

RECOMENDAÇÃO AUTOMÁTICA DE ESTRATÉGIAS DE APRENDIZADO ATIVO

Plano de Pesquisa de pós-doutorado

Candidato:
Davi Pereira dos Santos

Supervisor:
Prof. Dr. André C. P. L. F. de Carvalho

Universidade de São Paulo - USP
Instituto de Ciências Matemáticas e de Computação - ICMC
Departamento de Ciências de Computação

Abril/2015

Resumo

Apesar do crescente avanço no desenvolvimento de algoritmos capazes de induzir modelos preditivos numa grande diversidade de domínios, a participação humana ainda é necessária em muitas aplicações. A atividade de supervisão é essencial para a construção do conjunto de exemplos de treinamento que alimenta um algoritmo de aprendizado. Trata-se de uma atividade custosa, que requer tempo e conhecimento do supervisor, frequentemente na atividade de rotulação dos dados. Consequentemente, é desejável que apenas os exemplos mais importantes sejam considerados.

A amostragem de exemplos relevantes para rotulação é chamada de aprendizado ativo e pode ser empreendida segundo diferentes estratégias. Caso se deseje a melhor eficiência possível, é preciso recorrer a algum critério de escolha da melhor estratégia. Entretanto, diferentemente da tradicional escolha de algoritmos de aprendizado, que dispõe do conjunto de treinamento para compará-los por meio de validação, a escolha de estratégias de aprendizado ativo é anterior a essa possibilidade. Assim que uma alternativa é experimentada, o orçamento é gasto, impedindo que outras alternativas sejam testadas sem incorrer em custos adicionais.

Este projeto propõe a investigação aprofundada da possibilidade de recomendação automática de estratégias por meio de meta-aprendizado, dado que a única tentativa, de conhecimento do candidato, são seus próprios experimentos preliminares de conclusão do doutorado. Nesses experimentos, em fase de publicação, há evidências da viabilidade dessa abordagem, porém com implicações ainda desconhecidas para área de aprendizado ativo.

1 Introdução e Motivação

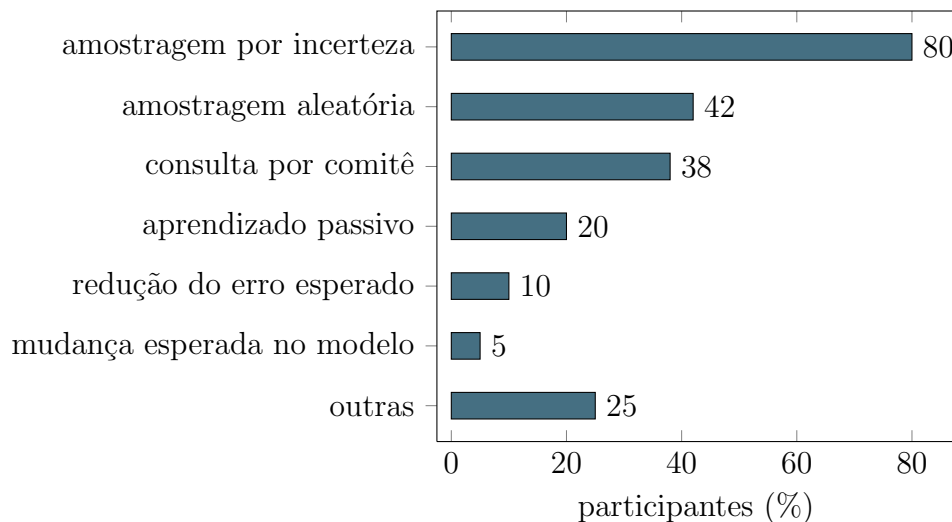
Atualmente, o aprendizado de máquina permeia o cotidiano humano provendo auxílio em tarefas diversas (Bishop, 2006). Seu bom desempenho na tarefa de classificação depende da existência de dados categorizados de qualidade para treinamento do sistema e de um algoritmo de aprendizado adequado. A categorização dos dados é um processo frequentemente custoso, pois é realizada por um supervisor humano que atribui categorias/classes para cada objeto/exemplo de interesse contido nos dados ou, dependendo da aplicação, é realizada por um processo químico ou mecânico (Bryant et al., 2001), por exemplo.

Dados os limites práticos de orçamento, esforço humano disponível, resistência à fadiga e crescente presença massiva de dados, faz-se necessária a *amostragem* de apenas um subconjunto de exemplos para a rotulação e subsequente construção do conjunto de treinamento. Quando se deseja eficiência no uso de recursos, essa amostragem de exemplos não é trivial, pois resultados teóricos e empíricos apontam para a existência de uma diferença de desempenho entre os métodos desenvolvidos na área de *aprendizado ativo* (Settles, 2012) frente à amostragem aleatória. Apesar de ser uma área promissora, a diversidade de métodos de amostragem ativa aliada às diferentes características de cada base de dados e aos vieses dos algoritmos de aprendizado dificulta a fase de projeto de um sistema de aprendizado de máquina. Existem assim, dois problemas de escolha: estratégia de amostragem e algoritmo de aprendizado.

