

Laboratory of Computational Intelligence and Operational Research – LINC/UFPA



Dr. Adamo Santana (coordinator)

adamo@ufpa.br

+55 91 3201-8112 / +55 91 8412-9598



Belém / Pará

- ◆ Belém
 - Area: 1,070 km²
 - Population: 2,08 Mi

- ◆ Pará
 - Area: 1,247,690 km²
 - Population: 7,793 Mi



Federal University of Pará - UFPA

- ◆ Area: 3,328,655.80 m²
- ◆ Num. students: over 36,000



Main Objectives

- ◆ Investigate new models and enhancements for systems and algorithms in the areas of mathematics, optimization and computational intelligence (CI)
- ◆ Provide logistics for the public and professional areas through the development of methodologies and systems for decision support
- ◆ Apply CI models for knowledge extraction, implementing and developing algorithms for pattern analysis, classification, clustering, correlation, prediction and inference for scenario analysis

Artificial/Computational Intelligence

- ◆ Knowledge Discovery and Data mining
- ◆ Models and techniques for:
 - Pattern analysis and diagnosis
 - Correlation / Causality
 - Scenario analysis
 - Inferences
 - Forecasting
 - Optimization

Recent Projects and Applications

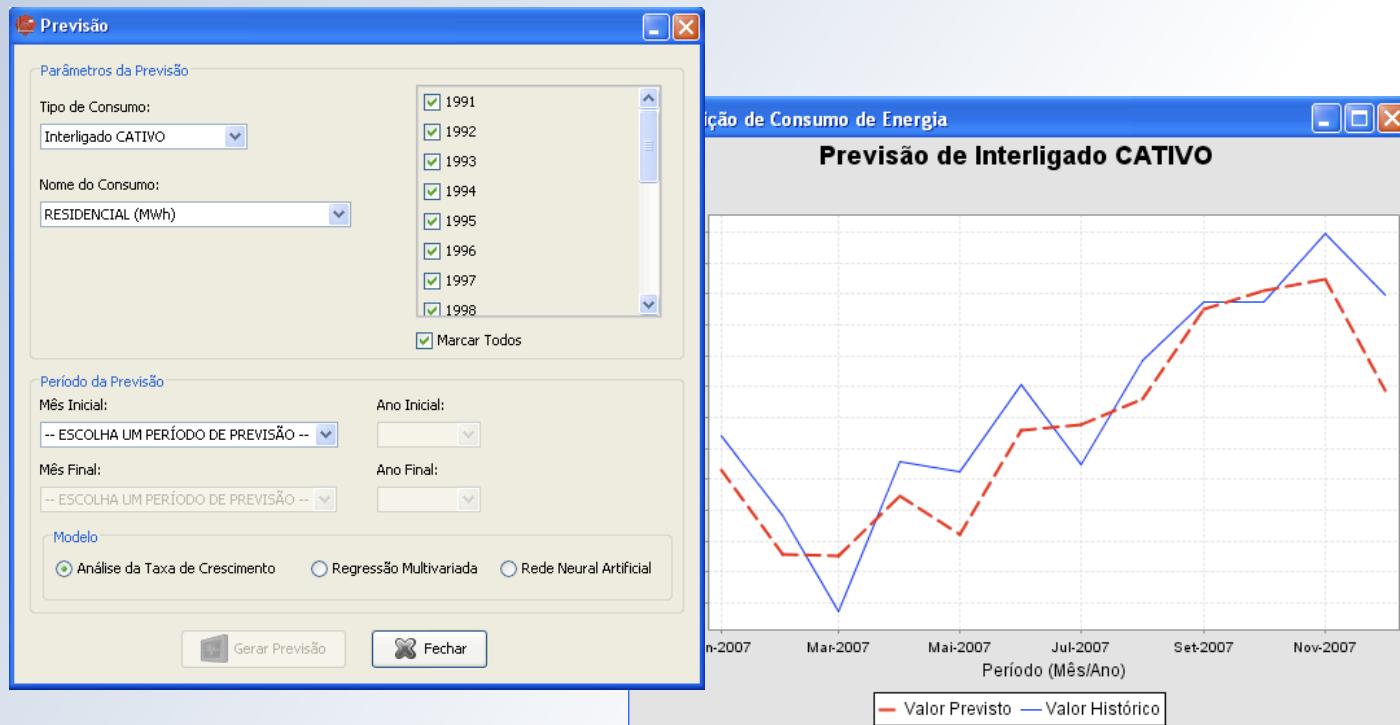
- ◆ Power systems [1],[2],[3]
- ◆ Public Safety [4]
- ◆ Bioinformatics [5]
- ◆ Health [6]
- ◆ Telecommunications [7]
- ◆ Livestock [8]
- ◆ Missing Values Treatments [9],[10]
- ◆ Automatic software testing

