



UNIVERSIDADE FEDERAL DE PERNAMBUCO

PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA

PROJETO DE PESQUISA – MESTRADO ACADÊMICO

ALUNO: PAULO FAGNER TENÓRIO BARROS DE MORAIS

ORIENTADOR: PROF. PAULO JORGE LEITÃO ADEODATO

Recife, março de 2012.

Recife (PE), 01 de Março de 2012.

Título Provisório :

**Gerção Dinâmica de Comitês de Classificadores Através da Ordenação de
Competências e Estabelecimento de Critério de Corte**

Resumo

O presente trabalho visa realizar um estudo sobre as técnicas de combinação de classificadores, em problemas de classificação, a teoria envolvida e a motivação para o emprego de abordagens baseada em comitês de classificadores. O projeto se propõe a apresentar uma metodologia para a geração dinâmica de comitês, onde a cada padrão de testes é atribuído, na fase de generalização, um comitê de classificadores. A abordagem proposta é baseada na definição e cálculo de regiões de competência, que, em conjunto com uma adaptação de outro algoritmo de seleção dinâmica existente, definem um critério de escolha do comitê.

Sumário

1. Introdução.....	1
2. Revisão de Literatura: Conceitos e Abordagens Existentes.....	2
2.1. Abordagens Existentes na Geração de Comitês de Classificadores.....	3
2.1.1. Abordagens e Categorias.....	4
3. Objetivos do Trabalho.....	8
4. Método Proposto.....	8
5. Metodologia.....	10
6. Resultados Esperados.....	10
7. Cronograma e Estado Atual da Pesquisa.....	11
8. Referências Bibliográficas.....	12

1. Introdução

Em problemas de classificação, uma das áreas que vem ganhando crescente atenção da comunidade de aprendizagem de máquina nos últimos anos é a combinação de classificadores via comitês, metodologia também conhecida por sistemas de múltiplos classificadores (*MCS – Multiple Classifier Systems*) ou conjunto de classificadores (*EoC – Ensemble of Classifiers*). A idéia é relativamente simples e é capaz de aumentar a capacidade de generalização de sistemas inteligentes (DIETTERICH, T., 2000).

Muitas vezes em nossas vidas, quando estamos diante de uma situação que exige uma tomada de decisão de grande impacto, buscamos recorrer a opiniões de especialistas no assunto em questão, de forma a embasar, com opiniões sensatas e fundamentas, nossa tomada de decisão. Buscamos, ainda, colher opiniões fundamentadas em diferentes pontos de vistas, ou opiniões de especialistas de áreas complementares. Como exemplo, no caso da compra de um imóvel, podemos recorrer a opinião de um corretor imobiliário que nos informará se o preço cobrado está de acordo com a realidade do momento; a opinião de um economista informará se a compra representa um bom investimento a médio prazo; um advogado poderá ser útil nos esclarecimentos sobre desembaraços legais; um engenheiro civil poderá emitir sua opinião sobre a solidez da estrutura edificada ou outros problemas aparentes; um morador da região nos informará sobre a segurança na vizinhança, serviços existentes no bairro, comércio etc. As opiniões emitidas tem caráter complementar e, sem dúvida, contribuirão para que a decisão tomada dificilmente venha a se converter em um arrependimento.

Em outras situações, podemos basear o processo de tomada de decisão a partir da reunião de um grupo de pessoas. Os membros deste grupo devem possuir um determinado nível de conhecimento sobre o tema em questão, e, ao mesmo tempo, devem deter opiniões diversas. Com estes pressupostos, espera-se que o grupo venha contribuir para um debate que possa listar os principais pontos relacionados a cada opção componente do espaço de escolhas.

De maneira análoga, comitês de classificadores buscam extrair informações de naturezas complementares e diversas sobre o problema. A geração de um comitê diverso, onde cada hipótese formulada por cada classificador é tão diferente das demais quanto possível, de forma a não perder a consistência com o conjunto de treinamento é um propriedade importante para a geração de um bom comitê (KROGH, A.; VEDELSBY, J., 1995).

É essencial que os erros cometidos pelos classificadores membros do comitê (componentes) possuam natureza complementar; do contrário, a utilização de um único classificador seria uma decisão mais sensata.

