

Inventário de itens de estrutura e de planejamento – V2 (16 de maio)

Discussão livre entre os envolvidos na disciplina PSI 2672 (Reconhecimento de Padrões e Neurocomputação), edição de 2011, em direção ao projeto de grupo final.

Desenvolvimento de pré-propostas Algumas perguntas que devem ser feitas na fase de pré-propostas

Primeiramente, algumas perguntas de definição de escopo, composição e objetivos do grupo:

Título Preliminar:

Composição do grupo e e-mails:

Descrição em poucas linhas da Temática:

Principais objetivos (para atingir no fim de semestre) do projeto – definição preliminar:

Ambientes de HW e/ou SW de entorno que devem ser desenvolvidos no projeto:

Ambientes de HW e/ou SW de entorno já prontos que serão ser desenvolvidos no projeto:

Caracterizações / ilustrações / medidas planejadas sobre o projeto final:

Alguns passos / algumas perguntas iniciais para a depuração da pré-proposta:

- Quem serão os componentes do vetor X (vetor de observáveis) – e qual o significado de cada variável no sistema sendo desenvolvido?

- Tipicamente, dos X elencáveis num primeiro passo, alguns podem ser descartados no decurso do projeto, por terem impacto marginal na variável y estimada / na decisão y estimada / no reconhecimento de padrões em questão. Numa primeira análise, qual a ordem de importância dos X elencados, do mais relevante ao que provavelmente tem papel mais marginal?

- Qual a natureza e o significado da variável y ? É a variável estimada? Prevista? Trata-se de decisão, de estimação? Qual seu significado? ...

- Como serão coletados os dados dos conjuntos de treino e teste?

- Quantos pares $(X;y)$ se espera ter para treino, neste projeto? E para teste? Em quanto tempo serão coletados tais dados? Há necessidade de desenvolvimento de HW e/ou SW específico para esta coleta? Quanto tempo isto levará? Note que estes tempos devem ser pequenos face ao tempo do restante do projeto.

- Há alternativa de obtenção de dados em bases públicas já prontas, caso os dados experimentais não possam ser coletados no final das contas? Quais seriam elas?

- Há sentido (e viabilidade) em se pensar num gerador de dados artificiais para uma fase inicial do desenvolvimento, com dados mais controlados que os dados reais?

- O projeto envolve um desafio de problema não estruturado (função escondida não prevista por alguma teoria ou por modelagem consagrada de sucesso)

- É razoável assumir que a função escondida depende apenas (ou principalmente) das variáveis escolhidas como observáveis? Ou seja, é razoável assumir que $y=F(X) + \text{erro}$, com erro zero ou pequeno? Há teoria ou modelagem anterior que suporta tal hipótese? Qual seria (discurse sobre ela)?

- Há uma componente randômica adicionada à função escondida? Qual sua origem? É possível estimar sua magnitude?

- E quanto às variáveis observáveis X ? Questões similares às acima se aplicam: há elementos randômicos em algum/s dos X s? Qual sua origem? É possível estimar sua/s magnitude/s?

- Qual seriam potenciais pré-processamentos que seriam aplicáveis às grandezas X e y ?

- Há pré-processamentos pretendidos visando redução de dimensionalidade? Quais seriam?

- Há pré-processamentos pretendidos visando melhor discriminabilidade de classes? Quais seriam?

- Há pré-processamentos pretendidos visando , etc etc (outros aspectos vistos em classe)? Quais seriam?

- Qual o suporte de especialistas da área você planeja para atacar as várias questões acima? Quem são esses especialistas? Já fez a sondagem sobre disponibilidade para ajudá-lo?

- Qual o suporte de literatura você planeja para atacar as várias questões acima, seja como reforço a sua parceria com especialista/s, seja como substituição? Quais são os títulos principais de livros / artigos / grupos de discussão / sites

- Como será caracterizado o resultado final? Há necessidade de desenvolvimento de HW e/ou SW específico?

ANOTAÇÕES ADICIONAIS DO GRUPO: