

SESHET-S: Um Sistema de Monitoria Virtual para o Ensino a Distância de Estatística

Hérico Gouveia de Souza¹, e Ronei Marcos de Moraes²

Resumo — Este artigo descreve a plataforma de monitoria virtual denominada SESHET-S e a implementação de um módulo adicional para o ensino de Amostragem. Esse sistema modela o conhecimento em módulos independentes de conhecimento usando regras “fuzzy”. A resposta a uma dada consulta ao sistema é fornecida levando-se em consideração todas as possíveis respostas e também apontando o valor percentual das suas possibilidades. O novo módulo de Amostragem contém o conhecimento dos diversos tipos clássicos de amostragem e inclui também o cálculo do tamanho da amostra para a amostragem aleatória simples usando uma média ou uma proporção

Palavras-chave — Monitoria Virtual, Sistemas Especialistas, Ensino a Distância, Ensino de Estatística.

INTRODUÇÃO

A profusão de análises estatísticas inadequadas no meio científico mostra a falta de um cuidado maior na manipulação de dados de pesquisa. Vários trabalhos publicados em periódicos de renome ou não, trazem a falta patente de um pesquisador da área de estatística, com suas análises insuficientes e conclusões seriamente prejudicadas. No meio comercial e empresarial, tal fato também ocorre e pesquisas de opinião no Brasil e em outros países trazem vícios de concepção que levam a conclusões no mínimo não-convencionais. Estas deficiências originam-se no aprendizado deficiente da estatística ministrada no ensino médio e superior, tanto pelo fato das ementas destas disciplinas não atenderem a cada curso especificamente, quanto pela falta de uma estrutura de apoio ao ensino que leve o aluno a um aprendizado mais profundo das técnicas estatísticas. Desta forma, a maioria dos profissionais sabe da existência da ciência, porém não a conhecem na profundidade suficiente para que ela lhes seja útil realmente. [12].

Dada esta realidade e com o advento do Ensino à Distância, onde ensinamos nossos alunos por correio, televisão e recentemente através de redes de computadores (notadamente pela Internet), observamos um sério problema: se o aprendizado nas disciplinas de estatística já é insuficiente no ensino presencial, como resolver as dúvidas de um estudante à distância? Na maioria das vezes, não é

possível colocar um professor em contato direto com o aluno.

No futuro, o ensino à distância reduzirá as limitações geográficas e os cursos atenderão a um grande número de estudantes. Esses estudantes necessitarão de orientadores e monitores para responder às suas questões e dúvidas [15]. Dentro desta ótica, um sistema especialista acoplado a um sistema de análise seria bem-vindo no sentido de orientar o correto uso das suas ferramentas, bem como no ensino delas, como Coob [3] sugere. Também seria interessante que esse sistema possa ensinar ao usuário os passos de uma análise e guiá-lo na sua execução, seguindo a tendência de Giraud [5].

Em várias áreas do conhecimento humano, os sistemas especialistas são utilizados para o auxílio em tomada de decisões: seja na medicina, economia, indústria, etc. As regras são usadas para expressar o conhecimento humano sobre um assunto e essas regras podem ser tratadas por vários tipos de lógica, como a lógica clássica, a lógica modal, a lógica *fuzzy*, etc de acordo com o tipo de aplicação [17].

Sob o ponto de vista didático e estatístico, não existe uma ferramenta de ensino que una um sistema especialista dedicado ao ensino a um “software” estatístico. Nem mesmo os “softwares” comerciais mais utilizados como SAS[®], SPSS[®], Statistica[®] ou SPlus[®], têm um sistema especialista para auxiliar o usuário na decisão sobre qual método é adequado ao seu problema. Estes sistemas estatísticos são capazes de análises complexas, mas não limitam ou guiam os usuários nas suas ações. Assim, esses “softwares” não podem auxiliar o usuário no uso de uma técnica em detrimento de outra e nem mesmo podem guiá-lo a não utilizar uma técnica equivocada em sua análise.

Uma possível solução para este problema seria a construção de um sistema especialista com conhecimento específico sobre “como ensinar” um determinado assunto. Assim, o estudante poderia interagir com este sistema para aprender e solucionar suas dúvidas. No entanto, esse sistema não pode ser apenas um sistema especialista para tomada de decisões, como os mencionados anteriormente; ele deve possuir características especiais para simplificar o seu uso e ser amigável, já que a sua utilização será feita por pessoas inexperientes na área.