2. Revisão da Literatura: Conceitos e Abordagens Existentes

Comitês de classificadores, em geral, apresentam uma melhor precisão que seus classificadores-base (componentes) se os erros individualmente cometidos por estes classificadores são descorrelacionados (HANSEN; SALAMON, 1990). Este pressuposto tem um apelo intuitivo, no caso de um problema de classificação: caso a decisão venha a ser tomada via votação, pressupondo que os erros sejam descorrelacionados, há uma boa chance de que os erros individualmente cometidos por um classificador, na rotulação de determinados padrões de teste, sejam compensados pelos acertos de outros classificadores na tarefa de rotulação daqueles mesmos padrões. Esta compensação de erro se deve ao poder de que componentes diferentes possuem de representar aspectos distintos e, ao mesmo tempo, relevantes para a resolução problema.

O paradigma de combinação de classificadores tem origem no trabalho de Hansen e Salamon (1990), onde a conclusão fundamental daquele trabalho é o aumento da capacidade de generalização de um sistema a partir da combinação, via voto, de redes neurais treinadas isoladamente em uma mesma base de dados. De grande relevância para na área é o trabalho de Schapire (1990), que mostra ser possível a construção de um sistema de classificação forte a partir de classificadores fracos. Este trabalho descreve a teoria relacionada ao algoritmo *boosting*, o qual foi posteriormente aperfeiçoado por Freund & Schapire (1995) e Freund & Schapire (1997), que apresenta sua versão mais conhecida, o *Adaboost*.

Outro trabalho de destaque na área foi o *Bagging* (BREIMAN, 1996). *Bagging* é um acrônimo para *Bootstrap Aggregating*, sendo um algoritmo simples e com um bom desempenho. O *bagging* é baseado na amostragem *bootstrapping* (EFRON; TIBSHIRANI, 1993) e gera um determinado número de amostras a partir do conjunto original. As amostras são geradas de maneira uniforme, com reposição, de maneira que os conjuntos gerados possuam o mesmo tamanho do conjunto original. Dentro de cada amostra alguns padrões podem ser selecionados mais de uma vez, ao passo que outros poderão não ser selecionados. Cada novo conjunto obtido desta forma servirá para o treinamento de um classificador. Com

isto, espera-se que o algoritmo gere diversidade entre os classificadores e atue na redução da variância do erro de classificação.

Ao contrário do *Bagging*, no algoritmo *Boosting* os conjuntos de treinamento não são gerados de maneira uniforme. A probabilidade de uma determinada amostra ser selecionada para compor um conjunto de treinamento é função de sua dificuldade de classificação pelos classificadores anteriormente treinados. Assim, uma dada amostra que não foi corretamente classificada pelos classificadores já treinados terá maior probabilidade de ser selecionada para compor o próximo conjunto de treinamento que as amostras corretamente classificadas. *Bagging* e *Boosting* são atualmente bastante utilizados para a geração de comitês de classificadores.

2.1. Abordagens Existentes Na Geração de Comitês de Classificadores

Para Dietterich (2000) existem razões, de três naturezas distintas, que justificam a construção de comitês de classificadores. A primeira razão é de natureza estatística, onde um algoritmo de aprendizagem pode ser interpretado como uma busca pela melhor hipótese num espaço de hipóteses H . Cada hipótese representa um classificador obtido no processo de treinamento. Neste caso, surge um problema quando a quantidade de dados existente é muito pequena, e o algoritmo pode encontrar mais de uma hipótese que leve a uma mesma taxa de acerto no conjunto de treinamento. Desta forma, uma melhor generalização pode ser obtida quando as hipóteses encontradas (classificadores distintos) são combinadas através da média de suas respostas, como forma de minimizar o risco da escolha de uma hipótese errada durante o treinamento.

A segunda razão, apontada por Dietterich, é de natureza computacional e pode ocorrer em casos de abundância de dados, ao contrário da razão anterior. Nesta segunda razão, os algoritmos podem atingir diferentes mínimos locais, não necessariamente alcançando um mínimo global na superfície de erro. Como forma de minimizar os efeitos do atingimento de diferentes mínimos locais durante o treinamento, e assumindo que tais mínimos são próximos ao mínimo global, uma melhor resposta de generalização pode ser obtida pela combinação das saídas individuais de cada classificador.