Power systems [1]-[3]

- ◆ Decision Support System (DSS) implemented for:
 - Forecast future demand and energy consumption [1]
 - Previous error of 4.5% by power suppliers
 - Correlate the data with socio economic and climate variables, and measure their impact [2]
 - Discover variables that mainly influence in the power consumption
 - Densification and analysis of outliers [3]

Power systems - *forecasting*

- ◆ Monthly mean error of $\sim 3 \times 10^{-7}\%$ with new hybrid CI models

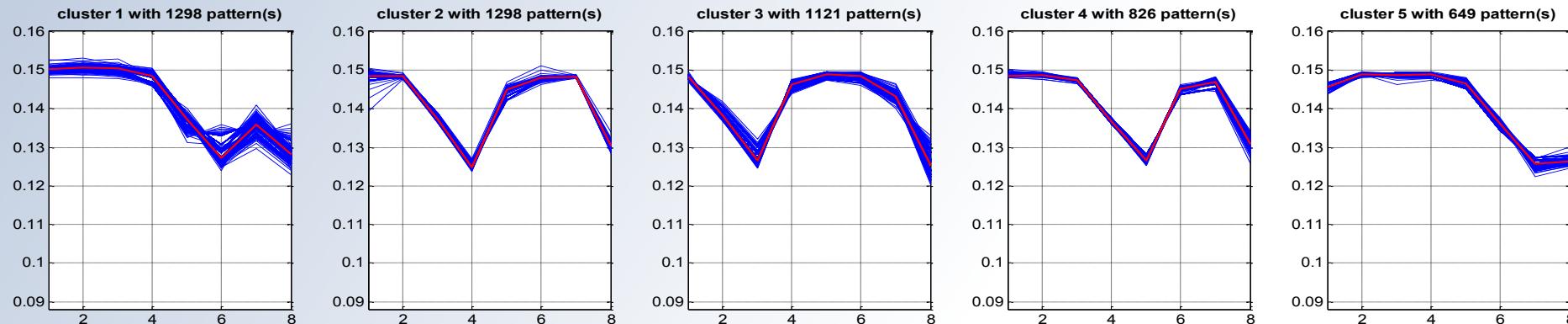
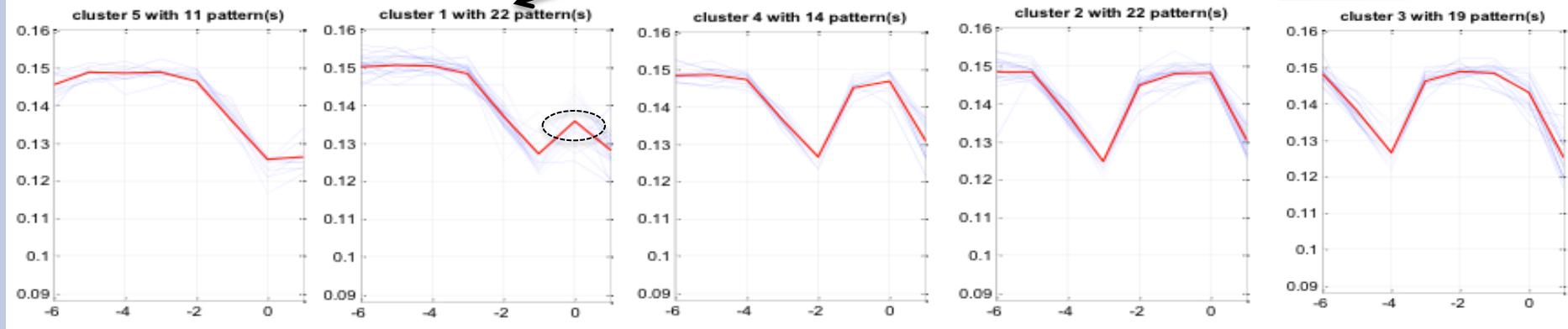


