes minoritárias também possuem informações importantes e têm a mesma relevância que as demais. Sendo assim, pode ser utilizado um filtro que adota o método de introdução de custos de classificação incorreta ou método de classificação sensível ao custo para realizar o balanceamento das classes.

Na maioria dos problemas de classificação costuma-se medir o desempenho dos classificadores por meio da taxa de erro (*error rate*). A taxa de erro é a proporção dos erros encontrados entre todas as instâncias e é um indicador de desempenho global do classificador. Muitos algoritmos de classificação assumem que todos os erros têm o mesmo custo e são,

normalmente, desenhados para minimizar o número de erros. Nesses casos, a taxa de erro é equivalente à atribuição do mesmo custo para todos os erros de classificação, ou seja, por exemplo, no caso de uma classificação binária, os falsos positivos teriam o mesmo custo que os falsos negativos. No entanto, em muitas situações, cada erro poderá ter um custo diferente associado. De fato, na maioria das situações do dia a dia, as decisões têm custos diferentes e uma má decisão pode ter consequências graves. Assim, é importante considerar os diferentes custos associados às decisões, ou seja, às classificações obtidas [38].

Desta forma, pode-se designar por classificação sensível ao custo (*cost-sensitive classification*) quando os custos são ignorados na construção do classificador e são usados na previsão de novos casos. Por outro lado, é possível falar em aprendizagem sensível ao custo (*cost-sensitive learning*) quando os custos são considerados durante a fase de treino, podendo posteriormente ser ignorados ou não na fase de previsão [39].

Neste trabalho, os custos devem ser calculados na fase de construção do classificador. Dessa forma, a aprendizagem sensível ao custo é considerada a opção mais ajustada e com melhores resultados [39].

Na aprendizagem sensível ao custo, é possível alterar a distribuição das classes, para que o classificador minimize num determinado sentido (com aumento da sensibilidade ou da especificidade) os custos com novas instâncias. Procura-se com isto que as classes tenham uma distribuição proporcional à sua importância, ou seja, ao seu custo [51].

O método *Resample* é um dos métodos de amostragem estatísticos capazes de estimar apropriadamente a taxa de erro verdadeira, mesmo que haja um conjunto reduzido de dados [47]. Esse tipo de método utiliza conjuntos de teste para estimar a taxa de erro verdadeira, os quais contêm exemplos que não foram utilizados no treinamento do sistema de aprendizado. Diferentes métodos de amostragem particionam de diferentes formas o conjunto de dados original em conjuntos de treinamento e teste.

O filtro *Resample* é disponibilizado na ferramenta Weka, que utiliza a sensibilidade ao custo aplicando um metaclassificador que manipula as instâncias de treino de forma a obter classificadores sensíveis ao custo em conjunto com o classificador base, tornando-o sensível a valores manipuláveis de custos [18].

Na classificação realizada pelo Weka, o desbalanceamento de classes está fortemente relacionado com a abordagem de aprendizado sensível ao custo [19]. Ou seja, um caso de falso negativo é mais custoso que um caso de falso positivo.

2.11 Método de Agrupamento

De acordo com os conceitos apresentados nas seções 2.6 e 2.7, é possível aplicar o método de Votação-distribuída na fase de reconstrução.

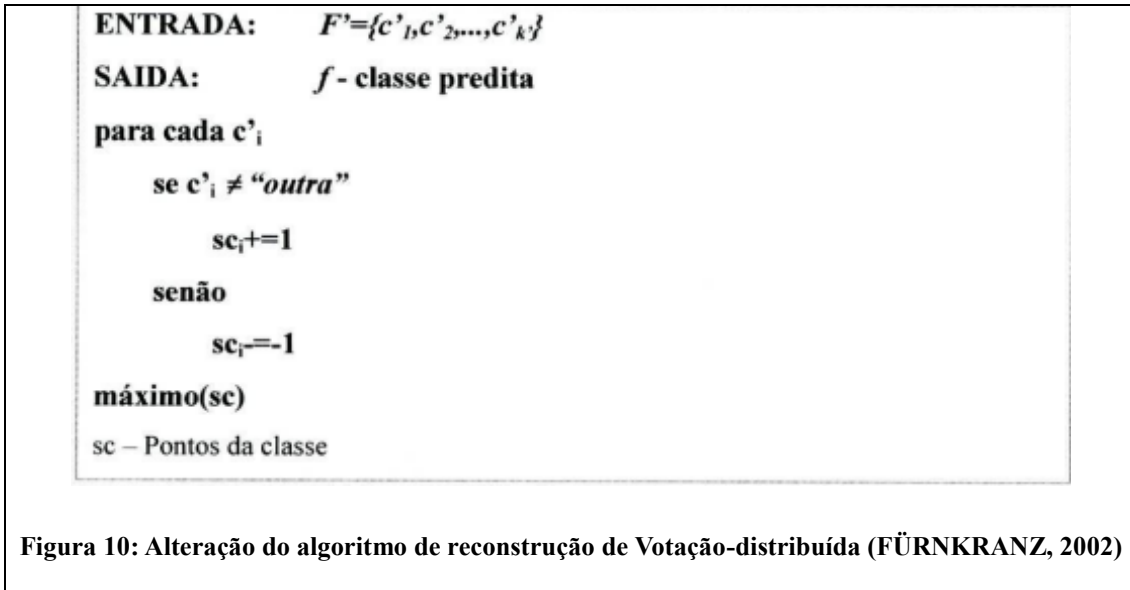
Para cada exemplo, e após a predição dos k' classificadores, a classe final predita é obtida por meio de uma votação segundo o algoritmo [12, 13, 20, 28]:

```
ENTRADA:  $F'=\{c'_1, c'_2, \dots, c'_k\}$   
SAIDA:  $f$  - classe predita  
para cada  $c'_i$   
  se  $c'_i \neq \text{"outra"}$   
     $sc_i += 1$   
  senão  
     $sc_j += 1/(k-1), j \neq i$   
máximo(sc)  
sc – Pontos da classe
```

Figura 9: Algoritmo de reconstrução Votação-distribuída (FÜRNKRANZ, 2002)

Portanto, se a classe predita for diferente de “outra”, então essa classe acumulará um ponto. Caso contrário, perderá um ponto.

Uma forma de votação ligeiramente diferente seria:



Com base nesse conceito, podemos aplicar as probabilidades ao método de reconstrução por meio da fórmula:

$$P_{outra-i} = \frac{1 - P_{i-outra}}{k' - 1}$$

Por exemplo, para o caso de 3 classes teríamos 3 modelos de decisão:

- b_1 : A-outra $m = A$
- b_2 : B-outra $m = B$
- b_3 : C-outra $m = C$

Cada classificador retornaria para cada exemplo a probabilidade de o exemplo ser da classe m :

$$b_1(a) = A, P_{A-outra} \quad \Rightarrow P_{outra-A} = (1 - P_{A-outra})/2$$

$$b_2(a) = B, P_{B-outra} \quad \Rightarrow P_{outra-B} = (1 - P_{B-outra})/2$$

$$b_3(a) = C, P_{C-outra} \quad \Rightarrow P_{outra-C} = (1 - P_{C-outra})/2$$

Com base nessa informação constrói-se a matriz de probabilidades (exemplificada para $k=3$):

$$P = \begin{pmatrix} P_{A-outra} & P_{outra-B} & P_{outra-C} \\ P_{outra-A} & P_{B-outra} & P_{outra-C} \\ P_{outra-A} & P_{outra-B} & P_{C-outra} \end{pmatrix}$$

A diagonal da matriz corresponde a probabilidades retornadas pelos classificadores.

A probabilidade aproximada do exemplo ser da classe c_i define-se como:

$$P_i = \frac{\sum_{j=1}^{k'} P_{ij}}{k'}$$

Que não é mais do que a média das diversas probabilidades de um exemplo ser da classe i .

Visando obter percentuais dos resultados obtido na fase de reconstrução, por meio do método de Votação-distribuída, foi aplicado o conceito de algoritmos de agrupamento.

Algoritmos de agrupamento têm como objetivo particionar uma massa de objetos em grupos de objetos similares. Essa tarefa pode ser dividida entre agrupamentos supervisionados, em que algum mecanismo externo forneça informações de como classificar corretamente um dado objeto. Por outro lado, temos o agrupamento não supervisionado, onde a classificação deve ser realizada sem referências externas.

Para este projeto, é utilizado o agrupamento não supervisionado, no qual a utilização de categorias pré-definidas é desnecessária. Portanto, o objetivo desta técnica é agrupar objetos com alto grau de semelhança, o que caracteriza alguma função de distância.

A classificação não supervisionada pode ser dividida em dois tipos: algoritmos de agrupamento hierárquico e algoritmos de agrupamento particionado [42, 50].

O algoritmo K-means, apresentado por J.B. MacQueen em 1967, tenta fornecer uma classificação de acordo com os próprios dados, sendo a classificação feita por similaridade de grupos e o objeto atribuído ao grupo a que mais se assemelha. [4, 42, 50].

Portanto, o algoritmo K-means busca minimizar a distância dos elementos a um conjunto de k centros dado por $x = \{x_1, x_2, \dots, x_k\}$ de forma iterativa. A distância entre um ponto p_i e um conjunto de *clusters*, dada por $d(p_i, x)$ é definida como sendo a distância do ponto ao centro mais próximo dele. A função a ser minimizada é dada por:

$$d(p, x) = \frac{1}{n} \sum_{i=1}^n d(p_i, x)^2$$

Assim, o algoritmo depende de um parâmetro (k = número de *clusters*).