E por fim, a terceira razão para a utilização de comitês de classificadores é de natureza representacional e está relacionada à natureza dos algoritmos de classificação e sua relação

com o conjunto de amostras apresentado para o treinamento. Por exemplo, caso o problema seja de natureza não-linear, e o classificador utilizado seja de natureza linear, este não poderá isoladamente representar a melhor hipótese dentre todas as possíveis hipóteses de H . Assim, a combinação de classificadores poderá ser uma alternativa para o problema. O aspecto representacional também poderá estar relacionado ao fato de que a amostragem utilizada para o treinamento pode não conter o caminho para se encontrar a hipótese ótima, justificando assim a utilização de uma combinação na tentativa de obter uma melhor aproximação àquela hipótese.

É oportuno ressaltar que os motivos expostos acima não são pressupostos suficientes para uma melhor resposta de generalização em um problema de classificação, não garantindo, portanto, que o processo de classificação via comitês irá sempre oferecer uma resposta melhor que a obtida com a utilização de um único classificador (o melhor classificador – *single best classifier*) ou mesmo elevar a média das taxas de acerto individuais dos classificadores. O que é garantido é que estas três deficiências existentes no processo de classificação via classificador único serão consideradas, e na pior das situações, atenuadas. Por outro lado, a teoria estudada e o trabalho experimental publicado até agora vêm demonstrando, de maneira geral, que o êxito na classificação via comitês é algo factível.

2.1.1. Abordagens e Categorias

No processo de geração de comitês de classificadores, duas principais abordagens poderão ser utilizadas: seleção ou fusão. A seguir as abordagens serão descritas.

Fusão de Classificadores

Esta abordagem, do ponto de vista conceitual do uso de comitês é a mais intuitiva, pois nas técnicas de fusão de classificadores (KITTLER *et al.*, 1998) todos os classificadores são, de alguma forma, utilizados no processo de classificação. Uma vez que, a depender da natureza do classificador, cada componente do comitê poderá gerar saídas discretas (rotulações de classes) ou contínuas, que em geral são normalizadas no intervalo $[0, 1]$ e podem ser interpretadas como o suporte do classificador a cada uma das possíveis valorações de classe. Na fusão, as saídas individuais de cada classificador são agregadas através de uma função ou regra. Um processo de fusão é esquematizado na figura 1.1.

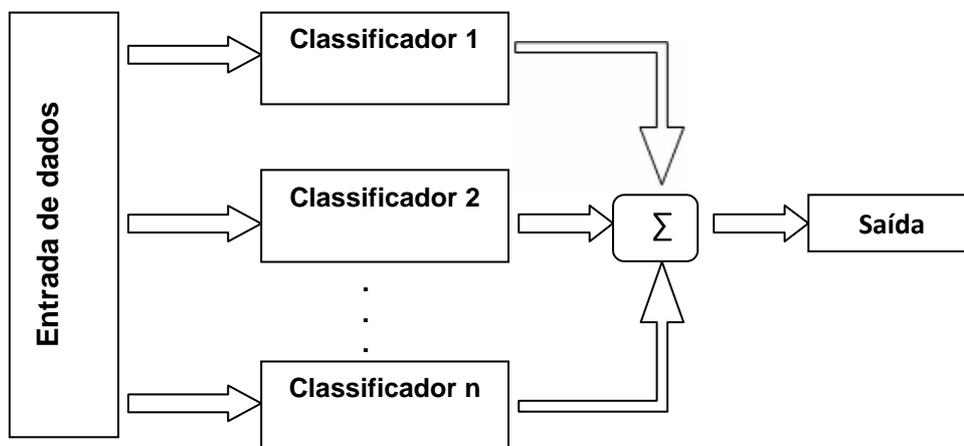


Figura 1: Estrutura de um comitê de classificadores. O símbolo “ Σ ” representa o processo de fusão.

Uma vez que saídas podem ser contínuas, representando o suporte do classificador a cada possível classe, a fusão pode se dar através de regras pré-estabelecidas. Como exemplos de regras utilizadas para a fusão de classificadores, pode-se citar a Regra da Soma, a Regra do Produto, a Regra do Máximo, Regra da Média, a Regra da Mediana e, para o caso de uma saída discreta, o Voto Majoritário (KITTLER *et al.*, 1998) (KUNCHEVA, 2002).

- **Regra da Soma**

Esta regra computa a saída do comitê a partir da soma de todos os suportes fornecidos pelo classificador para cada possível valoração de classe de saída. A classe que apresentar o maior a maior soma correspondente gerará a rotulação do padrão.