Moraes e Zuffo [12] propuseram uma arquitetura inteligente, usando lógica *fuzzy* para aplicações em Ensino à Distância com características específicas para atender a área

¹ Hérico Gouveia de Souza, Aluno Voluntário de Iniciação Científica. Departamento de Estatística, Universidade Federal da Paraíba, Cidade Universitária s/n CEP 58.051-900 João Pessoa – PB - Brasil, tel.: +55 83 216-7075, herico_gouveia@uol.com.br

² Ronei Marcos de Moraes, Departamento de Estatística, Universidade Federal da Paraíba, Cidade Universitária s/n CEP 58.051-900 João Pessoa – PB - Brasil, tel.: +55 83 216-7075, ronei@de.ufpb.br

de ensino, como: possibilidade para interpretar respostas vagas e imprecisas (como por exemplo: “Eu não sei”, ou “Deve ser...”); possibilidade para apresentar várias possíveis soluções para uma dúvida e apontar deficiências do conhecimento do estudante. Esta arquitetura foi denominada: *Monitor Virtual* dada a sua similaridade com a função exercida pelos monitores no ensino tradicional [12]. Foi proposta também uma aplicação desta arquitetura na construção de um sistema de apoio ao ensino à distância de Estatística, denominado SESHET-S [11].

A partir da implementação desta arquitetura o SESHET-S, em sua versão original, cobria os tópicos de Análise Exploratória de Dados [16], Noções de Probabilidade [8], Inferência Estatística Paramétrica [9], Inferência Estatística Não-paramétrica [7], Análise de Regressão e Análise de Variâncias [14].

Com este leque de assuntos o *software* se apresenta como uma ferramenta bastante útil para o esclarecimento de dúvidas pertinentes aos mesmos. Embora estes assuntos cubram uma quantidade significativa de análises, vários tópicos avançados ainda não foram implementados, como a Amostragem que é a base para os procedimentos de coleta de dados. Este trabalho propõe o aperfeiçoamento do SESHET-S e a implementação de um módulo que trate deste tema com o seguinte conteúdo: Amostragem Aleatória Simples, Amostragem Estratificada, Amostragem Sistemática, Amostragem por Conglomerados e Determinação do Tamanho de Amostra [2]. Com esse conteúdo, o sistema SESHET-S poderá se útil para um número maior de disciplinas e estudantes.

Este artigo está assim disposto: na próxima seção apresenta-se uma revisão da metodologia empregada na construção do SESHET-S, a qual está sendo utilizada novamente para a adição deste novo módulo. Na seção posterior, alguns resultados serão apresentados e por fim, a última seção apresenta as conclusões finais.

METODOLOGIA

Os sistemas especialistas tratam problemas que são usualmente resolvidos por especialistas humanos [13]. Estes sistemas trabalham sobre o conhecimento de um determinado domínio, de modo similar ao ser humano, conhecimento este que é codificado para ser tratado pelo computador. O conhecimento é geralmente caracterizado por um grupo de regras e fatos. As regras são a tradução do conhecimento de um especialista e os fatos podem ser adquiridos inicialmente ou posteriormente, durante o processo de dedução [1]. Usando regras e fatos, novos fatos são obtidos. Para obter novos fatos, uma ou mais formas de raciocínio podem ser usadas. Em geral, os sistemas especialistas podem mostrar ao usuário como deduzir algo no processo de decisão e como novos conhecimentos podem ser inseridos na base de conhecimento. Além disso, podem checar se um novo conhecimento pode gerar conflito com a base de conhecimento previamente armazenada.

Usualmente, os sistemas especialistas respondem a melhor solução lógica possível para a consulta do usuário e sistemas especialistas na área de estatística não são diferentes. Como a principal aplicação é o ensino, a base de conhecimento deste sistema deve ser capaz de tratar informação incompleta ou vaga. O usuário será um estudante e ele poderá não conhecer bem o assunto e por essa razão, utiliza-se a lógica *fuzzy* [6, 20]. Essa lógica combina regras e fatos usando um modo de raciocínio denominado raciocínio aproximado [4], podendo tratar respostas incompletas e apontar várias soluções lógicas para um problema e quantificá-las [10]. Para cada resposta, o sistema deverá mostrar as condições para sua aplicação e uma medida de possibilidade [19] para ela no contexto apresentado. Dependendo da qualidade da consulta processada pelo usuário, o sistema pode concluir que o mesmo deve estudar alguns fundamentos teóricos, apontando tópicos e sugerindo bibliografia correspondente, sem lhe dar uma resposta final a sua consulta, como faria um professor tradicional.