Tradicionalmente, a escolha do algoritmo de aprendizado é feita com base na experiência pessoal do especialista responsável pelo sistema ou em restrições próprias da aplicação. Outra possibilidade é o uso de um método automático de recomendação como o *meta-aprendizado* (Brazdil et al., 2009). Apesar dessas abordagens serem indicadas para o caso do aprendizado supervisionado convencional, em que todos os exemplos de treinamento já estão rotulados, elas não foram propostas para a situação de escassez ou ausência de rótulos, que é o caso de aplicações com orçamento limitado e ainda na fase de escolha da melhor forma de amostragem para rotulação.

Idealmente, a escolha do algoritmo de aprendizado é baseada em métodos de validação cruzada. Logo, a escolha se dá com o conjunto de treinamento completo, ou seja, amostrado e rotulado. Consequentemente, a escolha da estratégia de amostragem ativa, precede a determinação do algoritmo de aprendizado. Outra particularidade na escolha da estratégia é sua criticidade, tendo-se em vista que o sucesso da estratégia escolhida só pode ser determinado após ter-se incorrido em custos financeiros com a atividade de supervisão. Dessa forma, é preciso adotar uma estratégia com base em expectativas de desempenho ou de acordo com características da base de dados. Na prática, essa escolha tem sido arbitrária, tendendo a se concentrar no uso da estratégia mais simples, chamada

Figura 1: Frequência de uso de estratégias na competição de aprendizado ativo (alguns participantes adotaram mais de uma estratégia). *Adaptado de (Guyon et al., 2011).*



amostragem por incerteza. Essa preferência foi reportada numa competição de aprendizado ativo, onde também prevaleceu a ausência de estratégias possivelmente mais efetivas, como as baseadas em densidade, conforme Figura 1. Em consonância com esse panorama, está a ausência de estudos comparativos abrangentes que possam guiar o especialista na fase de amostragem.

Adicionalmente, diversas estratégias fazem uso interno de algoritmos de aprendizado para embasar a amostragem. Esse uso interno fecha um círculo impossível de dependências: a estratégia depende do algoritmo; o algoritmo depende da existência do conjunto de treinamento que requer rotulação; e, a rotulação é feita pela estratégia de aprendizado ativo. A forma trivial de se evitar o círculo de dependências, para além da amostragem aleatória, é a adoção de estratégias agnósticas. Essas estratégias não fazem suposições a respeito dos dados no que diz respeito ao viés de aprendizado, pois não requerem um algoritmo interno. Elas se baseiam, por exemplo, em medidas de densidade ou estatísticas de agrupamento (Dasgupta, 2011). A dispensa do algoritmo, entretanto, leva a amostragens que não consideram a fronteira de decisão que seria traçada durante o aprendizado. Essa fronteira contém informações que permitem consultas mais prospectivas que exploratórias, potenciais aceleradoras do processo de descoberta dos exemplos mais relevantes.

Assim, existem pelo menos dois problemas e respectivos subproblemas em aberto na área de aprendizado ativo:

- Que situações dispensam o uso de aprendiz (algoritmo interno)? Dentre as estratégias *agnósticas*, qual a mais indicada para um dado problema?

- Como quebrar o círculo de dependências? Dentre as estratégias *gnósticas*, qual a mais indicada para um dado problema?

Neste projeto, propõe-se investigar os desdobramentos da aplicação de meta-aprendizado como apoio ao usuário na definição do melhor par estratégia-algoritmo para seu problema. Essa forma de recomendação automática quebra o círculo de dependências, pois o aprendiz é recomendado em conjunto com a estratégia, isentando o usuário de definir um algoritmo compatível com ambas, base de dados e estratégia. Está implícito nessa abordagem que o algoritmo de aprendizado do aprendiz é desvinculado do algoritmo de aprendizado do classificador definitivo da aplicação que, por sua vez, é escolhido apenas após a rotulação.

O meta-aprendizado tem sido investigado e demonstrado efetivo no grupo de pesquisa do candidato na melhoria da classificação de dados de expressão gênica (Souza, 2010; Souza et al., 2008, 2010, 2009). O presente enfoque é verificar sua aplicabilidade na recomendação de estratégias e algoritmos visando uma rotulação eficiente.

Este plano de pós-doutorado é organizado como segue. Na Seção 2, os principais paradigmas de aprendizado ativo são expostos. Uma introdução ao meta-aprendizado é apresentada na Seção 3. Finalmente, na Seção 4, a proposta é detalhada juntamente com o plano de atividades e recursos necessários.

2 Aprendizado ativo

Aprendizado ativo é o estudo de máquinas de aprendizado capazes de se aprimorar fazendo perguntas (Settles, 2012). A ideia básica é, numa analogia com o aprendizado humano, aproveitar a individualidade de cada aluno e seu próprio ritmo de aprendizado. Assim, ele pode ser auxiliado pelo professor nos aspectos em que experimenta maior dificuldade. Da mesma forma, um algoritmo pode usufruir de uma atenção seletiva que priorize os exemplos mais difíceis para ele em um dado momento.