- **Regra do Produto**

Caso os classificadores sejam independentes, esta regra obtém bons resultados. A saída é obtida da mesma forma que a regra da soma, apenas substituindo a operação da soma pela multiplicação. Sua principal deficiência se dá quando o suporte de um classificador, para uma determinada classe, é zero, o que poderá conduzir o processo de fusão a um erro de classificação.

- **Regra da Média e da Mediana**

A saída final do classificador, na forma de um suporte individual para cada possível valoração de classe de saída, é dada pela média, ou mediana, calculada para todos os suportes.

- **Regra do Máximo**

Esta regra seleciona o classificador com o máximo valor de probabilidade *a posteriori*, caso o classificador gere como saída uma distribuição de probabilidades ou uma aproximação desta. A utilização desta regra, entretanto, pode ferir a razão estatística para a utilização de comitês, conforme descrito por Dietterich (2000): caso o classificador tenha sido sobreajustado (*overfitting*) a saída gerada pela regra pode não ser a melhor opção de combinação.

- **Voto Majoritário**

Para a utilização das regras descritas anteriormente, existe a preocupação de que os suportes gerados pelos classificadores membros de um comitê, para cada possível rotulação de classe, sejam informados sobre a forma de distribuição de probabilidade (ou uma aproximação desta), ou, ao menos, tenha suas saídas calibradas de alguma forma. Esta preocupação é fruto do processo de treinamento ao qual foi submetido cada classificador, que pode gerar, para cada componente do comitê, diferentes intervalos de saída. A regra do voto majoritário, até certo ponto, elimina esta preocupação visto que a regra trabalha com a combinação de saídas discretas (ou discretizadas) geradas pelos classificadores. Nesta regra, um rótulo de classe é atribuído a um padrão de teste via processo de votação, onde cada componente atribue uma rotulação ao padrão e a rotulação mais votada é, ao final, atribuída ao padrão.

Estas regras descritas não são as únicas. Existem várias outras regras, tais como o voto majoritário ponderado (KUNCHEVA, 2004) onde, no processo de votação, cada classificador tem seu voto geralmente ponderado por sua precisão em um conjunto de validação.

Regras, tais como as descritas acima, não levam em consideração o fato de que classificadores redundantes e/ou imprecisos podem afetar negativamente a performance do sistema. Redundância nos classificadores implica em baixa diversidade e, por conseguinte, aumento de complexidade do sistema (RUTA, D.; GABRYS, B., 2005). Para uma melhor ponderação destes problemas existe o processo de seleção de classificadores.

Seleção de Classificadores

O principal objetivo no processo de seleção de classificadores (WOODS *et al.*, 1997) (GIACINTO, G.; ROLI, F., 2001) (KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., 2008) é encontrar, a partir de um conjunto de classificadores, o subconjunto mais competente para a modelagem do problema. Por vezes este subconjunto é substituído por um único classificador. O subconjunto gerado é então utilizado, via processo de fusão, para produzir uma resposta aos dados de entrada. A idéia central aqui é selecionar e combinar um subconjunto de classificadores ao invés de combinar todo o conjunto. O processo de seleção pode ser dividido em duas categorias: seleção estática e seleção dinâmica.

No processo de seleção estática, um conjunto de classificadores é selecionado como parte do processo de treinamento, de forma que um mesmo comitê de classificadores responderá pela classificação de todo o conjunto de testes. Na seleção dinâmica de classificadores, o subconjunto (ou um único classificador) é selecionado durante a etapa de generalização, dinamicamente, a depender do padrão de teste atual. Este último processo é tipicamente baseado na definição de regiões de competência, *i.e.* regiões do espaço de características onde estão contidos padrões espacialmente próximos ao atual padrão de teste, em geral calculadas com o uso de um conjunto de validação separado para esta única finalidade.

A literatura recente tende a separar o processo de seleção dinâmica em dois grandes grupos: seleção dinâmica de classificadores e seleção dinâmica de comitês. A exceção de algumas poucas abordagens, tais como (DOS SANTOS, E. M.; SABOURIN, R.; MAUPIN, 2008) (DOS SANTOS, E. M.; SABOURIN, R., 2011), os processos funcionam de maneira bastante similar, de forma que o critério que define a escolha dinâmica de um único classificador pode ser estendido aos demais componentes para a seleção de um comitê, ao invés de um único classificador. Tipicamente o classificador que atinge o máximo valor do critério de competência adotado é escolhido no processo de seleção dinâmica de classificador, ao passo que este mesmo processo pode ser reformulado para a escolha de um comitê a partir da definição de um valor de corte em termos do critério de competência adotado, tal como ocorre no modelo probabilístico descrito em (WOLOSZYNSKI, T.; KURZYNSKI, M., 2011).