Power systems - *outliers*

- ◆ Short-term forecasting
 - Further improvement limited by existing variable
- ◆ Expand applicability with densification and outlier analysis
 - Analysis of special days
 - few data points available
 - holidays usually happen on different days every year
 - Densification with mean-shift

Power systems - *outliers*

Extended weekend



Power systems - *outliers*

- ◆ Initial results
 - Multivariate regression and ANN:
 - 7 days window input
 - Special day value as output
- ◆ With densification

Special day on	Test data available	NMAE	Std. dev.
monday	11	1.85%	1.76%
tuesday	22	2.82%	2.50%
wednesday	14	3.41%	1.90%
thursday	22	2.55%	1.66%
friday	19	3.92%	2.46%

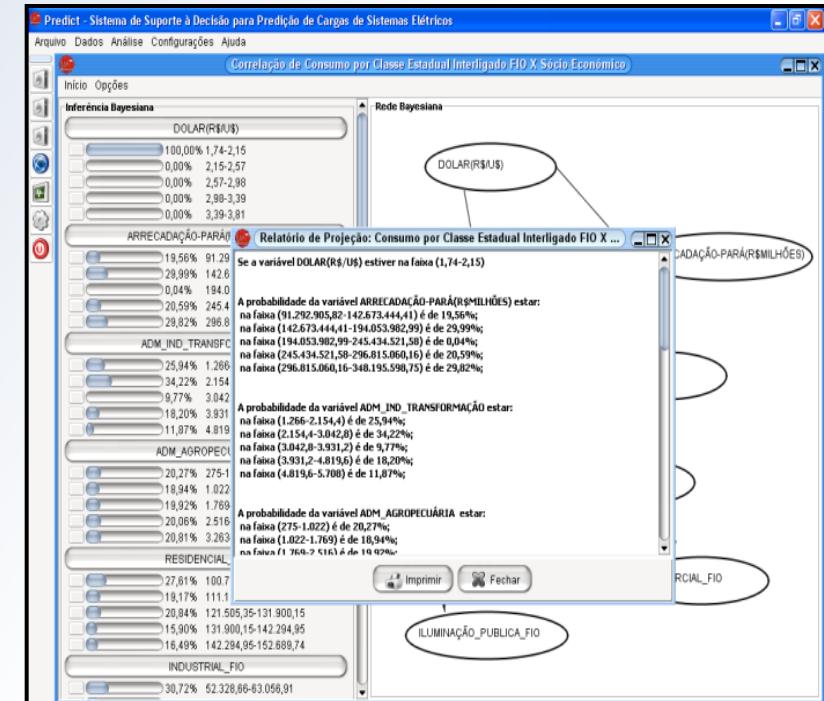
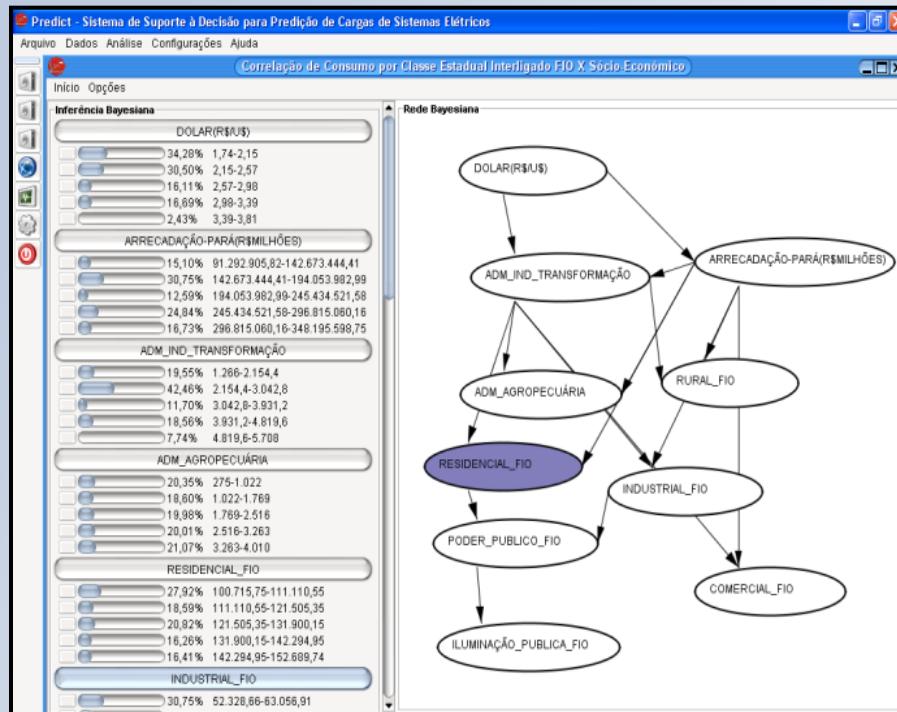
- ◆ Without densification