Dentre as perguntas que as máquinas de aprendizado ativo são capazes de elaborar, a mais direta é “*Qual é a classe do exemplo x ?*”. Dado que a obtenção de um rótulo confiável normalmente é um processo custoso, o problema de se decidir qual seria o melhor x é o cerne de uma estratégia de aprendizado ativo. Por exemplo, quando a consulta de um exemplo envolve reações químicas destrutivas, é desejável fazer o mínimo possível de consultas visando um reduzido custo material. Dessa forma, apenas uma parcela criteriosamente escolhida dos exemplos deve ser rotulada, ou *consultada*, na terminologia de aprendizado ativo. Assim, um direcionamento adequado do esforço de aprendizado tem como resultado um processo de rotulação mais econômico.

As estratégias podem ser baseadas em diferentes concepções de relevância de exemplos ou mesmo diferentes teorias do aprendizado. A amostragem por incerteza, por exemplo, assume que os exemplos e seus rótulos pertencem a uma distribuição de probabilidades;

Tabela 1: Características de cada estratégia. A ordem de complexidade é dada conforme a quantidade de exemplos a aprender para viabilizar cada consulta.

estratégia	forma de busca	presença de aprendiz	definição da sequência de consultas	dependência entre consultas	ordem de complexidade
Rnd	exploratória aleatória	agnóstica	autônoma	nenhuma	$\mathcal{O}(0)$
Clu	balanceada: exploratória e prospectiva	agnóstica	interativa	total	não especificada
Unc/Mar/Ent QBC/DW	prospectiva	gnóstica	interativa	total	$\mathcal{O}(1)$
SG	exploratória limitada e aleatória	gnóstica	interativa	total	$\mathcal{O}(Y)$
EGL	prospectiva	gnóstica (aprendiz específico)	interativa	total	$\mathcal{O}(Y \mathcal{U})$
EER	prospectiva	gnóstica	interativa	total	$\mathcal{O}(Y \mathcal{U} ^2)$
TU	balanceada: exploratória e prospectiva	gnóstica	interativa	após etapa exploratória	$\mathcal{O}(1)$

a amostragem por busca no espaço de hipóteses, por sua vez, assume a existência de hipóteses integrantes de um espaço de versões (Mitchell, 1997) que enquadram ou não cada exemplo. Estratégias agnósticas (Balcan et al., 2009), por outro lado, são independentes de algoritmo de aprendizado. Essa diversidade de embasamentos configura-se praticamente como um conjunto de paradigmas de amostragem ativa.

Os paradigmas normalmente se situam em um ou mais dos três principais cenários da literatura de aprendizado ativo (Settles, 2010): *síntese de consulta por associação* ou *consulta de exemplos sintetizados*¹; *amostragem baseada em reserva de exemplos*²; e, *amostragem seletiva baseada em fluxo*³. O cenário adotado neste trabalho é o baseado em reserva de exemplos. Esse cenário permite que a estratégia tenha acesso a todos os exemplos e escolha qual consultar. Nesse contexto, a variedade de abordagens fundamentais existentes é resumida na Tabela 1 cuja primeira coluna referencia cada estratégia pela abreviação de seu nome original em inglês, que é dado juntamente com os títulos das seções seguintes.

¹membership query synthesis

²pool-based sampling

³stream-based selective sampling

2.1 Amostragem aleatória - Rnd

Na amostragem aleatória não há uma ordem de preferência para a realização de consultas. Ela é puramente exploratória, ou seja, não enfoca nenhuma região em especial do espaço de exemplos; e, é agnóstica, pois não requer um aprendiz.

2.2 Amostragem por incerteza - Unc

A amostragem por incerteza decide quando selecionar um exemplo \mathbf{x} pela máxima probabilidade a posteriori de se obter a classe \mathbf{y} dada por um modelo probabilístico (Lewis, 1995); onde Y é o conjunto de classes possíveis:

$$P_{max}(\mathbf{x}) = \max_{\mathbf{y} \in Y} P(\mathbf{y}|\mathbf{x}) \quad (1)$$

A estratégia de amostragem por incerteza consiste em consultar o exemplo mais informativo \mathbf{x}^* , ou seja, aquele com a menor $P_{max}(\mathbf{x})$, com o intuito de se explorar a fronteira de decisão no espaço de exemplos, conforme Equação 2.

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{U}} P_{max}(\mathbf{x}) \quad (2)$$

Onde \mathcal{U} é a reserva de exemplos.