2.12 Sistema de Recomendação

Sistemas de Recomendação são ferramentas de *software* e técnicas que fornecem sugestões de itens para um usuário [24, 34]. As sugestões referem-se a vários processos de tomada de decisão, tais como, qual item comprar, qual música ouvir, ou quais notícias *online* ler. O “Item” é o termo geral usado para designar o que o sistema recomenda aos usuários. Um Sistema de Recomendação se concentra em um tipo específico de item (por exemplo, CDs ou notícias) e a técnica de recomendação central utilizada para gerar as recomendações são todas personalizadas para proporcionar sugestões úteis e eficazes para um tipo específico.

Um dos objetivos dos sistemas de recomendação é auxiliar no aumento da capacidade e da eficácia dos processos de indicação [3]. Em um sistema típico, as pessoas fornecem recomendações como entradas, sendo estas agregadas pelo sistema, que as direcionam para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos maiores desafios de tais sistemas é realizar a combinação adequada entre as expectativas dos usuários e os produtos, serviços, ou pessoas e a serem recomendados.

Alguns sistemas de recomendação não estimam totalmente a utilidade de um item antes de fazer uma recomendação, mas podem aplicar algumas heurísticas para a importância de um item para o usuário. Para isso, são utilizados algoritmos específicos e vários tipos de conhecimento sobre os usuários e os itens (por exemplo, o sistema pode supor que a função de utilidade é booleana e, portanto, ele vai apenas determinar se um produto é ou não é útil para o utilizador). Consequentemente, partindo do princípio de que existe algum conhecimento disponível sobre o usuário que está solicitando a recomendação, o conhecimento sobre itens, e as recomendações que outros usuários receberam, o sistema irá elevar o nível desse conhecimento com um algoritmo apropriado para gerar várias previsões de serviços e após isso, gerar as recomendações [11]. Os Sistemas de Recomendação podem ser classificados entre recomendação com base em conteúdo, recomendação de filtragem colaborativa, recomendação baseada em conhecimento e recomendação híbrida [10, 23, 31].

Reategui & Cazella (2005) realizaram um estudo em que existem diversas técnicas para descoberta de conhecimento aplicáveis nos sistemas de recomendação, que podem ser mais eficientes do que outras, dependendo do domínio de aplicação. Neste trabalho, são apresentadas três delas:

- *Regras de Associação*: são técnicas que procuram encontrar padrões entre conjuntos de valores ou em gráficos evolutivos de valores. Uma regra de associação é uma indicação de que existe um certo relacionamento entre fatos diferentes. Essa regra constitui-se em

um relacionamento $X \rightarrow Y$, onde X e Y são conjuntos de itens e a intersecção entre os mesmos constitui-se em um conjunto vazio. Cada regra de associação é associada a um fator de suporte, definido aqui como “FSup”, e a um fator de confiança definido por “FConf” [32].

$$FSup = \frac{\text{Frequência de } X \text{ e } Y}{\text{Total de amostras}} \quad \text{equação (1)}$$

$$FConf = \frac{\text{Frequência de } X \text{ e } Y}{\text{Frequência de } X} \quad \text{equação (2)}$$

- *Classificação*: a descoberta de conhecimento consiste em determinar a classe ou categoria de um elemento a partir das características, que podem ser os atributos deste elemento e com base em classes pré-definidas. Uma nova classe pode ser acrescida ao conjunto, caso seja detectado um elemento com características contraditórias que satisfazem duas classes, mas não podendo ser classificado assim por serem classes disjuntas, ou então quando as características do elemento não se enquadrem em nenhuma das classes pré-estabelecidas [32].

Exemplo: explicar o uso de classificação consiste em supor que um sistema de recomendação guarde algumas informações sobre os seus usuários, por exemplo, país e áreas de interesse. Estas informações podem ser interessantes para prever que tipo de usuário está mais interessado em certa área de pesquisa, desta forma, o sistema pode então concentrar esforços de recomendação para classificar os usuários por área de pesquisa [32].

- *Agrupamento*: este método consiste em associar um item a uma ou várias classes categóricas (*clusters*), determinando as classes pelos dados, independentemente da classificação pré-definida. Os grupos são definidos por meio do agrupamento de dados baseados em medidas de similaridade ou modelos probabilísticos, visando detectar a existência de diferentes grupos dentro de um determinado conjunto de dados e, em caso de sua existência, determinar quais são eles [32].

Para Silva [37], outra definição de agrupamento, que pode ser realizada de forma simplificada, é o processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos, sendo este o processo inverso da classificação. A partir de técnicas de agrupamento, são analisados os atributos de caracterização e os de discriminação para enquadrar

os elementos nas suas respectivas classes, que são definidas por essas técnicas. O algoritmo de agrupamento automaticamente deve criar grupos por meio de partições de um banco de dados em conjunto de tuplas similares. A partição acontece de forma que as tuplas de valores e atributos semelhantes sejam reunidas nos mesmos grupos. Um agrupamento de qualidade surge onde a similaridade intraclasse é alta e a interclasse é baixa.

O agrupamento é um tipo de aprendizado não supervisionado uma vez que fica a cargo do algoritmo a definição de quais atributos e seus valores serão utilizados para montar agrupamentos de dados similares [32].

2.13 Teoria das Inteligências Múltiplas

A teoria de Gardner [14, 15] afirma que todos os indivíduos normais dispõem de oito inteligências sendo elas: linguística, lógico-matemática, espacial, corporal-cenestésica, musical, interpessoal, intrapessoal e naturalista, e que cada indivíduo tem maior habilidade para desenvolver uma determinada inteligência. Essa teoria possibilita, então, trabalhar com estímulos específicos e com isso desenvolver de forma mais adequada determinada habilidade de cada um.

As inteligências Múltiplas foram mapeadas inicialmente como sendo sete, e segundo Armstrong [1], Gardner ofereceu um meio de mapear a ampla gama de capacidades dos seres humanos nessas sete categorias ou “inteligências”, que seriam:

Inteligência Lógico-Matemática: Capacidade de usar números de forma efetiva como um matemático, assim como a capacidade de raciocínio como um cientista. Esta inteligência inclui sensibilidade a padrões, afirmações, proposições, funções e outras abstrações relacionadas.

Inteligência Espacial: Capacidade de perceber com precisão o mundo visual, permitindo realizar transformações sobre essas percepções como um decorador de interiores. Esta inteligência envolve sensibilidade à cor, linha, forma, configuração e espaço, e às relações existentes entre esses elementos.

Inteligência Corporal Cenestésica: Habilidade no uso do corpo todo para expressar ideias e sentimentos, além de facilitar no uso das mãos para manipular, produzir ou transformar coisas. Esta inteligência inclui habilidades físicas específicas, tais como coordenação, equilíbrio, destreza, força flexibilidade e velocidade. Tendo como exemplo atletas, malabaristas e médicos cirurgiões.

Inteligência Musical: Facilidade em perceber, discriminar, transformar e expressar formas musicais. Esta inteligência inclui sensibilidade ao ritmo, tom ou melodia e timbre.

Inteligência Interpessoal: Capacidade de perceber e fazer distinções no humor, intenções, motivações e sentimentos das outras pessoas. Isso pode incluir sensibilidade a expressões faciais, voz e gestos; a capacidade de discriminar muitos tipos diferentes de sinais interpessoais; e a capacidade de responder efetivamente a estes sinais.

Inteligência Intrapessoal: Capacidade de autoconhecimento, ou seja, de se conhecer. Esta inteligência inclui possuir uma imagem precisa de si mesmo, tendo consciência dos estados de humor, intenções, motivações, temperamento e desejos; e a capacidade de autodisciplina, autoentendimento e autoestima.

2.14 Amostragem Aleatória Simples

Nos testes, foi utilizada uma pesquisa quantitativa, de caráter descritivo, por meio de coleta de amostras provenientes de questionário aplicado aos indivíduos selecionados. Essas pesquisas têm como objetivo a descrição das características de determinada população ou fenômeno ou então o estabelecimento de relações entre variáveis. Já os levantamentos consistem na solicitação de informações a respeito do problema estudado a um grupo significativo de pessoas, para que, na sequência, ocorra a realização das análises quantitativas Gil [16].

Para que seja possível fazer inferências para uma população, utilizando-se resultados obtidos a partir de uma amostra, é necessário que essa amostra seja coletada com alguns cuidados, como a atenção ao seu tamanho mínimo e à forma de coleta, que deve ser probabilística, ou seja, deve obedecer ao padrão AAS (amostragem aleatória simples). De acordo com esse padrão, cada elemento coletado tem tanta probabilidade de ser observado quanto qualquer outro elemento da população, e sua observação pode ocorrer uma única vez no processo de amostragem. Existem várias formas de se determinar o tamanho de um espaço amostral. A utilizada neste trabalho está descrita pela equação:

$$n = \frac{N \cdot Z^2}{(N - 1) \cdot E^2 + Z^2}$$

Considerando a equação, tem-se que “Z”=1,645, sendo que “Z” é o desvio do valor médio aceitável para alcançar o nível de confiança desejado. Em função do nível de confiança que se pretende alcançar, deve ser utilizado um valor determinado que é dado pela forma da distribuição de Gauss [43].

Os valores de confiança mais utilizados e os valores de Z correspondentes podem ser encontrados na Tabela 1.