Model	Special days	Normal days
MvR	9.82%	2.49%
ANN	9.01%	2.29%

Power systems - *correlation*

- ◆ DSS implementing functionalities for data correlation and scenario analysis:
 - Find best scenario to achieve the company's goals
 - Correlate variables and quantify the impact toward achieving a given goal
 - Consider the expert knowledge in the scenarios discovery process
 - Estimate the behaviour of the systems/domain

Power systems



Related application...

- ◆ Adaptive control in electric power systems
 - hybrid approach for maintaining synchronization between generators (for both, steady state and during the occurrence of disturbances)

Public Safety (Gov.) [4]

- ◆ DSS developed as a focus for analyzing and identifying patterns and ranges for criminality based on areas, degree of violence and types of crimes
 - Red - districts with high rates of violent crimes
 - Yellow - districts with moderate violent crimes
 - Greens - districts with low rates of violent crimes
- ◆ Implementation of models for scenario analysis and data visualization
- ◆ Dataset with over 1Mi records
 - Several records with inconsistency and missing data

Public Safety

Sistema Web de Suporte à Decisão para Gestão Governamental

Início Dados Análise Configurações Ajuda

PRODEPA

Pará GOVERNO DO ESTADO

PRAD

Sistema Web de Suporte à Decisão para Gestão do Estado do Pará

Início Dados Análise Configurações Ajuda

Ocorrências

Emissão	Idend.Bo	Ident.Bairro	Bairro	Cod. Local.Ocorrência	Local da Ocorrência	Logradoiro	Cod. Crime	Crime	Unidade Policial
4/2003	2003000010	4593	Umanzal	7	Café,Bar,Etc	Senador Lemos, Aven...	230	Art.1 - Omr informa...	35
4/2003	2003000011	4579	Pedreira	20	Outros	Prímeiro de Dezembr...	68	Furto Sem Arrombamen...	6
4/2003	2003000012	4545	Campina	17	Via Pública	Pedre Prudêncio, Tra...	228	Art.1 - Negar/delivrar...	35
4/2003	2003000013	4570	Imigrante	15	Residência Particular	Bernel do Couto, Rua...	40	Art.149 - Lesão corporal	5
4/2003	2003000013	4593	Umanzal	15	Residência Particular	Bernel do Couto, Rua...	47	Art.129 - Lesão corporal	5
4/2003	2003000015	4579	Pedreira	16	Transporte	Barão do Trinúfo, Tra...	50	Outros crimes contra...	5
4/2003	2003000016	4545	Campina	20	Outros	Castro França, Bou...	68	Furto Sem Arrombamen...	6
4/2003	2003000018	4579	Pedreira	4	Estabelecimento de E...	Pedro Miranda, Avenida	39	Agressão	5
4/2003	2003000019	4579	Pedreira	4	Estabelecimento de E...	Humaitá, Travessa	39	Agressão	5
4/2003	2003000020	4545	Campina	17	Via Pública	Assis de Vasconcelos...	68	Furto Sem Arrombamen...	6