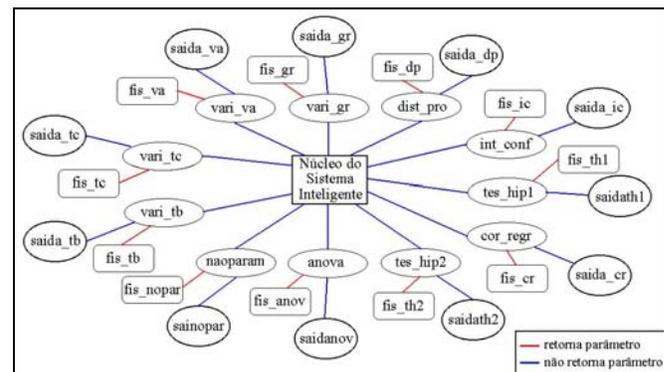


FIGURA. 1

ESTRUTURA MODULAR DO SISTEMA ESPECIALISTA SESHET-S.

Para combinar três propriedades importantes da Engenharia de Software que são: modularidade, facilidade de manutenção e facilidade de implementação de novas rotinas, o sistema foi implementado de acordo com a estrutura modular mostrada na Figura 1. O núcleo do sistema contém apenas as chamadas para sub-rotinas que implementam cada um dos tópicos abordados pelo sistema e uma parte da interface de comunicação com o usuário. Dentro das sub-rotinas (vari_va, vari_tc, dist_pro, etc), uma segunda parte da interface de comunicação com o usuário atua na coleta das informações relevantes e as transfere para uma rotina de manipulação do conhecimento. Essas rotinas de manipulação (fis_va, fis_tc, fis_dp, etc) apenas verificam a compatibilidade de cada possível resposta com os dados de entrada e retorna um vetor de possibilidades à rotina que a iniciou. Esta comunicação que retorna um parâmetro está denotada no diagrama pelas ligações em vermelho. Esta arquitetura possibilita um acesso rápido ao conjunto de regras envolvidas em cada módulo evitando assim, interações desnecessárias do sistema. Outra vantagem é a

possibilidade de evitar que problemas de manutenção em um dado módulo venham a se propagar pelo sistema.

Para implementar essa arquitetura, foi utilizado o *software* Matlab que possui uma *toolbox* para o manuseio do conhecimento dado por regras *fuzzy*, as quais manipulam o conhecimento modelando conjuntos *fuzzy* [18]. Além disso, o Matlab também possui uma *toolbox* de estatística que embora não tenha sido utilizada na construção do sistema foi um dos atrativos para a sua utilização em futuras versões.

O conjunto de regras e funções de pertinência necessárias à implementação do conhecimento foi fornecido por um especialista humano. Esse conjunto de regras e seus respectivos assuntos estão descritos na Tabela 1, onde também podem ser visualizadas as disciplinas oferecidas pela UFPB que poderão utilizar o sistema.

TABELA 1

MÓDULOS DO SISTEMA SESHET-S COM SEUS RESPECTIVOS NÚMEROS DE REGRAS, ASSUNTO E A QUE DISCIPLINA SE DESTINA.

| Disciplinas a que se destina | Assunto | Módulos implementados | Número de regras |
|--|---------------------------------|-----------------------|------------------|
| Cálculo das Probabilidades e Estatística I | - Tabelas de Frequência | vari_tb | 22 |
| | - Medidas de Tendência Central | vari_tc | 46 |
| Estatística Vital | - Med. de Variabilidade | vari_va | 45 |
| | - Gráficos | vari_gr | 15 |
| | - Distribuição de Probabilidade | dist_pro | 23 |
| Estatística III | - Intervalo de Confiança | int_conf | 36 |
| | - Testes de Hipóteses | tes_hip1 | 48 |
| | - Correlação e Regressão | cor_regr | 32 |
| Estatística IV | - Testes de Hipóteses | tes_hip2 | 80 |
| | - Análise de Variâncias | Anova | 32 |
| | - Métodos Não-paramétricos | Naoparam | 35 |
| | | Total | 414 |

FONTE: DADOS DO SISTEMA SESHET-S NA SUA VERSÃO ATUAL.