2.3 Amostragem por margem ou entropia - Mar ou Ent

Em problemas com mais de duas classes, o menor $P_{max}(\mathbf{x})$ pode não indicar o melhor exemplo. A medida de margem, apresentada nas Equações 3 e 4, utiliza o valor da diferença entre as duas maiores probabilidades. Outra possibilidade é a medida de entropia normalizada (Lewin et al., 2004), apresentada na Equação 5; onde $z(\mathbf{x})$ mapeia o exemplo \mathbf{x} com sua classe correspondente.

$$z(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y} P(\mathbf{y}|\mathbf{x}) \quad (3)$$

$$M(\mathbf{x}) = P(z(\mathbf{x})|\mathbf{x}) - \max_{\mathbf{y} \in Y \setminus \{z(\mathbf{x})\}} P(\mathbf{y}|\mathbf{x}) \quad (4)$$

$$E(\mathbf{x}) = -\log^{-1} |Y| \sum_{\mathbf{y} \in Y} P(\mathbf{y}|\mathbf{x}) \log P(\mathbf{y}|\mathbf{x}) \quad (5)$$

2.4 Consulta por comitê - QBC

Comitês são combinações de modelos com o objetivo de superar as predições de modelos únicos. Em aprendizado ativo tem-se interesse no grau de desacordo entre os membros (Abe e Mamitsuka, 1998). A divergência de Jensen-Shannon (Lin, 1991) é uma medida

comumente utilizada em comitês para avaliar o grau de desacordo (Melville e Mooney, 2004). A divergência não-ponderada de Jensen-Shannon é definida em termos da entropia das distribuições na Equação 6.

$$JS(\{\theta^{(m)} \forall m \in Y\}) = E(\sum_{m \in Y} P(\theta^{(m)}, \mathbf{x})) - \sum_m E(P(\theta^{(m)}, \mathbf{x})) \quad (6)$$

Onde $\theta^{(m)}$ é o modelo membro de número m . Quanto maior o valor de JS , mais distante os membros estão de um consenso. Assim, o exemplo com o maior valor deve ser consultado primeiro.

2.5 Busca no espaço de hipóteses - SG

A intuição da busca no espaço de hipóteses é que os exemplos mais importantes residem na região onde as hipóteses se contradizem. Isso equivale a consultar os exemplos que reduziriam o espaço de versões (Mitchell, 1997) depois de inseridos no conjunto de treinamento. A busca no espaço de hipóteses é feita pelo acompanhamento das hipóteses mais específicas e as mais gerais pertencentes aos conjuntos S e G de todas as hipóteses possíveis, denominadas $h_S \in S$ e $h_G \in G$, respectivamente (Cohn et al., 1994).

2.6 Redução do erro esperado - EER

A estratégia de redução de erro é um método que busca pelo exemplo que mais reduz uma função objetivo, por exemplo a entropia, na predição geral do modelo para todo o conjunto de dados (Guo e Greiner, 2007). Para cada exemplo candidato com vetor descritivo \mathbf{x} obtido da reserva \mathcal{U} , sua classe mais provável, representada pelo vetor preditivo \mathbf{y}' , é calculada de forma otimista:

$$\mathbf{y}' = \arg \min_{\mathbf{y}} \sum_{\mathbf{u} \in \mathcal{U}} O(\mathbf{x}, \theta_{\mathcal{L} \cup \{\langle \mathbf{u}, \mathbf{y} \rangle\}}) \quad (7)$$

onde O é a função objetivo e \mathcal{L} é a parcela já rotulada da reserva de exemplos.

2.7 Impacto esperado no modelo - EMC

O futuro impacto de um exemplo sobre o modelo, chamado de *mudança esperada no modelo*⁴, é uma indicação de sua possível contribuição para o aprendizado e pode ser obtido na variação do gradiente de uma rede neural, por exemplo (Settles et al., 2007). Como o rótulo não é sabido de antemão, usa-se a soma das contribuições de cada rótulo ponderada pelas respectivas probabilidades condicionais.

⁴ *expected model change*

2.8 Amostragem ponderada por densidade - DW

As amostragens ponderadas por densidade utilizam a medida de *densidade de informação*, que atribui diferentes pesos à medida de informatividade $Inf(\mathbf{x})$ conforme o nível de concentração de exemplos não rotulados no entorno de \mathbf{x} dado por uma medida de similaridade $sim(\mathbf{x}, \mathbf{u})$ (Settles, 2008):

$$ID(\mathbf{x}) = Inf(\mathbf{x}) \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u} \in \mathcal{U}} sim(\mathbf{x}, \mathbf{u}) \quad (8)$$

Uma melhoria é o acréscimo da *utilidade de treinamento* (Fujii et al., 1998), que é a densidade de informação inversamente ponderada pela concentração de exemplos rotulados:

$$TU(\mathbf{x}) = ID(\mathbf{x}) \left(\sum_{\mathbf{l} \in \mathcal{L}} sim(\mathbf{x}, \mathbf{l}) \right)^{-1} \quad (9)$$

Qualquer medida de similaridade e de informatividade podem ser adotadas.

2.9 Amostragem por agrupamento - Clu

A amostragem por agrupamento explora agrupamentos naturais na reserva de exemplos (Dasgupta, 2011). Essa abordagem é uma alternativa à realização de consultas que enfocam a fronteira de decisão ou o manejo de hipóteses citados anteriormente. Uma importante representante desse paradigma é baseada em agrupamento hierárquico. O método de agrupamento hierárquico (Murtagh, 1983) organiza os exemplos numa hierarquia que pode ser representada por uma árvore. A amostragem hierárquica faz uso da árvore para definir a relevância dos exemplos. Eles têm maior probabilidade de serem consultados se pertencerem aos grupos mais impuros e estatisticamente representativos.