Tabela 1: Valores críticos associados ao grau de confiança na amostra

| Grau de Confiança | α | Valor Crítico $Z_{\alpha/2}$ |
|--------------------------|----------------------------|--|
| 90% | 0,10 | 1,645 |
| 95% | 0,05 | 1,96 |
| 99% | 0,01 | 2,575 |

O nível de confiança adotado é de 90% (noventa por cento), assim, o valor de “Z” deve ser igual a 1,645 justificando desta forma o valor adotado e “E” é a margem de erro máximo admitido, considerando que o número total de indivíduos é relativamente baixo, e devido à dificuldade de se conseguir entrevistar 100% dos indivíduos devido a alguns alunos estarem matriculados, mas não frequentarem as aulas, foi adotado que a margem de erro é de 20% (vinte por cento).

2.15 Matriz Confusão

O resultado obtido após a classificação é expresso por meio de uma matriz de confusão que contém informações sobre as classificações atuais e as previsões realizadas pelo algoritmo de classificação. O desempenho do classificador é comumente avaliado utilizando os dados da matriz que tem seus valores gerados a partir do treinamento da SVM, assim gerando uma matriz dos subconjuntos entre si.

Entre as informações obtidas no treinamento, é a matriz de confusão que oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas versus as classificações previstas, ou seja, que podem ocorrer para cada classe. O número de acertos, para cada classe, se localiza na diagonal principal da matriz. Os demais elementos para $L \neq C$ representam erros na classificação.

A matriz de confusão da figura 12 pode ser lida na horizontal e vertical. Por meio da tabela 2 podem ser visualizados os dados referentes a leitura realizada na horizontal e na tabela 3 na vertical.

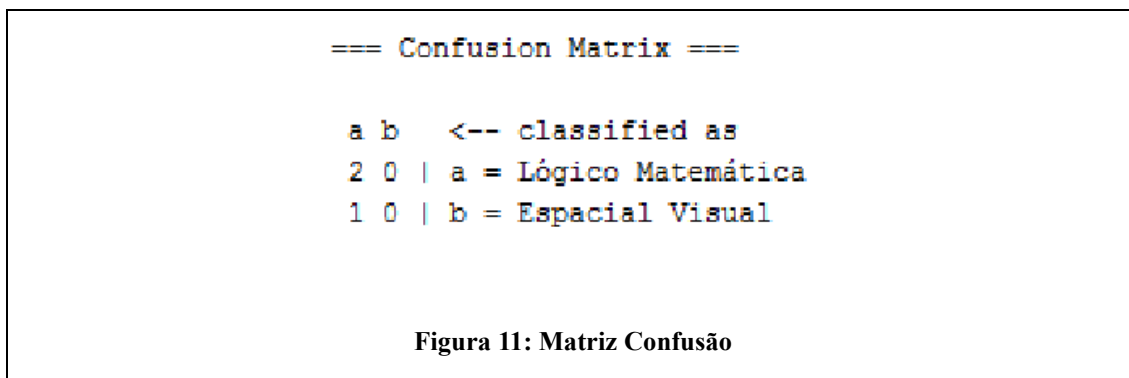


Tabela 2: Leitura da Matriz Confusão na Horizontal

| Classe | Total de itens por linha | Itens corretamente classificados | Itens incorretamente classificados |
|-------------------|---------------------------------|---|---|
| Lógico-Matemática | 2 | 2 | 0 |
| Espacial Visual | 2 | 1 | 0 |

Tabela 3: Leitura da Matriz Confusão na Vertical

| Classe | Total de itens por coluna | Itens corretamente classificados | Itens incorretamente classificados |
|-------------------|----------------------------------|---|---|
| Lógico-Matemática | 2 | 3 | 0 |
| Espacial Visual | 2 | 0 | 0 |

Pode-se constatar que existem alguns erros na classificação que serão analisados visando melhorar a precisão do classificador.

2.15.1 Exatidão das informações para classificação

Além da matriz de confusão, o classificador apresenta os valores detalhados referentes à exatidão das informações para as classes utilizadas no processo de classificação, tais como, a taxa de verdadeiros positivos, taxa de falsos positivos, precisão, entre outros.

As informações disponibilizadas na figura 12 são definidas conforme informações fornecidas abaixo [35] e [49]:

| === Detailed Accuracy By Class === | | | | | | | |
|------------------------------------|---------|---------|-----------|--------|-----------|----------|-------------------|
| | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
| | 1 | 1 | 0.667 | 1 | 0.8 | 0.5 | Lógico Matemática |
| | 0 | 0 | 0 | 0 | 0 | 0.5 | Espacial Visual |
| Weighted Avg. | 0.667 | 0.667 | 0.444 | 0.667 | 0.533 | 0.5 | |

Figura 12: Valores de Exatidão das Informações do Classificador após Treinamento

✓ *TP rate* (taxa de verdadeiros positivos) é a proporção de casos positivos que foram corretamente identificados.

$$TP Rate = \frac{\text{verdadeiros positivos}}{\text{total de positivos}}$$

✓ *FP rate* (taxa de falsos positivos) é a proporção de casos negativos que foram incorretamente classificados como positivos.

$$FP Rate = \frac{\text{falsos positivos}}{\text{total de negativos}}$$

✓ *Precision* (precisão) é a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas.

$$Precision = \frac{\text{verdadeiros positivos}}{\text{verdadeiros positivos} + \text{falsos positivos}}$$

- ✓ *Recall* (sensibilidade) é a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas.

$$Recall = \frac{\textit{verdadeiros positivos}}{\textit{verdadeiros positivos} + \textit{falsos negativos}}$$

- ✓ *F-Measure* é a média ponderada da precision com o recall.

$$F_{Measure} = \frac{2x(Precision \times Recall)}{(Precision + Recall)}$$

- ✓ ROC Área é a área da curva ROC (*Receiver Operating Characteristic Curve* ou Curva Característica de Operação do Receptor), os gráficos ROC são uma outra maneira além de matrizes de confusão para examinar o desempenho de classificadores. Um gráfico ROC é um gráfico com a taxa de falsos positivos no eixo X e a taxa verdadeiros positivos no eixo Y. Quanto maior a área da curva, maior é a exatidão das respostas para o teste.

Após o processo de classificação, é necessário acessar a opção *cluster* para aplicação do método de agrupamento *K-means*. Este método tem a função de separar os valores obtidos no processo de classificação e agrupá-los em classes. Este procedimento se faz necessário para que se possa obter o resultado da classificação por inteligências em um percentual, facilitando o entendimento e a visualização dos resultados.

Os percentuais definem quais são as inteligências predominantes para o indivíduo e os tipos de inteligência classificados.

Capítulo 3- Metodologia

Neste capítulo são detalhados o uso dos conceitos do capítulo 2, de maneira que possa se desenvolver o mecanismo de classificação.

3.1 Aplicação da Teoria das Inteligências Múltiplas para identificação de Perfis

Na tabela 4, são apresentados os valores e tipos de respostas que cada indivíduo pode fornecer ao responder o questionário. Estas respostas são conciliadas a pesos usados pela SVM para definição das fronteiras para os vetores de suporte.

Tabela 4: Tipos de Respostas

| Valor Respostas | Tipos Respostas |
|------------------------|------------------------|
| 4 | Sempre |
| 3 | Muitas Vezes |
| 2 | Raramente |
| 1 | Nunca |

O Questionário possui três questões para cada tipo de inteligência, portanto, o indivíduo pode fornecer apenas três respostas dentre os quatro tipos existentes.

Essas respostas, junto às inteligências múltiplas, compõem o Vetor de Características das SVMs utilizadas neste experimento.

3.2 Definição da Ferramenta de Desenvolvimento para o Mecanismo de Classificação

O Weka é um *software* livre do tipo *open source* para mineração de dados, desenvolvido por um grupo de pesquisadores da Universidade de Waikato, Nova Zelândia. Ao longo dos anos, o Weka se consolidou como a ferramenta de *data mining* mais utilizada em ambiente acadêmico. Apesar de seu ponto forte ser a tarefa de classificação, ele também é capaz de minerar regras de associação e *clusters* de dados [49].

O software utiliza para a entrada de dados os arquivos de extensão *arff*, *csv*, *xrff* entre outros. A extensão utilizada neste trabalho é a *csv* que pode ser obtida por meio da conversão de planilhas do Microsoft Excel do formato *xls* ou *xlsx* para o formato *csv*. A justificativa para esta escolha é que o questionário utilizado para fornecer os parâmetros de entrada foi desenvolvido utilizando a ferramenta Microsoft Excel e a conversão para o formato *csv* permite que os dados inseridos na planilha sejam aceitos e processados pelo Weka.

3.2.1 Técnicas Utilizadas pelo WEKA

- *Regras de Associação*: identificação de grupos de dados que apresentam coocorrência entre si.
- *Regressão ou Predição*: aprendizado de uma função que pode ser usada para mapear os valores associados aos dados em um ou mais valores reais.
- *Agrupamento ou clustering*: identificação de grupos de dados onde os dados têm características semelhantes aos do mesmo grupo e onde os grupos tenham características diferentes entre si.
- *Classificação*: aprendizado de uma função que pode ser usada para mapear dados em uma de várias classes discretas definidas previamente, como por exemplo, as inteligências múltiplas.

Para este trabalho, a técnica selecionada será a de classificação, que, como entrada, recebe instâncias para as quais as classes são conhecidas. Com isso, cria-se um classificador ou modelo (fase de treinamento).

Como entrada em uma segunda fase, tem-se vários dados para os quais as classes não são conhecidas. Desta forma, utilizasse o classificador para indicar classes para estes dados, assim, podendo avaliar o modelo classificando instâncias com classes conhecidas. Essa é fase a de teste, que consiste em obter predições da classe a que pertence cada exemplo do conjunto de teste.