Alguns estudos vem demonstrando que a seleção dinâmica de classificadores (ou comitês) tende a obter melhores resultados que a abordagem estática de seleção, uma vez que muitos padrões de testes estão relacionados a diferentes graus de dificuldade de classificação, requerendo assim tratamentos diferenciados.

3. Objetivos do Trabalho

O objetivo deste trabalho é realizar um estudo sobre a concepção de um sistema de multi-classificadores baseado na abordagem dinâmica de construção de comitês. O sistema é baseado numa métrica de competência relativamente simples, descrita em (GIACINTO, G.; ROLI, F.; FUMERA, G., 2000) e nos métodos apresentados em (KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., 2008).

4. Método Proposto

O método proposto é baseado na métrica descrita em (GIACINTO, G.; ROLI, F.; FUMERA, G., 2000), que representa o valor médio do suporte a uma dada classe que o classificador fornece para um padrão de teste x_i . Este suporte é calculado tomando os k -vizinhos mais próximos a x_i obtidos num conjunto de validação independente. Representa a média ponderada dos suportes que o classificador fornece à rotulação de classe real de cada um dos k -vizinhos mais próximos, onde os pesos são as distâncias entre cada um dos k -vizinhos mais próximos e o padrão. A métrica original estabelece que o suporte a cada classe representa a probabilidade *a posteriori* de o padrão pertencer àquela classe. No presente trabalho esta probabilidade será substituída por uma aproximação obtida por um processo de calibração realizada nas saídas de uma rede neural do tipo MLP (HAYKIN, S, 2002).

O método é baseado na abordagem *Overproducing and Choose Strategy* (OCS) (PARTRIDGE, D.; YATES, W., 1996), que gera um conjunto inicial de classificadores (*classifier pool*) baseado em alguma técnica de geração de comitês. Tipicamente é utilizado o algoritmo *Bagging* na geração do conjunto inicial, tal como em (DOS SANTOS, E. M.; SABOURIN, R.; MAUPIN, P., 2008). Outras técnicas para a geração do conjunto inicial de classificadores também são utilizadas, a exemplo do *Random Subspace Method* (HO, T., 1998) que gera classificadores com base em diferentes subconjuntos do espaço de características (diferentes subconjuntos de atributos originais).

A figura 2 ilustra a divisão do conjunto de dados para o método. O conjunto de treinamento é utilizado na geração do conjunto inicial de classificadores, seguindo a abordagem OCS utilizando o algoritmo *Bagging*, como forma de gerar um conjunto diverso de classificadores. São utilizados classificadores do tipo redes neurais MLP. O conjunto V é

utilizado na extração das competências associadas a cada padrão de testes, seguindo a abordagem descrita em (GIACINTO, G.; ROLI, F.; FUMERA, G., 2000).

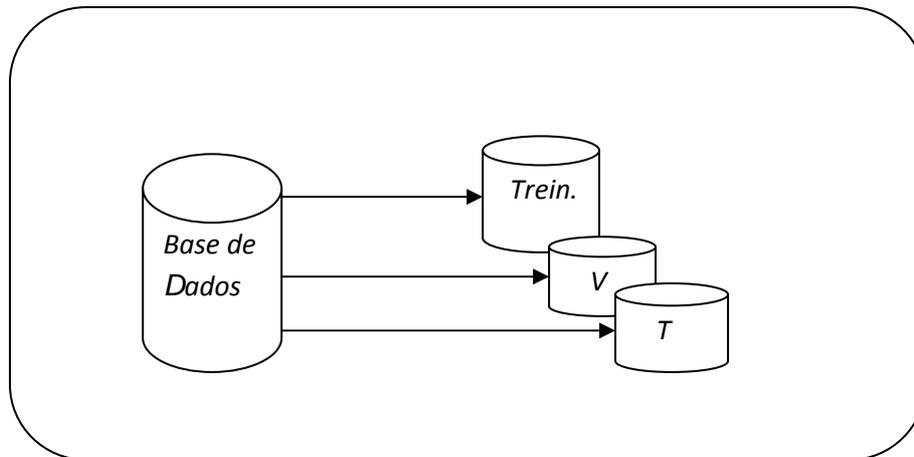


Figura 2: Divisão do conjunto de dados para a abordagem proposta. Cada base de dados é dividida em três conjuntos: treinamento, validação (V) e teste (T).