Atualmente, as disciplinas apresentadas na Tabela 1 e suas variações (disciplinas específicas com ementa ligeiramente menor) cobrem cerca de 40% das disciplinas oferecidas pelo Departamento de Estatística da UFPB, ou seja, cerca de 1000 alunos podem se beneficiar deste sistema a cada semestre letivo.

Com a implementação do novo módulo, duas novas disciplinas e cerca de 200 novos alunos irão se somar a esses potenciais usuários, elevando o percentual a valores próximos a 50% das disciplinas oferecidas pelo Departamento. O módulo sobre Amostragem segue essa arquitetura e traz um conteúdo sobre [2]:

- Amostragem Aleatória Simples - é o processo de amostragem em que as combinações de n diferentes elementos, dos N que compõem a população, possuem igual probabilidade de vir a ser a amostra efetivamente sorteada;

- Amostragem Estratificada - é o processo de amostragem no qual a população constituída por N elementos é previamente dividida em grupos distintos ou

mutuamente exclusivos, que são denominados estratos, daí o seu nome. Dentro de cada extrato podemos extrair as amostras segundo os critérios da amostragem aleatória simples.

- Amostragem Sistemática - é o processo de amostragem no qual se considera os N elementos da população numerados de 1 até N , em alguma ordem. Para selecionar uma amostra de n unidades, tomamos uma unidade aleatoriamente das primeiras k unidades e a partir desta, a cada k unidades sucessivamente.

- Amostragem por Conglomerados - é o processo de amostragem em que, os elementos da população fazendo parte de grupos, são sorteados alguns desses grupos para comporem a amostra.

- Determinação do Tamanho de Amostra - Mostrar a metodologia de cálculo do tamanho da amostra para a amostragem aleatória simples nos casos de se estimar uma média ou uma proporção [2]. No caso da média, temos dois casos a se considerar: o primeiro, quando se conhece a variância populacional e o segundo, quando ela não é conhecida. No caso da proporção, o cálculo depende do próprio parâmetro a ser estimado [2] e por isso também temos dois casos a considerar: o primeiro quando temos alguma informação sobre a proporção a ser pesquisada e o segundo, quando nenhuma informação é disponível.

A limitação do novo módulo a esse escopo específico é devido às ementas das disciplinas a serem atendidas. É sabido que outros tópicos poderiam também ser abordados, como por exemplo, o sorteio das amostras, os estimadores e seus diferentes tipos, a determinação do tamanho da amostra para os outros tipos de amostragem, os diferentes tipos de conglomerados e a correção para populações finitas, entre outras, porém estes tópicos pertencem a disciplinas mais avançadas no estudo da Amostragem. Futuramente, pretendemos ampliar o módulo de amostragem para atender a essas disciplinas mais avançadas, tanto para atender melhor aos alunos de graduação, como aos de pós-graduação.

RESULTADOS

Da concepção original do sistema a versão atual, o mesmo passou por várias melhorias e acréscimos no conjunto de regras que foram dando origem a novas versões. Estas melhorias permitiram chegar aperfeiçoar o conjunto de regras para melhor tratar formas específicas de dúvidas e também a interface, onde são utilizadas janelas com perguntas de múltipla resposta, como pode ser visto na Figura 2. Atualmente o sistema está na versão pré-release SESHET-S versão 0.4 e está em fase de testes de utilização por algumas turmas. Estes testes visam a evolução do sistema, correções de problemas e imperfeições e ainda a adaptações gráficas para facilitar a sua utilização e torná-lo mais amigável ao usuário. Ao mesmo tempo, novos conhecimentos e módulos estão sendo inseridos,

aprimorando o sistema e tornando-o mais útil para os alunos e professores das disciplinas mencionadas anteriormente.

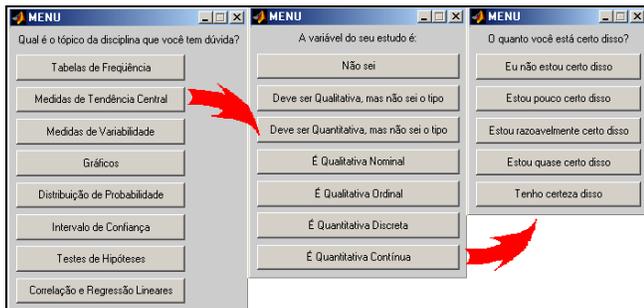


FIGURA. 2

MENU PARA ENTRADA DE INFORMAÇÕES NO SISTEMA SESHET-S.