3.2.2 Biblioteca de Implementação - LibSVM

O LibSVM é uma biblioteca de implementações de SVM desenvolvida por Chin Chung Chang (2001) com várias finalidades: classificação, regressão e estimativa de distribuição. Um dos algoritmos de classificação implementado na biblioteca leva o nome de C-SVC [5].

A LibSVM possui um recurso que permite encontrar automaticamente os melhores parâmetros para SVM pelo uso de validação cruzada. Quando realizada, a otimização dos parâmetros da SVM provoca um ganho substancial no desempenho de classificação e isto motivou a escolha dessa biblioteca.

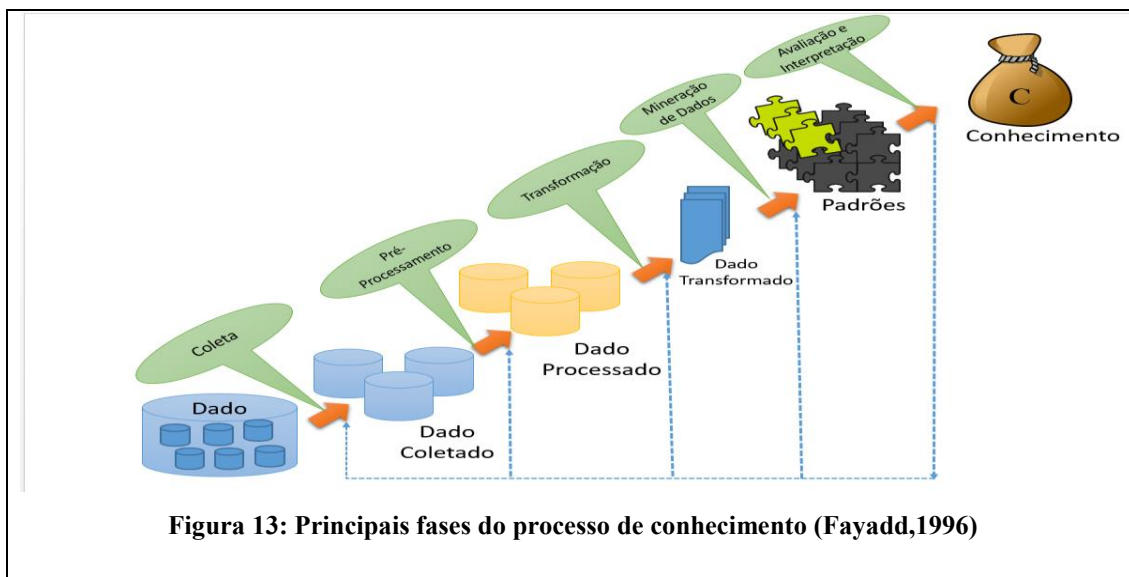
Em comparação com outros algoritmos de treinamento, o LibSVM também suporta múltiplas classes e a probabilidade estimada com base na escala de Platt aos valores de confiança são adequados após a aplicação do modelo de classificação aprendido [5]. Essa vantagem comprova o grau de confiança nos experimentos desenvolvidos neste trabalho.

3.3 Definição do Modelo Classificador para Sistemas de Recomendação

As técnicas utilizadas para a descoberta do conhecimento no Weka são as mesmas que as descritas neste trabalho: sistemas de recomendação, regras de associação, classificação e agrupamento. Dentro do contexto em que este trabalho está inserido, faz-se necessária a utilização da classificação como técnica de descoberta do conhecimento, estando relacionada aos princípios do treinamento da SVM. Vale ressaltar que, conforme conceitos abordados, esta fase é a de treino e que as etapas descritas a seguir compõem o treinamento.

A classificação no Weka passa por diversas etapas, sendo a de processamento realizada para a obtenção do conhecimento, podem ser visualizadas por meio da figura 13. Inicialmente, na etapa de coleta, são selecionados, na base de dados, os dados a serem analisados; na etapa de processamento, estes dados são processados e entendidos pelo sistema que, na etapa de transformação, os converte em dados que possam ser minerados e entendidos pelo Weka; na etapa

de mineração, é realizado o treinamento que deverá definir os padrões a serem localizados e, logo em seguida, é realizada uma busca por padrões que foram previamente informados como sendo os parâmetros para a definição das respostas; o resultado da mineração dos dados é interpretado e, como saída, é fornecida uma base de conhecimento. É importante ressaltar que ao término da última etapa o Weka pode realimentar uma das etapas anteriores refazendo o processo a partir da etapa em que ele retornou até que um padrão seja obtido e uma resposta seja gerada.



Este modelo é adaptado a sistemas de recomendação, pois, o mecanismo desenvolvido está inserido nesse contexto por possui características similares.

Ainda não se pode dizer que este mecanismo seja um sistema de recomendação por não possuir todas as funcionalidades necessárias e, conseqüentemente, o processo de recomendação não é inteiramente automatizado.

Assim, o mecanismo que utiliza este modelo classificador, não indica materiais automaticamente aos usuários, no entanto, ele fornece subsídios que auxiliam o professor a definir quais materiais serão recomendados, baseado no perfil dos alunos.

3.4 Filtro Resample aplicado em Balanceamento de Classes

O método de amostragem Resample no Weka balanceia um conjunto de dados por meio de uma amostragem com reposição. A utilização dessa técnica permite variar a proporção de

exemplos entre as classes, que podem ser similares à distribuição inicial ou próxima distribuição balanceada.

Este filtro é do tipo supervisionado, fator relevante pelo fato de realizar o tratamento do conhecimento obtido na Mineração dos Dados, sendo responsável por balancear os conjuntos de dados por meio de uma amostragem de reposição. A utilização dessa técnica permite variar a proporção de exemplos entre as classes, que pode ser similar à distribuição inicial ou próxima da distribuição balanceada e cada valor utilizado para o balanceamento pode ser escolhido mais de uma vez. Isso ocorre devido a uma das características da amostragem de reposição que, após cada etapa de seleção de amostras para o balanceamento, possibilita que o objeto selecionado retorne à classe, podendo ser selecionado novamente.

3.5 Aplicação Algoritmo SimpleKMeans para Agrupamento de Classes

A proposta do modelo de classificação consiste em se fazer um agrupamento de dados após a classificação, que ocorre ao final do treinamento das SVM's. Para isso, é aplicado o algoritmo de particionamento SimpleKMeans, que, dado um conjunto D de dados com n registros e k o número de agrupamentos desejados, organizam os objetos em k agrupamentos, tal que $k < = n$.

Dado um conjunto de dados, o algoritmo supracitado, usando o conceito de centroide, seleciona de forma aleatória k registros, cada um representando um agrupamento. Para cada registro restante, é calculada a (apenas o artigo é necessário) similaridade entre o registro analisado e o centro de cada agrupamento. O objeto é inserido no agrupamento com a menor distância, ou seja, maior similaridade. O centro do cluster é recalculado a cada novo elemento inserido. Diferentes variações surgiram: implementando otimizações para escolha do valor do k , novas medidas de dissimilaridade e estratégias para o cálculo do centro do agrupamento.

Proposto por J. MacQueen em 1967, o algoritmo de Análise de Agrupamento k-means (ou algoritmo das k-médias) é um dos mais conhecidos e utilizados, além de ser o que possui o maior número de variações. O algoritmo inicia com a escolha dos k elementos que formaram as sementes iniciais. Essa escolha pode ser feita de muitas formas, entre elas: selecionando as k primeiras observações; selecionando k observações aleatoriamente; e escolhendo k observações de modo que seus valores sejam bastante diferentes. Por exemplo, ao se agrupar uma população em três grupos de acordo com a inteligência predominante, seria possível escolher um indivíduo de inteligência Lógico-Matemática, um de linguística e um espacial visual, entre outros.

Escolhidas as sementes iniciais, é calculada a distância de cada elemento em relação às sementes, agrupando o elemento ao grupo que possui a menor distância (mais similar) e recalculando o centroide do mesmo. O processo é repetido até que todos os elementos façam parte de um dos clusters.