Para cada classificador, utilizando cada padrão de teste, o método proposto gera um par associado ao padrão. Este par é composto pela competência do classificador na classificação do padrão de teste x_i e por uma rotulação atribuída. Esta rotulação recebe valor 1 caso o classificador atual rotule corretamente uma determinada fração de seus k -vizinhos mais próximos em V , e 0 caso o classificador não atinja este limiar de acertos em V . Este processo de rotulação tem inspiração no funcionamento do algoritmo *k-nora Eliminate*, descrito em (KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., 2008). O processo de geração de pares é repetido para cada classificador, sempre tomando por base o mesmo padrão x_i atual, de forma que será gerado um conjunto de m pares, onde m é a quantidade de componentes do conjunto inicial de classificadores gerado. A partir destas rotulações, utilizando-se uma métrica de corte empregada em problemas de classificação binária e baseada no teste estatístico de Kolmogorov-Smirnov (CONNOVER, W. J., 1999), um limiar de corte é estabelecido. Os classificadores, cujos valores de pares estejam ranqueados acima deste limiar, são selecionados para classificar o padrão x_i .

5. Metodologia

O trabalho será desenvolvido a partir do estudo das técnicas e métodos estatísticos e matemáticos relacionados com a teoria envolvida, além da análise e entendimento profícuo das técnicas de combinação de classificadores existentes, onde para tanto, uma extensa revisão de literatura foi realizada.

A tecnologia Java é utilizada no desenvolvimento dos algoritmos. O desempenho das técnicas estudadas e desenvolvidas será mensurado: i) a partir dos custos computacionais associados ao processo de seleção de atributos e ao treinamento da rede construída com estes atributos de entrada e ii) pelos desempenhos dos comitês, medidos pelos métodos de geração da curva ROC (FAWCETT, 2003) e pelo teste estatístico de Kolmogorov-Smirnov (KS) (CONNOVER, W. J., 1999), para bases de dados de problemas binários e iii) pela taxa de acertos de classificação. Em todos os casos, um teste estatístico pareado será empregado para comparar o desempenho do método com outras abordagens.

O desempenho será testado sobre os atributos de bases de dados públicas envolvendo problemas de classificação (inicialmente classificação binária), e comparado com outras técnicas de construção de comitês de classificadores, em especial com o *k-nora Eliminate* (KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., 2008) e com o algoritmo *Bagging*.

6. Resultados Esperados

O método utilizado já foi codificado, em linguagem Java. Os resultados inicialmente obtidos demonstram que o algoritmo gera resultados satisfatórios, com desempenho similar a um dos métodos nos quais ele se baseia, apresentando resultados ligeiramente superiores ao *k-nora Eliminate* (KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., 2008). O método ainda possui espaço para melhorias, onde um prévio *ranking* de classificadores, com base em critérios de diversidade e relevância, será testado.

7. Cronograma e Estado Atual da Pesquisa

O cronograma abaixo destaca as atividades programadas para os próximos 6 (seis) meses do curso. Até o presente momento, foram realizadas atividades referentes à conclusão dos créditos necessários para conclusão do mestrado, revisão bibliográfica, estudo aprofundado da bibliografia encontrada, codificação do método proposto em linguagem computacional. Foram iniciados os experimentos com o método, que evidenciaram a existência de espaço para melhorias, as quais estão sendo executadas.

Os prazos podem sofrer modificações no decorrer da evolução e aprofundamento do trabalho.

Atividades	Março	Abril	Mai	Junho	Julho	Agosto
Levantamento bibliográfico						
Estudo da bibliografia encontrada						
Codificação do método em linguagem computacional						
Refinamento do método	█	█				
Realização de experimentos	█	█				
Testes e análise dos resultados		█	█			
Escrita da dissertação	█	█	█	█	█	█
Preparação da defesa						█
Defesa da dissertação						█

Referências Bibliográficas

BREIMAN, L. "Bagging Predictors". **Machine Learning**, v. 24, n. 2, p. 123–140, 1996.

CONOVER, W. J. **Practical Nonparametric Statistics**. John Wiley & Sons, 1999.