3 Meta-aprendizado

Um sistema de classificação baseado em aprendizado de máquina depende de um modelo que é induzido por algoritmos de aprendizado. Diante da infinidade de vieses possíveis de aprendizado, muitos algoritmos têm sido propostos e alguns são frequentemente empregados de forma generalizada na solução dos mais diversos problemas, como é o caso das redes neurais artificiais (Haykin e Network, 2004). Entretanto, nenhum algoritmo pode ser adequado a todos os domínios, pois um desempenho positivo em algumas situações de aprendizado precisa ser compensado por um igual grau de desempenho negativo em outras (Schaffer, 1994). Isso decorre da existência de um viés necessário na forma de representação (árvores de decisão e neurônios artificiais entre outras) e busca de hipóteses sobre um dado problema (busca gulosa e otimização de funções entre outros). A existên-

cia do viés de aprendizado é essencial para a capacidade de generalização do algoritmo (Mitchell, 1980).

Dessa forma, um sistema de aprendizado de máquina requer uma escolha criteriosa de qual algoritmo deva ser empregado. Normalmente, o problema da escolha do algoritmo é resolvido por um especialista em aprendizado de máquina que se baseia em conhecimentos sobre os dados e sobre os algoritmos disponíveis para escolher manualmente o melhor. Essa escolha é feita segundo alguma métrica de desempenho (Brazdil et al., 2009). Uma maneira de se evitar a escolha manual é a adoção de algum tipo de *meta-aprendizado*, que é o estudo do aperfeiçoamento dos algoritmos de aprendizado por meio da experiência. Esse aperfeiçoamento se dá no nível *meta*, que é um nível acima do aprendizado convencional, chamado de nível *base* (Vilalta e Drissi, 2002). No nível base, o viés de aprendizado é fixo, enquanto que, no nível meta, o viés normalmente é escolhido dinamicamente. Existem diferentes formas de meta-aprendizado. As mais relevantes são apresentadas nas seções seguintes. A Seção 3.4, em especial, descreve a abordagem mais aplicável ao problema de recomendação de estratégias de aprendizado ativo. Dependendo do conjunto de meta-atributos escolhidos, ela permite caracterizar as bases de dados com poucos rótulos ou na ausência deles.

3.1 Generalização em pilha

Na *generalização em pilha*⁵ (Wolpert, 1992) o meta-aprendiz lida com uma metabase que consiste de um conjunto de treinamento transformado por aprendizes no nível base. O resultado dessa transformação são metaexemplos cujos valores dos atributos são as previsões de cada modelo base. Uma particularidade dessa abordagem é o viés estático, pois ocorre uma combinação de algoritmos ao invés de uma seleção.

3.2 Caracterização por modelos

A própria estrutura dos modelos do nível base pode ser explorada na construção dos metaexemplos. Uma representante da *caracterização por modelos* é a *indução de modelos tipados de ordem maior*⁶. Ela gera - de acordo com exemplo dado no trabalho de (Bensusan et al., 2000) - uma árvore de decisão para cada base de dados. As árvores são completamente representadas por estruturas complexas que servem de metaexemplos que são aprendidos por algoritmos especialmente desenvolvidos para esse tipo de tarefa.

⁵*stacked generalization*

⁶*typed higher-order induction*

3.3 Marcadores de referência

Os *marcadores de referência*⁷ (Pfahringer et al., 2000) são um conjunto diverso de algoritmos simples usados como referência para algoritmos mais complexos. A acurácia de cada um dos modelos associados fornece o valor de um meta-atributo. Embora não diretamente aplicável, o conceito da geração de meta-atributos por meio de processamentos que representem uma simplificação da tarefa base pode ser adaptável ao cenário de aprendizado ativo.

3.4 Caracterização direta

A caracterização direta consiste na obtenção de medidas diretamente dos exemplos que compõem uma base de dados, ou seja, sem o intermédio de um algoritmo de aprendizado. A primeira caracterização de bases de dados foi feita por (Rendell et al., 1987) com o intuito de prever acurácia e tempo de processamento. Ela era baseada no número de exemplos e de atributos. O próximo conjunto de meta-atributos, proposto no projeto STATLOG (Brazdil e Henery, 1994), era composto de quinze medidas:

- número de exemplos, atributos binários e não binários e classes;
- entropia das classes, informação mútua entre classe e atributos e razão sinal-ruído;
- entropia, curtose, assimetria, correlação e razão entre os desvios padrão entre atributos;
- primeira correlação canônica e variância pelo primeiro discriminante canônico.