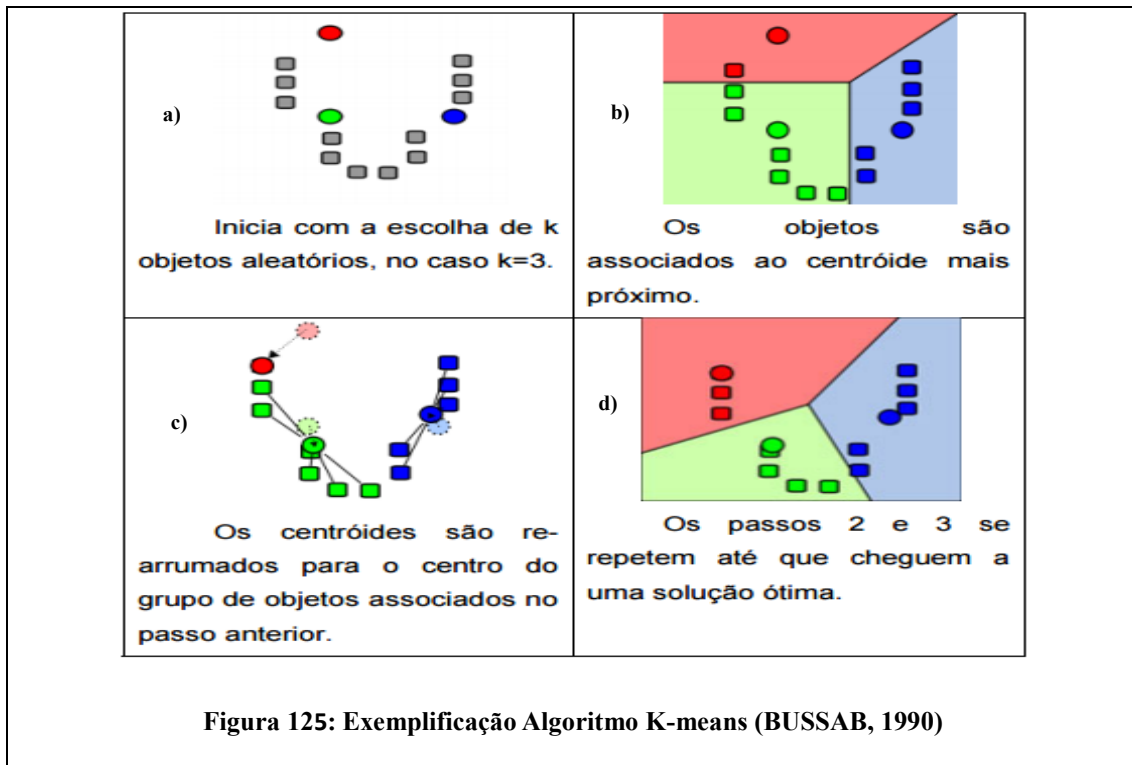
A ideia por trás deste algoritmo é escolher k objetos (aleatoriamente ou com alguma heurística) que serão à base de cada grupo (denominados centroides), os demais objetos são associados ao centroide mais próximo. A cada passo, os centroides são recalculados dentre os objetos de seu próprio grupo e estes são realocados para o centroide mais próximo. Esse procedimento é repetido até que o nível de convergência seja satisfatório de acordo com alguma heurística estabelecida.

O pseudocódigo do algoritmo é descrito, como:

| Algoritmo K-means |
|--|
| 1- Selecione K pontos como centroides iniciais |
| 2- Repita |
| 3- Atribua cada objeto ao cluster mais próximo |
| 4- Recalcula cada centroide de cada cluster |
| 5- Até que (até que os centroides permaneçam estáveis) |

Figura 14: Algoritmo K-means (BUSSAB, 1990)

Na figura 15 é possível ver uma simulação do algoritmo:



Assim, no K-means, o centróide é dado pela média de todos os objetos dentro de um agrupamento.

3.6 Aplicação de Métodos de Decomposição de problemas multiclasse

Para este trabalho são criadas duas SVMs não lineares e multiclases, com propósitos e métodos de decomposição distintos:

- A classificação da primeira SVM, é do tipo Multiclases usando o método de *k* problemas binários, sendo, para esse caso, o de classificação por *Decomposição Um por Classe*.
- A classificação da segunda SVM, é do tipo Multiclases usando o método de *k* problemas binários, sendo, para esse caso, o de classificação por *Decomposição Duas a Duas*.

Para ambas SVMs é aplicada a função de *kernel* polinomial, tornando mais simples o cálculo e aprimorando a capacidade de representar espaços muito abstratos.

$$K(x_i, x_j) = ((x_i \cdot x_j) + a)^p$$

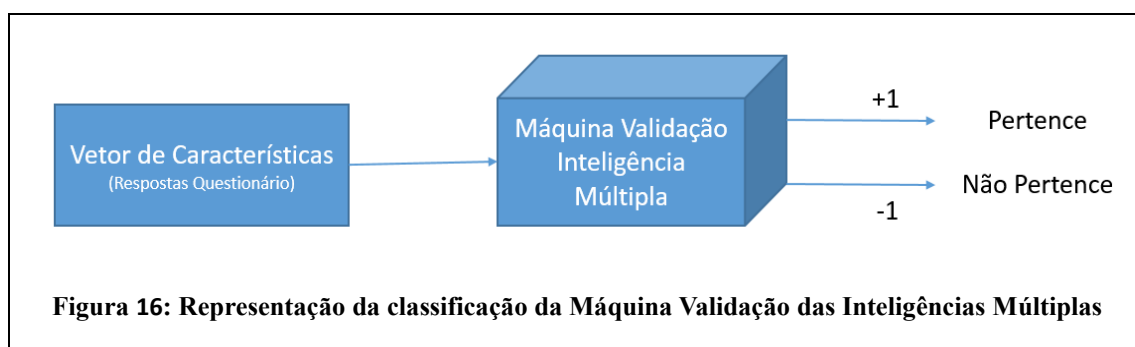
Na fase de treinamento, foi aplicado o balanceamento de classes em ambas SVM's por meio do filtro Resample.

Na fase de teste, é aplicado o conceito do método de Votação-distribuída em ambas SVM's, o que foi definido como método de agrupamento, com objetivo de facilitar entendimento dos resultados, por meio de percentuais gerados pelo algoritmo SimpleKMeans.

Os problemas de classificação que as SVMs visam solucionar neste trabalho são:

SVM para validação das Inteligências de um indivíduo: considerando que a SVM é um método binário, serão recebidas três Vetores de Características compostos pelas respostas do questionário aplicado durante o experimento, sendo a inteligência do indivíduo validada apenas em duas classes: *Pertence* e *Não Pertence*. Vale ressaltar, que este contexto deve ser aplicado para cada uma das sete inteligências deste experimento.

Têm-se a representação da classificação a partir das características extraídas na figura 16:



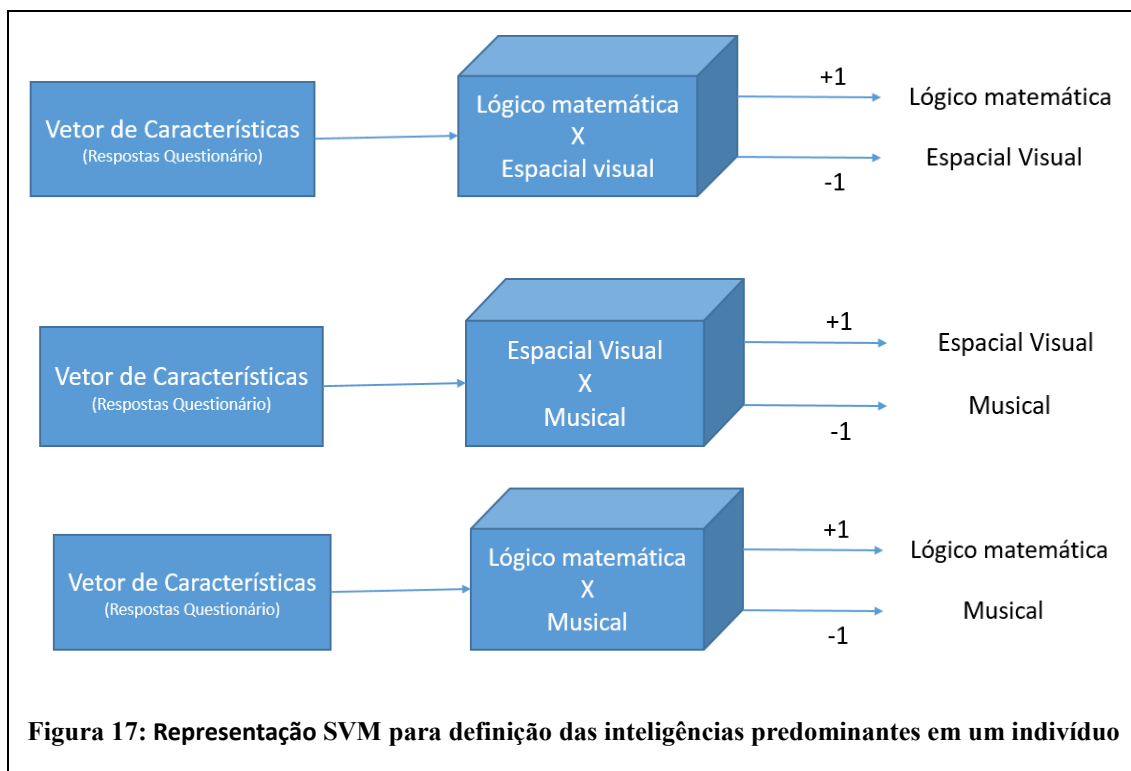
Para a classificação desta SVM são utilizados três vetores de características, por exemplo, (R1, R2 e R3), definidos como entrada para fase de treinamento e testes. Assim, se o algoritmo obtiver o valor +1, a Inteligência está contida na classe *Pertence*, ou na classe *Não Pertence* caso o valor retorne -1. Os resultados obtidos pela utilização de cada vetor de característica são analisados para verificar quais entre as sete inteligências obteve melhores resultados na classificação, assim, tornando-as aptas para a próxima etapa, em que serão treinadas na segunda SVM. Portanto, cada uma das sete inteligências é classificada entre as classes *Pertence* e *Não Pertence*.

SVM para definição das inteligências predominantes em um indivíduo: essa SVM indicará quais dentre os melhores resultados classificados pela primeira SVM são as inteligências múltiplas com maior predominância em um indivíduo. Para isso, são determinados dois vetores

de características compostos pelas respostas do questionário. São treinados dois vetores por vez, até que se obtenha os melhores resultados. Assim, a máquina classificará de acordo com o vetor de características a que obtiver melhores resultados, indicando qual a inteligência múltipla é a de maior importância, ou seja, predominante.