O usuário passa as informações sobre a sua dúvida através de um sistema de menus. Esses menus direcionam o sistema para um dos módulos de conhecimento específico do SESHET-S. O sistema quantifica a certeza do aluno em cada uma das suas informações e ao final coloca todas as possíveis respostas em uma nova janela, bem como as possibilidades de ocorrência de cada uma delas, como pode ser observado na Figura 3. Devemos notar que a medida de possibilidade [19], diferentemente da medida de probabilidade [8], não precisa ter uma soma total fechada em 100%.

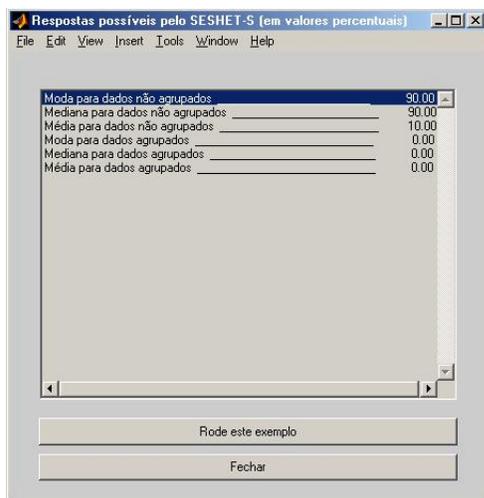


FIGURA. 3

RESPOSTAS POSSÍVEIS PARA UMA DADA CONSULTA NO SISTEMA SESHET-S.

Para ilustrar o processo de consulta, tomemos como exemplo seqüência dada pelas Figuras 2 e 3. Na Figura 2, no primeiro Menu, o aluno selecionou uma dúvida no tópico “Medidas de Tendência Central”. No segundo Menu, a variável do estudo é declarada como “É Qualitativa Ordinal” e no terceiro Menu, ele responde que “Estou quase certo disso”. O sistema continua a questionar o aluno através de

Menus, análogos ao da Figura 2 para definir melhor o escopo da resposta e as possibilidades de cada uma delas. Ao final dessas interações com o aluno, o sistema exhibe todas as respostas possíveis para o tipo de questão e as possibilidades de cada resposta, inclusive as respostas não possíveis. O objetivo na menção das respostas não possíveis é mostrar ao aluno também os caminhos que não devem ser seguidos. Na Figura 3, o aluno receberia como resposta a sua consulta que a possibilidade para a resposta “Moda para dados não agrupados” é de 90%, bem como a da “Mediana para dados não agrupados”. A “Média para dados não agrupados” possui uma possibilidade de apenas 10%. As demais respostas para as mesmas medidas para dados agrupados possuem possibilidade zero. Então o aluno pode perceber que as Medidas de Tendência Central: Moda e Mediana podem se aplicar ao seu caso com alta possibilidade. Com uma baixa possibilidade, a Média também poderia se aplicar a esse caso. A imprecisão da resposta deve-se a imprecisão do questionamento do aluno, dado que ele não tem certeza sobre o tipo da variável, ainda que aponte que ela possa ser Qualitativa, em um primeiro momento. O sistema nesse caso aponta direções para o aluno, mas não esclarece definitivamente a dúvida do aluno, assim como um professor real também não poderia fazê-lo. Um professor real necessariamente diria: “Você precisa saber qual é o tipo de variável para que eu possa esclarecer quais tipos de Medidas de Tendência Central se aplicam ao seu caso”.

CONCLUSÕES FINAIS E TRABALHOS FUTUROS

Este trabalho descreve a plataforma de monitoria virtual denominada SESHET-S e a implementação de um módulo adicional para o ensino de Amostragem. O conteúdo deste módulo inclui as técnicas clássicas de Amostragem, e a forma de se calcular o tamanho da amostra para a amostragem aleatória simples.

Esse sistema pode apoiar o ensino à distância e também pode ser aplicado como um reforço no ensino presencial. Auxiliado pelo sistema especialista *fuzzy*, o usuário pode ser treinado em várias situações a encontrar a melhor resposta para a sua consulta ou confrontar suas conclusões com as conclusões apresentadas pelo sistema especialista. Este sistema pode também verificar o conhecimento do estudante sobre algum assunto específico e assuntos correlatos.

Vale notar que em termos de arquitetura, esta é suficientemente genérica para ser aplicada ao ensino à distância de outros tópicos, relacionados ou não com a Estatística.