Variações desse conjunto são propostas em trabalhos posteriores (Brazdil et al., 2009), como a adoção de histogramas para evitar a perda de informações que ocorre quando se adota a média das medidas nos diferentes atributos base (Kalousis, 2002); ou a binarização de medidas, como o grau de dispersão do atributo alvo em tarefas de regressão (Gomes et al., 2012). Outros conjuntos visam a recomendação automática de algoritmos não supervisionados. Essa tarefa é mais próxima da recomendação de estratégias de aprendizado ativo pela ausência de rótulos. Alguns trabalhos, por exemplo, adotam medidas como o grau de normalidade da distribuição e o percentual de pontos aberrantes e de atributos mais relevantes (Ferrari e de Castro, 2015; Souto et al., 2008). Também há trabalhos direcionados a: otimização (Kanda et al., 2011), fluxos de dados (Rossi et al., 2014), predição de ranqueamentos (de Souza et al., 2010) e detecção de ruído (Garcia et al., 2015). Não é do conhecimento do candidato a existência de publicações referentes à aplicação de meta-aprendizado a aprendizado ativo. Experimentos preliminares têm apresentado indícios da efetividade de tal abordagem e estão sendo submetidos em veículo apropriado

⁷*landmarkers*

na época de escrita deste projeto. Ela têm sido investigada de forma pioneira pelo candidato na fase final de seu doutorado e sua exploração em maior profundidade está em sua tese como um possível desdobramento futuro que espera-se contemplar com a proposta da Seção 4.

4 Proposta

A literatura atual de aprendizado ativo segue pelo menos duas tendências de criação: novos cenários, como a rotulação massiva por meio de supervisores de baixo custo (Vijayanarasimhan e Grauman, 2014); e, novas estratégias como a síntese de exemplos de fronteira (Wang et al., 2015) dentre outras. Apesar das novas propostas de estratégias contribuírem para o avanço da área, elas tornam a tarefa de escolha de estratégia ainda mais complexa, pois aumenta a quantidade de opções, sem necessariamente aumentar a eficiência da amostragem. Adicionalmente, cada estratégia requer uma implementação compatível com o sistema do usuário, inviabilizando, assim, a experimentação de grande parte das abordagens existentes.

O candidato do presente projeto comparou, em seu doutorado, estratégias já consolidadas na literatura e seus nichos de aplicação conforme publicado recentemente em (Santos e Carvalho, 2014a) e (Santos e Carvalho, 2014b). As possibilidades iniciais de recomendação automática são apresentadas na tese, ainda em conclusão, antes do fechamento do presente texto. A proposta deste projeto é manter esse conjunto de estratégias já implementadas (Santos e Carvalho, 2015), que é representativo da diversidade de paradigmas existentes, como as alternativas disponíveis ao usuário. Mais precisamente, será considerado cada par estratégia-algoritmo como uma alternativa.

O objetivo primário é explorar as possibilidades de recomendação automática desses pares por meio da aplicação de meta-aprendizado.

Especificamente, espera-se investigar as seguintes questões:

1. Quais são as implicações do uso de meta-aprendizado?
2. Ele pode representar uma *quebra de paradigma* na área (Kuhn, 2012)?
3. Existe um subconjunto ideal de estratégias a se considerar?
4. Qual o efeito no meta-aprendizado apenas para agnósticas (ou gnósticas)?
5. Qual tipo de recomendação é mais adequado: predição de ranqueamento, predição da vencedora, predição de desempenho ou outro?

4.1 Infraestrutura

O trabalho será desenvolvido no Laboratório de Computação Bioinspirada (BIOCOM), do ICMC/USP. Ele dispõe de computadores pessoais e servidores para a implementação

e execução de experimentos. Conta-se ainda com as bibliotecas das unidades de Computação, Física e Engenharia do Campus de São Carlos.

4.2 Metodologia

As técnicas desenvolvidas no projeto serão investigadas experimentalmente e avaliadas principalmente quanto à viabilidade na recomendação automática de estratégias. Serão empregados métodos quantitativos amplamente aceitos na literatura de aprendizado de máquina (Bishop, 2006) e testes estatísticos como o teste de Friedman/Nemenyi (Demsar, 2006).

4.3 Plano de atividades

As atividades previstas para a execução deste projeto de pós-doutorado são as seguintes:

1. **Pesquisa bibliográfica.** Pesquisa bibliográfica com o objetivo de monitorar o surgimento de eventuais publicações que conciliem aprendizado ativo e meta-aprendizado. Atualização de conhecimentos e estudo de alternativas para cumprir os objetivos apontados.
2. **Implementação de técnicas.** A partir da pesquisa bibliográfica, é possível que novas estratégias se mostrem interessantes do ponto de vista de sua representatividade e contribuição para a diversidade de opções do sistema de recomendação. Elas podem ser implementadas para integrar o sistema de meta-aprendizado.
3. **Experimentação e validação.** Por meio da metodologia comentada anteriormente serão conduzidos experimentos para avaliação das propostas.
4. **Redação do relatório final.** Redação do relatório final expondo os resultados obtidos no decorrer deste projeto.
5. **Elaboração de artigos científicos.** Serão redigidos artigos que reflitam o grau de desenvolvimento do projeto e as contribuições realizadas.

4.4 Cronograma

O cronograma de desenvolvimento das atividades compreende dois anos. Ele é apresentado na Tabela 2.