Para exemplificar o experimento, foram selecionadas as respostas de um dos indivíduos participantes. Essas respostas irão compor o vetor de características e, por meio desta SVM, será classificada a inteligência predominante entre Lógico-Matemática, Musical e Espacial Visual.

Têm-se a representação da classificação a partir das características extraídas na figura 17:



O resultado da primeira classificação indicou a inteligência múltipla Lógico-Matemática. Já na segunda classificação foi a inteligência múltipla Musical. Por fim, na classificação entre as de maior predominância, a inteligência múltipla Lógico-Matemática foi classificada como a de maior predominância. Assim, neste exemplo é possível identificar que as duas inteligências múltiplas de maior predominância são: Lógico-Matemática e Musical.

É importante destacar que na fase de teste são gerados os percentuais, conforme descrito no início desta seção. Portanto, este trecho é apenas um exemplo para entender o funcionamento desta SVM.

Capítulo 4- Experimentos

Neste capítulo são apresentados os experimentos e os detalhes de sua realização. Além disso, são apresentadas análises dos resultados.

4.1 Descrição do experimento

Para realizar o experimento, foi desenvolvido um questionário com perguntas de todos os tipos de inteligências definidas por Gardner [14, 15].

Este questionário (APÊNDICE A), fundamentado na Teoria das Inteligências Múltiplas, foi elaborado pelo próprio autor e contém três questões para cada uma das sete inteligências, totalizando 21 questões, uma vez que não é considerada a inteligência naturalista. As questões foram trabalhadas de forma a contextualizar situações que aplicassem as características das inteligências de acordo com a teoria elaborada por Gardner.

O questionário original (APÊNDICE B), contém setenta questões, sendo dez questões para cada Inteligência Múltipla com quatro possibilidades de resposta. As opções são: agradável, pouco agradável, pouco desagradável e desagradável.

As Inteligências Múltiplas analisadas no questionário original são as mesmas que as utilizadas neste trabalho. São elas: linguística, lógico-matemática, musical, corporal sinestésica, espacial visual, interpessoal e intrapessoal, sendo estas relacionadas aos aspectos cognitivos que serão utilizados como parâmetros de classificação dos indivíduos.

A diminuição no número de questões de setenta para vinte e uma reduz o tempo para responder o questionário. Essa alteração foi realizada após a apresentação do questionário a pessoas escolhidas para os testes, tendo estas realizado críticas ao fato de aquele ser extenso e cansativo de se responder.

Com relação às possibilidades de resposta para as questões, os indivíduos concordaram que quatro níveis de resposta são suficientes para expressar a sua opinião em relação a cada questão.

O questionário foi elaborado no Microsoft Excel e aplicado a alunos das áreas de Humanas, Exatas e Biológicas dos cursos do Centro Paula Souza em virtude da viabilidade do acesso às informações e da necessidade de se obter os tipos de inteligência que foram estimulados ao longo do curso. A amostra utilizada foi composta pelos respondentes do último período dos cursos de Informática, Manutenção de Computadores, Informática para Internet Administração, Enfermagem, Zootecnia e Recursos Humanos. A autorização para aplicação dos questionários foi concedida informalmente pelos coordenadores dos respectivos cursos.

Todos os indivíduos responderam o questionário, cada um em um computador sem influência dos demais indivíduos ou do aplicador e não foi determinado um tempo limite para responder. Antes do início foram passadas as informações sobre como deveria ser realizado o preenchimento das respostas e como o gabarito deveria ser salvo. O teste foi disponibilizado por meio de planilha desenvolvida com a ferramenta Microsoft Excel disponibilizada aos indivíduos em seus computadores. Assim, o tempo médio que os indivíduos levaram para responder o questionário foi de 25 minutos.

Os dados foram coletados entre novembro de 2015 e março de 2016, sendo que a amostra se caracterizou naqueles que estavam presentes em sala de aula nos dias em que os questionários foram aplicados.

O número de questionários respondidos por curso, está apresentado na tabela 5:

Tabela 5: Coleta de Amostras

| Área | Curso | Questionários Respondidos | Total de Alunos Matriculados por Curso |
|--------------|----------------------------|----------------------------------|---|
| Exatas | Informática para Internet | 23 | 30 |
| Exatas | Manutenção de Computadores | 12 | 14 |
| Exatas | Informática | 35 | 37 |
| Biológicas | Enfermagem | 59 | 68 |
| Biológicas | Zootecnia | 21 | 28 |
| Humanas | Administração | 16 | 21 |
| Humanas | Recursos Humanos | 23 | 30 |
| Total | | 189 | 228 |

O total é dividido em duas partes:

- ✓ Total de questionários respondidos
- ✓ Total de alunos matriculados

4.1.1 Análise Estatística

O questionário empregado tem como base na proporção de respostas, ou seja, o resultado para cada um dos aspectos cognitivos é uma proporção de itens assinalados que variam de 0 a 100% e um plano amostral teve que ser desenvolvido a partir da disponibilidade dos indivíduos em responder ao questionário. Por meio da tabela 6, é possível observar as informações que representam o espaço amostral.

Tabela 6: Informações sobre espaço amostral

| Área de Conhecimento | Cursos | Total de Alunos | Nº Necessário de Alunos | Nº de Alunos analisados |
|----------------------|----------------------------|-----------------|-------------------------|-------------------------|
| Exatas | Informática para Internet | 30 | 21 | 23 |
| Exatas | Manutenção de Computadores | 14 | 12 | 12 |
| Exatas | Informática | 37 | 23 | 35 |
| Biológicas | Enfermagem | 68 | 34 | 59 |
| Biológicas | Zootecnia | 28 | 20 | 21 |
| Humanas | Administração | 21 | 16 | 16 |
| Humanas | Recursos Humanos | 30 | 21 | 23 |

4.2 Análise dos Resultados

Na tabela 7 são apresentados os percentuais dos respondentes em relação ao total de amostras e com isso pode-se ver que o curso Enfermagem é o de maior representatividade com 18,5% das amostras. Outra informação identificada é que o curso de maior representatividade pertence à área de Biológicas.

A área que obteve maior percentual de representatividade nas amostras foram as biológica com 42,5%, seguida por exatas 37% e humanas 20,5%.

Tabela 7: Percentual de respondentes por curso

| Curso | Quantidade de Respondentes | Percentual de respondentes em relação a amostra |
|----------------------------|-----------------------------------|--|
| Informática para Internet | 23 | 12% |
| Manutenção de Computadores | 12 | 6,5% |
| Informática | 35 | 18,5% |
| Enfermagem | 59 | 31% |
| Zootecnia | 21 | 11,5% |
| Administração | 16 | 8,5% |
| Recursos Humanos | 23 | 12% |
| Total | 189 | 100% |

Na tabela 8, é possível observar quais dentre os sete tipos de inteligências trabalhadas no experimento entre os cursos pesquisados são as de maior incidência nos alunos.

Tabela 8: Percentuais Inteligências por Curso

| Inteligências | Cursos | | | | | | |
|---------------------------|--------------------|-----------------|-------|------------|-----------|---------|-----|
| | Inf. para Internet | Manut. de Comp. | Info. | Enfer. | Zootecnia | R.H | ADM |
| | Exatas | | | Biológicas | | Humanas | |
| Linguística | - | 16% | 19% | 17% | 10% | 29% | 31% |
| Lógico-Matemática | 48% | 25% | 24% | 8% | 15% | 22% | 25% |
| Espacial visual | 18% | 17% | 14% | - | 20% | 21% | - |
| Corporal cenestésica | 4% | - | 5% | 17% | 15% | - | 13% |
| Musical | 26% | 25% | 14% | - | 15% | 21% | 19% |
| Interpessoal | 4% | 17% | 14% | 33% | 15% | 7% | 12% |
| Intrapessoal | - | - | 10% | 25% | 10% | - | - |
| Total Respondentes | 23 | 12 | 35 | 59 | 21 | 23 | 16 |

Conforme dados da tabela 8, é possível identificar que a de maior incidência nos cursos de exatas foi a inteligência Lógico-matemática, nos de Humanas, Linguística, em Enfermagem, interpessoal e, em Zootecnia, Espaço visual.

Conclusão

Neste trabalho, após experimentos realizados, pôde-se notar que é possível classificar os dados coletados de acordo com as inteligências múltiplas. Como proposta de trabalhos futuros, esse mecanismo poderá ser adaptado em ferramentas de apoio ao ensino.

A integração desse classificador com uma ferramenta de apoio ao ensino à distância contribuirá para melhorar a qualidade dos testes realizados e isso deverá fornecer mais parâmetros para validar esta proposta. Porém, para essa integração, será necessário melhorar os métodos de treinamento do classificador devido ao aumento do número de amostras, bem como ajustar o formulário das questões a serem respondidas pelos indivíduos, realizar mais testes e simulações para que uma análise mais detalhada possa ser realizada e conclusões mais significativas possam ser obtidas.

Um dos objetivos desta classificação consiste em permitir que os tutores responsáveis por gerar o material para as aulas em ferramentas de apoio ao ensino à distância possam utilizar este mecanismo de classificação para direcionar o conteúdo ao aluno de forma a explorar sua inteligência múltipla predominante, com a tentativa de potencializar o ensino, permitindo que o aluno possa realizar os estudos por meio de técnicas que poderão facilitar o aprendizado de forma individualizada. No entanto, esta ferramenta de classificação pode ser aplicada a vários outros sistemas de recomendação e não somente ao proposto, sendo necessário ajustar os parâmetros de entrada para que se possa obter os parâmetros de saída desejados.