DIETTERICH, T., "Ensemble methods in machine learning". In **J. Kittler and F. Roli, editors, the 1st International Workshop on Multiple Classifier Systems**, Lecture Notes in Computer Science. Springer-Verlag, 2000.

DOS SANTOS, E. M.; SABOURIN, R.; MAUPIN, P. "A dynamic overproduce-and-choose strategy for the selection of classifier ensembles". **Pattern Recognition**, v. 41, p. 2993–3009, 2008.

DOS SANTOS, E. M.; SABOURIN, R., "Classifier ensembles optimization guided by population oracle". In **Proceedings of IEEE Congress on Evolutionary Computation**. pp.693–698, 2011

EFRON, E.; TIBSHIRANI, R. **An Introduction to the Bootstrap**. Chapman & Hall, 1993.

FAWCETT, T. **ROC Graphs: Notes and Practical Considerations for Researchers.**, Palo Alto, p.1-38, 16 mar. 2004. Disponível em: <<http://www.binf.gmu.edu/mmasso/ROC101.pdf>>. Acesso em: 15 nov. 2010.

FREUND, Y.; SCHAPIRE, R. E. "A decision-theoretic generalization of the on-line learning and an application to boosting". In **Proceedings of the 2nd. European Conference on Computational Learning Theory**, (pp. 23–27). Barcelona, Spain, (1995).

FREUND, Y.; SCHAPIRE, R. E. "Experiments with a New Boosting Algorithm". **Proc. 13th International Conference on Machine Learning (ICML-96)**. [S.l.]: [s.n.]. 1996. p. 148–156.

GIACINTO, Giorgio; ROLI, Fabio; FERRI, "Selection of Classifiers Based on Multiple Classifier Behaviour." **Advances in Pattern Recognition: Joint IAPR International Workshops SSPR'2000 and SPR 2000**, Alicante, Spain, August/September 2000: Proceedings. By Giorgio Giacinto. Ed. Francesc Ferri. Berlin: Springer, 2000. 87-93. Print. 2000.

GIACINTO, G.; ROLI, F., "Dynamic Classifier Selection Based On Multiple Classifier Behaviour", **Pattern Recognition**, vol. 34, pp. 1879–1881, 2001.

HANSEN, L.K.; SALAMON, P., "Neural network ensembles". **IEEE Transactions on Pattern Analysis and Machine Intelligence** 12 (10) (1990) 993–1001.

HAYKIN, Simon. **Redes Neurais: Princípios e Prática**. 2. ed. Porto Alegre: Bookman, 2002.

HO, T., "The random subspace method for constructing decision forests", **IEEE Transactions on PAMI** 20 (8) (1998) 832–844.

KITTLER, Josef; HATEF, Mohamad; DUIN, Robert P. W.; MATAS, Jiri, "On combining classifiers", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 20, pp.226–239, 1998.

KO, A. H. R.; SABOURIN, R.; BRITTO JR., A. S., "From dynamic classifier selection to dynamic ensemble selection," **Pattern Recognition**, vol. 41, pp. 1735–1748, may 2008.

KROGH, A.; VEDELSBY, J., "Neural network ensembles, cross validation and active learning", **Advances in Neural Information Processing Systems** 7 (1995) 231–238.

KUNCHEVA, L.I.; "A theoretical study on six classifier fusion strategies", **IEEE Transactions on Pattern Analysis and Machine Intelligence** 24 (2) 281–286, 2002.

KUNCHEVA, L.I, **Combining Pattern Classifiers: Methods and Algorithms**, Wiley-Interscience, 2004.

PARTRIDGE, D.; YATES, W., "Engineering multiversion neural-net systems", **Neural Computation** 8 (4) (1996) 869–893.

RUTA, D.; GABRYS, B., "Classifier selection for majority voting", **Information Fusion** 6 (1) 63–81, 2005.

SCHAPIRE, R. E. "The Strength of Weak Learnability". **Machine Learning**, v. 5, n. 2, p. 197–227, 1990.

WOLOSZYNSKI, T.; KURZYNSKI, M. "A probabilistic model of classifier competence for dynamic ensemble selection", **Pattern Recognition**, Volume 44, Issues 10-11, October-November 2011, Pages 2656-2668

WOODS, Kevin; KEGELMEYER, PHILIP Jr., W.; BOWYER, Kevin, "Combination of multiple classifiers using local accuracy estimates," **IEEE Transactions on Pattern Analysis Machine Intelligence**, vol. 19, pp. 405–410, April 1997.