Atividades	Ano 1						Ano 2					
	01 - 02	03 - 04	05 - 06	07 - 08	09 - 10	11 - 12	01 - 02	03 - 04	05 - 06	07 - 08	09 - 10	11 - 12
1) Pesquisa bibliográfica												
2) Implementação												
3) Experimentação e validação												
4) Redação do relatório												
5) Elaboração de artigos												

Tabela 2: Cronograma de atividades

Referências

- ABE, N.; MAMITSUKA, H. Query learning strategies using boosting and bagging. In: SHAVLIK, J. W., ed. *ICML*, Morgan Kaufmann, 1998, p. 1–9.
- BALCAN, M.-F.; BEYGEZIMER, A.; LANGFORD, J. Agnostic active learning. *J. Comput. Syst. Sci.*, v. 75, n. 1, p. 78–89, 2009.
- BENSUSAN, H.; GIRAUD-CARRIER, C. G.; KENNEDY, C. J. A higher-order approach to meta-learning. In: CUSSENS, J.; FRISCH, A. M., eds. *ILP Work-in-progress reports*, CEUR-WS.org, 2000 (*CEUR Workshop Proceedings*, v.35).
- BISHOP, C. M. *Pattern Recognition and Machine Learning*, v. 4. 738 p., 2006.
- BRAZDIL, P.; GIRAUD-CARRIER, C. G.; SOARES, C.; VILALTA, R. *Metalearning - applications to data mining*. Cognitive Technologies. Springer, I-X, 1–176 p., 2009.
- BRAZDIL, P.; HENERY, R. Analysis of results. *Machine learning, neural and statistical classification*, p. 175–212, 1994.
- BRYANT, C. H.; MUGGLETON, S.; OLIVER, S. G.; KELL, D. B.; REISER, P. G. K.; KING, R. D. Combining inductive logic programming, active learning and robotics to discover the function of genes. *Electron. Trans. Artif. Intell.*, v. 5, n. B, p. 1–36, 2001.
- COHN, D. A.; ATLAS, L. E.; LADNER, R. E. Improving generalization with active learning. *Machine Learning*, v. 15, n. 2, p. 201–221, 1994.
- DASGUPTA, S. Two faces of active learning. *Theoretical Computer Science*, v. 412, n. 19, p. 1767–1781, 2011.
- DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.
- FERRARI, D. G.; CASTRO, L. N. Clustering algorithm selection by meta-learning systems: A new distance-based problem characterization and ranking combination methods. *Information Sciences*, v. 301, n. 0, p. 181 – 194, 2015.
- FUJII, A.; INUI, K.; TOKUNAGA, T.; TANAKA, H. Selective sampling for example-based word sense disambiguation. *Computational Linguistics*, v. 24, n. 4, p. 573–597, 1998.
- GARCIA, L. P. F.; CARVALHO, A. C. P. F.; LORENA, A. C. Noise detection in the meta-learning level. *Neurocomputing*, in press 2015, 2015.
- GOMES, T. A. F.; PRUDÊNCIO, R. B. C.; SOARES, C.; ROSSI, A. L. D.; CARVALHO, A. C. P. L. F. Combining meta-learning and search techniques to select parameters for support vector machines. *Neurocomputing*, v. 75, n. 1, p. 3–13, 2012.