Para realizar o treinamento da SVM, foi necessário definir o método de classificação baseado em k problemas binários, sendo este um método de classificação para múltiplas classes. Além disso, para gerar os percentuais das inteligências, foi aplicado o algoritmo SimpleKMeans, assim, tornando a interpretação dos resultados mais fácil de ser interpretada.

O uso da ferramenta WEKA e seu ambiente gráfico, facilitou a configuração do algoritmo de treinamento, que utilizou recursos da biblioteca LibSVM. O algoritmo foi engendrado com dados reais coletados por meio de um questionário baseado no de Gardner e aplicado em instituições de ensino público que pertencem ao Centro Paula Souza.

Os resultados foram analisados e por meio dos percentuais gerados nos treinamentos, pôde-se identificar o perfil de cada aluno que participou do experimento, bem como, as

inteligências predominantes por área e curso. Portanto, esses dados mostraram em percentuais a quantidade de predominância das inteligências múltiplas nos alunos.

Analisando os resultados do treinamento, foi possível notar que, aplicando o filtro para balanceamento de amostras *Resample*, obteve-se ganho de até 15% nas taxas de precisão nos melhores casos de treinamento. Com isso, os índices de incorreções variaram entre 5 a 10%. Em alguns treinamentos as taxas de incorreções se aproximaram a 1%. Portanto, pode-se considerar que o mecanismo de classificação proposto neste trabalho é eficaz.

A partir da nova coleta de amostras, realiza-se uma integração com ferramenta de apoio ao ensino, ajustando os sistemas de recomendação. Com isso, pode-se realizar novas formulações para proporcionar novas análises dos dados que comprovem a eficácia do mecanismo de classificação.

Para trabalhos futuros, propõe-se a realização de novos estudos e a tentativa de aprimorar o questionário, que talvez possa ser utilizado como escolha padrão nesse tipo de teste.

Por fim, uma das maiores dificuldades para realização dos experimentos foi conseguir uma base de dados. Portanto, as amostras coletadas para este experimento podem servir como base para outros trabalhos similares.

Referências

- [1] ARMSTRONG, T. J. Inteligências Múltiplas na Sala de Aula. ARTMED, Porto Alegre, 2001. 192p.
- [2] ANDREOLA, R.; HAERTEL, V. Classificação de Imagens Hiperespectrais Empregando Support Vector Machines **Boletim de Ciências Geodésicas**. Curitiba, v. 16, n. 2, 2010.
- [3] BURKE, R. Integrating knowledge-based and collaborative-filtering recommender systems. In: **Proceedings of the Workshop on AI and Electronic Commerce**.1999. p. 69-72.
- [4] BUSSAB, W. Introdução à análise de agrupamentos. ABE, p. 105, 1990.
- [5] CHANG, C.; LIN, C. {LIBSVM}: A Library for Support Vector Machines (Version 2.3). 2001.
- [6] COSTA NETO, A. Ambiente virtual de apoio ao ensino com ênfase na Teoria das Inteligências Múltiplas e sua aplicação em sistemas digitais. 2009 – Dissertação – Universidade Estadual Paulista – Unesp – São José do Rio Preto – SP – Brasil [75] BALDI, Pierre; LOPES, Crista. The Universal Campus: An open virtual 3-D world infrastructure for research and education.eLearn, v. 2012, n. 4, p. 6, 2012.
- [7] DE ALMEIDA, M.B.; DE PÁDUA BRAGA, A.; BRAGA, J.P.SVM-KM: speeding SVMs learning with a priori cluster selection and k-means. p. 162-167In: Neural Networks, 2000. Proceedings. **Sixth Brazilian Symposium** on. IEEE, 2000.
- [8] DOS SANTOS, T. F. S.; BOTTENTUIT, J. B. J.: Perspectivas da Teoria das Inteligências Múltiplas para a Educação a Distância: um estudo no curso de licenciatura em pedagogia do NEAD-UFMA. Revista Educação Online, São Luís– ISSN 1983-2664 – volume 9 – jan. 2015.
- [9] FAYADD, U.; PIATESKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. AAAI. 1996.
- [10] FELFERNIG, A; BURKE, R. Constraint-based recommender systems: technologies and research issues. In: **Proceedings of the 10th international conference on Electronic commerce**. ACM, p. 3, 2008.
- [11] FREUND, Y.; IYER, R.; SCHAPIRE, R.E., Singer, Y. An efficient boosting algorithm for combining preferences. **The Journal of Machine Learning Research**, 4:933–969, 2003.
- [12] FÜRNKRANZ, J. Round robin classification. **The Journal of Machine Learning Research**, v. 2, p. 721-747, 2002.
- [13] FÜRNKRANZ, J. Round robin ensembles. Intelligent Data Analysis, v. 7, n. 5, p. 385-403, 2003.

- [14] GARDNER, H. Estruturas da Mente: A Teoria das Inteligências Múltiplas. Artes Médicas. Porto Alegre, p. 340, 1994.
- [15] GARDNER, H. Inteligências Múltiplas: A Teoria na Prática. Artes Médicas. Porto Alegre, 1995.
- [16] GIL, A. C.; MÉTODOS, C. Técnicas de pesquisa social. São Paulo. 1999.
- [17] GIRARDELLO, A. D. Um Estudo sobre o Uso de Máquinas de Vetores de Suporte em Problemas de Classificação. 2010. 75 f. TCC (Graduação) - Curso de Curso de Bacharelado em Ciência da Computação, Colegiado de Ciência da Computação, Universidade Estadual do Oeste do Paraná, Cascavel, 2010. Disponível em: <<http://www.inf.unioeste.br/~tcc/2010/TCC-AdrianoDouglasGirardello.pdf>>. Acesso em: 01 fev. 2016.
- [18] HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1, 2009.
- [19] HAN, J.; KAMBER, M. Data Mining. Concepts and Techniques. Second edition. The Morgan Kaufmann Series in Data Management Systems. Elsevier Inc., 2006.
- [20] HASTIE, T. et al. Classification by pairwise coupling. The annals of statistics, v. 26, n. 2, p. 451-471, 1998.
- [21] HAYKIN, S. S Neural Networks -A Comprehensive Foundation. Prentice-Hall, New Jersey, 2nd edition, p. 842, 1999.
- [22] JO, T.; JAPKOWICZ, N. Class imbalances versus small disjuncts. ACM SigkddExplorations Newsletter, v. 6, n. 1, p. 40-49, 2004.
- [23] KONSTAN, J. A. et al. Group Lens: applying collaborative filtering to Usenet news. **Communications of the ACM**, v. 40, n. 3, p. 77-87, 1997.
- [24] MAHMOOD, T.; RICCI, F. Improving recommender systems with adaptive conversational strategies. In: **Proceedings of the 20th ACM conference on Hypertext and hypermedia**. ACM, p. 73-82, 2009.
- [25] MAYORAZ, E.; MOREIRA, M. On the decomposition of polychotomies into dichotomies. In: **Proceedings of the Fourteenth International Conference on Machine Learning**. Morgan Kaufmann, 1997.
- [26] MITCHELL, T. M. Machine Learning. McGraw Hill, p.414, 1997.
- [27] MONARD, M. C.; BARANAUSKASJ. A. Conceitos de aprendizado de máquina. In S. O. Rezende, editor, Sistemas Inteligentes - Fundamentos e Aplicações, p.89–114. Editora Manole, 2003.
- [28] MOREIRA, L. M. The use of Boolean concepts in general classification contexts. p. 173, 2000.

- [29] MÜLLER, K.-R. et al. An introduction to kernel-based learning algorithms. *Neural Networks, IEEE Transactions on*, v. 12, n. 2, p. 181-201, 2001.
- [30] OLIVEIRA, R. B. Método de detecção e classificação de lesões de pele em imagens digitais a partir do modelo Chan-Vese e máquina de vetor de suporte. 2012. 134 f. Dissertação (Mestrado) - Curso de Programa de Pós-graduação em Ciência da Computação, Departamento de Ciência da Computação e Estatística, Universidade Estadual Paulista, São José do Rio Preto, 2012. Cap. 6. Disponível em: <http://repositorio.unesp.br/bitstream/handle/11449/98657/oliveira_rb_me_sjrp.pdf?sequence=1&isAllowed=y>. Acesso em: 01 fev. 2016. (antigo 51)
- [31] PAZZANI, M. J. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, v. 13, n. 5-6, p. 393-408, 1999.
- [32] REATEGUI, E. B.; CAZELLA, S. C. Sistemas de Recomendação. **XXV Congresso da Sociedade Brasileira de Computação. Universidade do Vale do Rio dos Sinos (UNISINOS)**. São Leopoldo, 2005.
- [33] RICCI, F.; ROKACH, L.; SHAPIRA, B.; KANTOR, P. B. *Recommender Systems Handbook*. Springer Science Business Media, New York, 2011.
- [34] RICCI, F.; ROKACH, L.; SHAPIR A, B. *Introduction to recommender systems handbook*. Springer US, 2010.
- [35] SANTOS, R. Weka na Munheca: Um guia para uso do Weka em scripts e integração com aplicações em Java. *Apostila Princípios e Aplicações de Mineração de Dados*. SI, 2005.
- [36] SEMOLINI, R. Support Vector Machines, Inferência Transdutiva e o Problema de Classificação. 2002. 142 f. Dissertação (Mestrado) - Curso de Programa de Mestrado em Engenharia Elétrica, Departamento de Engenharia de Computação e Automação Industrial, Universidade Estadual de Campinas, Campinas, 2002. Cap. 8. Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/semolini_mest/semolini_tese_mest.pdf>. Acesso em: 01 fev. 2016.
- [37] SILVA, I. A. F. Descoberta de Conhecimento em Base de Dados de Monitoramento Ambiental para Avaliação da Qualidade da Água. Programa de Pós-Graduação em Física e Meio Ambiente. Universidade Federal de Mato Grosso, Cuiabá, 2007.
- [38] SHENG, V. S.; LING, C. X. Feature value acquisition in testing: a sequential batch test algorithm. In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pp. 809-816, 2006.
- [39] SHENG, V. S.; LING, C. X.; NIA.; ZHANG, S. Cost-Sensitive Test Strategies. In: **Proceedings of the 21st National Conference on Artificial Intelligence**, 2006.
- [40] SMOLA, A. J.; SCHÖLKOPF, B. *Learning with kernels*. GMD - Forschungszentrum Informationstechnik, 1998.
- [41] SOUTO, M. C. P.; LORENA A. C.; DELBEMA. C. B.; CARVALHOA. C. P. L. F. Técnicas de Aprendizado de Máquina para problemas de Biologia Molecular, p.103–152. *Minicursos de*

Inteligência Artificial, **Jornada de Atualização Científica em Inteligência Artificial, XXIII Congresso da Sociedade Brasileira de Computação**, 2003.