O sistema SESHET-S estará brevemente disponível para *download* via Internet a todos os alunos, em uma versão compilada independente do *software* Matlab. O próprio Matlab permite a compilação em separado de sistemas criados em sua plataforma. Assim, os alunos não necessitarão de ter uma licença desse *software* instalado para poder utilizar o SESHET-S. Além da facilidade de utilização para os alunos, essa solução também permite que o sistema

continue sendo desenvolvido sob a plataforma Matlab. Estuda-se também uma possível versão em linguagem Java que poderia ser operada pela Web, diretamente do servidor do Departamento de Estatística da UFPB.

Futuramente, pretendemos analisar a contribuição desse sistema ao ensino de estatística, comparando-se turmas que utilizem o sistema e que não o façam. Várias outras particularidades no estudo da amostragem poderão também ser abordadas, para tornar esse módulo um subsistema maior, dedicado ao estudo de técnicas mais avançadas dos processos de Amostragem.

REFERÊNCIAS

- [1] Buchanan, B. G.; Shortliffe, E. H. *Rule-based expert systems, the MYCIN experiments of Stanford heuristics programming project*. Reading: Addison-Wesley, 1985.
- [2] Cochran, W. G., *Sampling Techniques*. New York: John Wiley & Sons, 1977, 3rd ed.
- [3] Cobb, G. W.; Reconsidering Statistics Education: National Science Foundation Conference. *Journal of Statistics Education* v.1, n.1, 1993. [on line] <<http://www.stat.ncsu.edu/info/jse/homepage.html>>, Maio de 1999.
- [4] Dubois, D.; Prade, H. *Fuzzy sets and systems - theory and applications*. New York: Academic Press, 1980.
- [5] Giraud, G.; Cooperative Learning and Statistics Instruction. *Journal of Statistics Education* v.5, n.3, 1997. [on line] <<http://www.stat.ncsu.edu/info/jse/homepage.html>>, May of 1999.
- [6] Kruse, R.; Schwecke, E.; Heinsohn, J.; *Uncertainty and vagueness in knowledge based systems*. Berlin: Springer-Verlag, 1993.
- [7] Lehman, E. L. e D 'Abrera, H. J. M.; *Nonparametrics - Statistical Methods Based on Ranks*. San Francisco, Holden-Day, 1975.
- [8] Meyer, P. L., *Probabilidade: Aplicações à Estatística*. Rio de Janeiro: Livros Técnicos e Científicos Editora, 1983.
- [9] Mood, A. M.; Grabyll, F. A.; Boes, D. C.; *Introduction to the Theory of Statistics*. Tokyo. McGraw Hill, 1977, 3rd. ed.
- [10] Moraes, R. M.; *Uma Arquitetura de Sistemas Especialistas Nebulosos para Classificação de Imagens Utilizando Operadores da Morfologia Matemática*. Tese de Doutorado em Computação Aplicada. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998.
- [11] Moraes, R. M.; *Desenvolvimento de Sistemas Inteligentes com Interface Visual Aplicados ao Ensino Básico de Estatística*. Relatório Final de Atividades de Pós-doutorado, Departamento de Estatística - UFPB, João Pessoa, 2002.
- [12] Moraes, R. M.; Zuffo, J. A.; Intelligent Systems with Visual Interface Applied to the Basic Teaching of Statistics, *In Proceedings of 2001 International Conference on Artificial Intelligence*, Las Vegas, EUA, Junho, 2001.
- [13] Rich, E.; Knight, K.; *Artificial Intelligence*. New York: McGraw, 1993.
- [14] Snedecor, G. W. e Cochran, W. G.; *Statistical Methods*. Ames, Iowa University Press, 1967.
- [15] Synnes, K.; Parnes, P.; Widn, J.; Schefstrm, D.; *Net-based Learning for the Next Millenium*. Lule: Research Report 1999-05 (Lule University of Technology, Sweden, 1999).
- [16] Tukey, J. W.; *Exploratory Data Analysis*. New York. McGraw Hill, 1977.
- [17] Turban, E.; Aronson, J. E.; *Decision Support Systems and Intelligent Systems*. New Jersey, Prentice Hall, 5th ed., 1998.
- [18] Zadeh, L. A.; Fuzzy Sets. *Information and Control*, v. 8, pp338--353, 1965.
- [19] Zadeh, L. A.; Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, v. 1, n.1, pp3--28, 1978.
- [20] Zadeh, L. A.; Fuzzy Logic. *Computer*, v. 1, pp83--93, 1988.