- GUO, Y.; GREINER, R. Optimistic active-learning using mutual information. In: VELOSO, M. M., ed. *IJCAI*, 2007, p. 823–829.
- GUYON, I.; CAWLEY, G. C.; DROR, G.; LEMAIRE, V. Results of the active learning challenge. In: *Active Learning and Experimental Design @ AISTATS*, JMLR.org, 2011, p. 19–45.
- HAYKIN, S.; NETWORK, N. A comprehensive foundation. *Neural Networks*, v. 2, n. 2004, 2004.
- KALOUSIS, A. *Algorithm selection via meta-learning*. Tese de Doutorado, Universidade de Geneve, 2002.
- KANDA, J.; CARVALHO, A. C. P. L. F.; HRUSCHKA, E. R.; SOARES, C. Selection of algorithms to solve traveling salesman problems using meta-learning. *Int. J. Hybrid Intell. Syst.*, v. 8, n. 3, p. 117–128, 2011.
- KUHN, T. S. *The structure of scientific revolutions*. University of Chicago press, 2012.
- LEWIN, J.; SCHMITT, A. O.; ADORJÁN, P.; 0002, T. H.; PIEPENBROCK, C. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics*, v. 20, n. 17, p. 3005–3012, 2004.
- LEWIS, D. D. A sequential algorithm for training text classifiers: Corrigendum and additional data. *SIGIR Forum*, v. 29, n. 2, p. 13–19, 1995.
- LIN, J. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, v. 37, n. 1, p. 145–151, 1991.
- MELVILLE, P.; MOONEY, R. J. Diverse ensembles for active learning. In: *Proceedings of the Twenty-first International Conference on Machine Learning*, New York, NY, USA: ACM, 2004, p. 74–.
- MITCHELL, T. M. The need for biases in learning generalizations. In: SHAVLIK, J. W.; DIETTERICH, T. G., eds. *Readings in Machine Learning*, Morgan Kaufman, p. 184–191, book published in 1990., 1980.
- MITCHELL, T. M. *Machine learning*. McGraw Hill Series in Computer Science. McGraw-Hill, I–XVII, 1–414 p., 1997.
- MURTAGH, F. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.*, v. 26, n. 4, p. 354–359, 1983.
- PFARRINGER, B.; BENSUSAN, H.; GIRAUD-CARRIER, C. Tell me who can learn you and i can tell you who you are: Landmarking various learning algorithms. In: *Proceedings of the Seventeenth International Conference on Machine Learning (ICML2000)*, 2000, p. 743–750.
- RENDELL, L. A.; SHESHU, R.; TCHENG, D. K. Layered concept-learning and dynamically variable bias management. In: MCDERMOTT, J. P., ed. *IJCAI*, Morgan Kaufmann, 1987, p. 308–314.
- ROSSI, A. L. D.; CARVALHO, A. C. P.; SOARES, C.; SOUZA, B. F. Metastream: A meta-learning based method for periodic algorithm selection in time-changing data. *Neurocomputing*, v. 127, p. 52–64, 2014.
- SANTOS, D. P.; CARVALHO, A. C. Viabilidade do aprendizado ativo em máquinas extremas. In: *Brazilian Conference on Intelligent Systems and Encontro Nacional de Inteligência Artificial e Computacional*, 2014a.
- SANTOS, D. P.; CARVALHO, A. C. P. L. F. Comparison of active learning strategies and proposal of a multiclass hypothesis space search. In: POLYCARPOU, M. M.; CARVALHO, A. C. P. L. F.; PAN, J.-S.; WOZNIAK, M.; QUINTIÁN, H.; CORCHADO, E., eds. *HAIS*, Springer, 2014b, p. 618–629 (*Lecture Notes in Computer Science*, v.8480).
- SANTOS, D. P.; CARVALHO, A. C. P. L. F. An Active Learning Library for Scala. *GitHub Software Repository*, DOI:10.5281/zenodo.13733. URL:http://dx.doi.org/10.5281/zenodo.13733, 2015.
- SCHAFER, C. A conservation law for generalization performance. In: COHEN, W. W.; HIRSH, H., eds. *ICML*, Morgan Kaufmann, 1994, p. 259–265.

- SETTLES, B. *Curious machines: active learning with structured instances*. Tese de Doutorado, University of Madison Wisconsin, 2008.
- SETTLES, B. Active learning literature survey. *University of Wisconsin, Madison*, v. 52, p. 55–66, 2010.
- SETTLES, B. *Active learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2012.
- SETTLES, B.; CRAVEN, M.; RAY, S. Multiple-instance active learning. In: PLATT, J. C.; KOLLER, D.; SINGER, Y.; ROWEIS, S. T., eds. *NIPS*, Curran Associates, Inc, 2007.
- SOUTO, M. C. P.; PRUDÊNCIO, R. B. C.; SOARES, R. G. F.; ARAUJO, D. S. A.; COSTA, I. G.; LUDERMIR, T. B.; SCHLIEP, A. Ranking and selecting clustering algorithms using a meta-learning approach. In: *IJCNN*, IEEE, 2008, p. 3729–3735.
- SOUZA, B. F. *Meta-aprendizagem aplicada á classificação de dados de expressão gênica*. Tese, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos - SP, 2010.
- SOUZA, B. F.; CARVALHO, A.; SOARES, C. Metalearning for gene expression data classification. In: *HIS '08: Proceedings of the 2008 Eighth International Conference on Hybrid Intelligent Systems*, Washington, DC, USA: IEEE Computer Society, 2008, p. 441–446.
- SOUZA, B. F.; CARVALHO, A.; SOARES, C. Empirical evaluation of ranking prediction methods for gene expression data classification (aceito). In: *IBERAMIA*, 2010, p. 1–8.
- SOUZA, B. F.; CARVALHO, A. C. P. L. F.; SOARES, C. Empirical evaluation of ranking prediction methods for gene expression data classification. In: MORALES, Á. F. K.; SIMARI, G. R., eds. *IBERAMIA*, Springer, 2010, p. 194–203 (*Lecture Notes in Computer Science*, v.6433).
- SOUZA, B. F.; SOARES, C.; CARVALHO, A. Meta-learning approach to gene expression data classification. *International Journal of Intelligent Computing and Cybernetics*, v. 2, n. 2, p. 285–303, 2009.
- VIJAYANARASIMHAN, S.; GRAUMAN, K. Large-scale live active learning: Training object detectors with crawled data and crowds. *International Journal of Computer Vision*, v. 108, n. 1-2, p. 97–114, 2014.
- VILALTA, R.; DRISSI, Y. A perspective view and survey of meta-learning. *Artif. Intell. Rev*, v. 18, n. 2, p. 77–95, 2002.
- WANG, L.; HU, X.; YUAN, B.; LU, J. Active learning via query synthesis and nearest neighbour search. *Neurocomputing*, v. 147, p. 426–434, 2015.
Disponível em: <http://dx.doi.org/10.1016/j.neucom.2014.06.042>
- WOLPERT, D. H. Stacked generalization. *Neural Networks*, v. 5, n. 2, p. 241–259, 1992.