[42] STEINLEY, D. K-Means Clustering: A Half-Century Synthesis. *British J. of Mathematical and Stat. Psychology*, V. 59, 2006.

[43] TRIOLA, M.F. et al. *Introdução à estatística*. Rio de Janeiro: Ltc, 2005.

[44] VAPNIK, V. N. *Statistical learning theory*. New York: Wiley, p. 736, 1998.

[45] VAPNIK, V. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[46] VERT, J.-P. *Introduction to Support Vector Machines and Applications to Computational Biology*. (draft). 2001.

[47] WEISS, S. M.; KULIKOWSKI, C. A. *Computer Systems That Learn: Classification and Prediction Methods from Statistics. Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann Publishers, CA, p. 223, 1991.

[48] WEISS, G. M.; PROVOST, F. Learning when training data are costly: the effect of class distribution on tree induction. ***Journal of Artificial Intelligence Research***, p. 315-354, 2003.

[49] WITTEN, I.H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[50] WU, X.; KUMAR, V. *The Top Ten Algorithms in Data Mining*. Chapman & Hall/CRC, p. 208, 2009.

[51] ZADROZNY, B.; LANGFORD, J.; ABE, N. Cost-Sensitive Learning by Cost-Proportionate Example Weighting. In: **Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)**, 2003.

APÊNDICE A – Questionário

A seguir, consta a tabela com as questões classificadas de acordo com as inteligências definidas por Gardner e adaptadas de seu questionário.

| QUESTÕES | |
|-------------------|--|
| INTELIGÊNCIA | PERGUNTA |
| Musical | Faça uma lista das músicas que você costuma ouvir e associe o ritmo musical ao seu estado de espírito. As músicas que você ouve representam esse estado? |
| Musical | Ao estudar, ler ou realizar alguma atividade em casa ou no trabalho, com que frequência você faz esta atividade ouvindo alguma música de fundo ou mesmo cantando alguma? |
| Musical | Quando você está ouvindo música e percebe os sons gerados por instrumentos musicais, com que frequência você sente vontade de tocar algum instrumento musical, mesmo? |
| Lógico-Matemática | Quando você precisa resolver problemas que envolvem cálculos matemáticos, você prefere tentar resolver mentalmente ou procura alguma ferramenta de cálculo que facilite? |

| | |
|-------------------|---|
| Lógico-Matemática | Você costuma guardar seu dinheiro e ir gastando ele aos poucos comprando somente o necessário e guardando o restante para uma emergência? |
| Lógico-Matemática | Quando você vai realizar alguma tarefa sozinho (a), você costuma ser detalhista mesmo que isso possa fazer você perder muito tempo em detalhes pequenos? |
| Interpessoal | Quando você percebe que alguma pessoa da família, ou um amigo mais próximo ou até mesmo um desconhecido não está bem, estando ele com aparente tristeza, com que frequência você procura saber o que está acontecendo e se dispõe a ajudar? |
| Interpessoal | Se você fosse um atleta profissional e tivesse que escolher um esporte para praticar, com que frequência você iria escolher esportes coletivos, tendo mais chances de ser reconhecido e beneficiado da ajuda dos outros atletas? |
| Interpessoal | Quando você não está se sentindo bem emocionalmente, com que frequência você comenta seus problemas para algum amigo (a) e permite que ele dê conselhos a você? |
| Espacial | Quando você vai ler revistas, livros ou até mesmo visitar sites na Internet, com que frequência você olha primeiro as imagens e depois lê o texto? |
| Espacial | Quando você está na sala de aula e o professor esta explicando a matéria utilizando gestos e palavras, se o professor gesticular demais, |

| | |
|----------------------|---|
| | com que frequência isso acaba distraindo você? |
| Espacial | Quando você viaja para lugares que você ainda não conhece, com que frequência você repara em detalhes nos formatos de árvores, cores e formas geométricas das casas, modelos dos carros, entre outras coisas? |
| Corporal cenestésica | Ao consertar ou montar equipamentos e objetos, com que frequência você prefere tentar montar de acordo com a forma que considera ser a melhor deixando de lado o manual de instruções? |
| Corporal cenestésica | Com que frequência você se oferece como voluntário (a) para realizar tarefas manuais, tais como utilizar chaves de fenda, cortar arames com alicate, pregar pregos, cortar papéis, lavar a louça, passar roupa, entre outras? |
| Corporal cenestésica | Quando o professor está explicando um conteúdo novo e utiliza a fundamentação teórica, com que frequência isso é o suficiente para que você possa aprender e resolver os exercícios propostos ao final da aula? |
| Intrapessoal | Se você tivesse que anotar tudo que se passa na sua vida em um diário para que um dia pudesse se lembrar de tudo que já fez e deixar anotações para outras pessoas. Com que frequência você anotaria as coisas? |
| Intrapessoal | Quando você não está bem emocionalmente e alguma coisa o está incomodando, com que frequência você prefere estar junto de várias |

| | |
|--------------|---|
| | <p>peças independentemente do que estejam conversando, mesmo que seja sobre você?</p> |
| Intrapessoal | <p>Com que frequência você prefere realizar tarefas que o deixam feliz independentemente de qual seja, mesmo sabendo que existem outras que você poderia realizar que não lhe fariam mal, mas também são necessárias?</p> |
| Linguística | <p>Para se manter atualizado com relação as notícias do cotidiano, com que frequência você busca notícias escritas em jornais, revistas e na Internet?</p> |
| Linguística | <p>Quando você está em ambientes com várias pessoas e os assuntos vão surgindo, você gosta de debater mesmo que não tenham perguntado sua opinião, considerando assim, que sua opinião é importante?</p> |
| Linguística | <p>Suponha que você esteja participando de um concurso musical e você tem facilidade em cantar músicas nacionais e em outro idioma independentemente de qual seja. Com que frequência você faria a opção por cantar músicas internacionais?</p> |

APÊNDICE B – Questionário Gardner

| Multiple Intelligences Test |
|---|
| I like to learn more about myself |
| I can play a musical instrument |
| I find it easiest to solve problems when I am doing something physical |
| I often have a song or piece of music in my head |
| I find budgeting and managing my money easy |
| I find it easy to make up stories |
| I have always been physically well co-ordinated |
| When talking to someone, I tend to listen to the words they use not just what they mean |
| I enjoy crosswords, word searches or other word puzzles |
| I don't like ambiguity, I like things to be clear |
| I enjoy logic puzzles such as 'sudoku' |
| I like to meditate |
| Music is very important to me |
| I am a convincing liar |
| I play a sport or dance |
| I am very interested in psychometrics (personality testing) and IQ tests |

| |
|---|
| People behaving irrationally annoy me |
| I find that the music that appeals to me is often based on how I feel emotionally |
| I am a very social person and like being with other people |
| I like to be systematic and thorough |
| I find graphs and charts easy to understand |
| I can throw things well - darts, skimming pebbles, frisbees, etc |
| I find it easy to remember quotes or phrases |
| I can always recognise places that I have been before, even when I was very young |
| I enjoy a wide variety of musical styles |
| When I am concentrating I tend to doodle |
| I could manipulate people if I choose to |
| I can predict my feelings and behaviours in certain situations fairly accurately |
| I find mental arithmetic easy |
| I can identify most sounds without seeing what causes them |
| At school one of my favourite subjects is / was English |
| I like to think through a problem carefully, considering all the consequences |
| I enjoy debates and discussions |
| I love adrenaline sports and scary rides |
| I enjoy individual sports best |
| I care about how those around me feel |
| My house is full of pictures and photographs |
| I enjoy and am good at making things - I'm good with my hands |
| I like having music on in the background |
| I find it easy to remember telephone numbers |
| I set myself goals and plans for the future |
| I am a very tactile person |
| I can tell easily whether someone likes me or dislikes me |

| |
|--|
| I can easily imagine how an object would look from another perspective |
| I never use instructions for flat-pack furniture |
| I find it easy to talk to new people |
| To learn something new, I need to just get on and try it |
| I often see clear images when I close my eyes |
| I don't use my fingers when I count |
| I often talk to myself – out loud or in my head |
| At school I loved / love music lessons |
| When I am abroad, I find it easy to pick up the basics of another language |