

MAC – 5701 Tópicos em Ciência da Computação
Prof. Responsável: Yoshiharu Kohayakawa

Plano de Estudos

Aluno: **Germano Capistrano Bezerra**

Prof. Orientador: **Marcelo Finger**

Tema: Aprendizado Computacional Não-Supervisionado e Métodos de *Clustering*

Introdução

No contexto de sistemas distribuídos, surge o conceito das grades de computação. As grades de computação têm por objetivo principal otimizar a utilização de recursos computacionais ociosos, distribuídos ao longo de diversas redes de computadores. A idéia básica da grade é permitir que máquinas sejam a ela conectadas, compartilhando e disponibilizando seus recursos, socializando o seu uso pelos diversos processos a serem executados na rede.

O projeto de pesquisa InteGrade, em desenvolvimento no IME-USP, visa o estudo e implementação de uma infra-estrutura para grades de computação. Há diversas linhas de pesquisa sendo abordadas, a saber:

- requisitos e arquitetura geral do sistema;
- paralelização de problemas computacionalmente difíceis;
- segurança;
- identificação de padrões de acesso de usuários e de disponibilidade de recursos; e
- mobilidade de código e computação ubíqua.

Particularmente, a identificação de padrões de uso dos recursos disponíveis é um processo de aprendizado constante. Esse aprendizado tenta estabelecer métricas que permitam à Grade a identificação da disponibilidade dos recursos compartilhados, possibilitando uma melhor distribuição de carga entre os nós envolvidos no sistema.

Para tanto, é necessário monitoramento de alguns dos recursos das máquinas envolvidas na grade, como:

- memória física disponível;
- memória disponível na área de *swap*;
- quantidade de espaço livre em disco; e
- percentual de utilização do processador.

Com essas informações, deverá ser possível o estabelecimento de categorias de comportamento das máquinas. Em um dado momento de solicitação de recursos de uma dada máquina, é identificado qual a categoria em operação corrente, permitindo uma decisão sistemática sobre a possibilidade de utilização do recurso compartilhado.

Objetivo do Estudo

Este estudo tem por objetivo a conceituação de métodos de aprendizado computacional visando a implementação da identificação de padrões de acesso de usuários e de disponibilidade de recursos, no escopo do projeto de pesquisa InteGrade.

Escopo do Estudo

Esse estudo deverá abordar os seguintes tópicos:

1- Métodos de Aprendizado Computacional

- a. Aprendizado Supervisionado – onde um especialista fornece informações de saída prévias, que podem ser utilizadas pelo algoritmo de aprendizado para avaliação do resultado obtido, permitindo iterações de retroalimentação e ajuste.
- b. Aprendizado Não-Supervisionado – não há a indicação inicial de algum resultado esperado. Uma das técnicas utilizadas no aprendizado não-supervisionado, e que será abordada nesse estudo, é o método de *clustering* ou agrupamento.

2- Métodos de Agrupamento (*Clustering*)

As técnicas de *clustering* objetivam organizar o conjunto de dados de entrada em diversas categorias que agrupam os dados que mantêm semelhança entre si. Uma medida de similaridade entre os dados é estabelecida e utilizada para a determinação das diversas categorias. A análise de um novo evento é feita de forma a determinar a qual categoria esse evento pertence ou estabelecer uma nova categoria de dados a partir do reagrupamento da estrutura existente.

Como os algoritmos de *clustering* são caros, esse estudo também prevê a análise de algumas das aproximações utilizadas na literatura.

Bibliografia Proposta

Cruz, F. Aprendizado de máquina aplicado à fragmentação vertical de bases de objetos. Tese de Mestrado, UFRJ, 2002.

David J. Hand. Construction and Assessment of Classification Rules. John Wiley and Sons, 1997.

Navathe, S., Ceri, S., Wiederhold, G., Dou, J. Vertical Partitioning Algorithms for Database Design. ACM Transactions on Database Systems, vol. 9 (4), pp. 680 – 710, 1984.

Özsu, M., Valduriez, P. Principles of Distributed Database Systems. Prentice-Hall, 1999.

Pat Langley and Michael B. Morgan, editors. Elements of Machine Learning. Morgan Kaufmann, 1995.

Richard A. Johnson and Dean Wichern. Applied Multivariate Statistical Analysis. Prentice-Hall, 1983.

Tom Mitchell. Machine Learning. McGraw Hill, 1997.

Vazirani, V. Approximation Algorithms. Georgia Institute of Technology, 1997.

.