4/2003	2003000020	4579	Pedreira	17	Via Pública	—	66	Furto de Veículo	6
4/2003	2003000022	4579	Pedreira	17	Via Pública	Pedro Miranda, Avenida	270	Art.156 - Furto de coi...	6
4/2003	2003000023	4579	Pedreira	17	Via Pública	Pedro Miranda, Avenida	229	Extravio de Documentos	25
4/2003	2003000024	4545	Campina	17	Via Pública	Assis de Vasconcelos...	68	Furto Sem Arrombamen...	6
4/2003	2003000026	4579	Pedreira	17	Via Pública	Marquês de Herval, A...	39	Agressão	5
4/2003	2003000027	4545	Campina	17	Via Pública	Assis de Vasconcelos...	70	Art.156 - Furto de coi...	6
4/2003	2003000028	4545	Campina	23	Loyerado Público	Castilhos França, Bou...	68	Furto Sem Arrombamen...	6
4/2003	2003000030	45254	Cabangen	17	Via Pública	Principal, Rua	40	Art.147 - Ameaça	5
4/2003	2003000030	4579	Pedreira	15	Residência Particular	Barão do Trinúfo, Tra...	68	Furto Sem Arrombamen...	6
4/2003	2003000031	4593	Umanzal	20	Outros	Dom Pedro I, Travessa	81	Art.171 - Estelionato	6
4/2003	2003000031	4579	Batista Campos	20	Outros	Presidente Vargas, B...	219	Advertência a Boletim...	21
4/2003	2003000032	4579	Pedreira	16	Transporte	Pedro Miranda, Avenida	68	furto de Veículo	6
4/2003	2003000034	4579	Pedreira	17	Via Pública	Campos Sales, Travessa	68	Outro Sem Arrombamen...	6
4/2003	2003000035	4545	Campina	17	Via Pública	Chaco, Travessa	73	Art.157 - Roubo	6
4/2003	2003000037	4579	Pedreira	9	Casa Comercial	Pedro Miranda, Avenida	40	Art.147 - Ameaça	5
4/2003	2003000038	4579	Pedreira	15	Residência Particular	Portugal, Avenida	68	Furto Sem Arrombamen...	6
4/2003	2003000039	4545	Campina	17	Via Pública	Augusto Corrêa, Rua	40	Destrução de Bens (CIV)	21
4/2003	2003000040	4579	Pedreira	9	Casa Comercial	Barão do Trinúfo, Tra...	73	Art.157 - Roubo	6
4/2003	2003000041	4579	Pedreira	17	Via Pública	Pedro Miranda, Avenida	229	Extravio de Documentos	25
4/2003	2003000042	4579	Pedreira	17	Via Pública	Gentil Bitencourt, Av...	231	Outros Motivos	21
4/2003	2003000043	4545	Batista Campos	17	Via Pública	Alferez Costa, Travessa	40	Art.147 - Ameaça	5
4/2003	2003000043	4541	Pedreira	17	Via Pública	Augusto Corrêa, Rua	40	Art.157 - Roubo	6
4/2003	2003000044	4579	Pedreira	15	Residência Particular	Augusto Corrêa, Rua	40	Art.147 - Ameaça	5
4/2003	2003000044	4545	Reduto	15	Via Pública	Caixa, Av. das Amazô...	40	Art.157 - Roubo	6
4/2003	2003000045	4579	Pedreira	17	Casa Comercial	Quinz de Novembro, ...	73	Art.147 - Roubo	6
4/2003	2003000049	4545	Pedreira	17	Via Pública	Pedro Miranda, Avenida	47	Art.129 - Lesão corporal	5
4/2003	2003000049	4579	Pedreira	17	Via Pública	Augusto Corrêa, Rua	229	Extravio de Documentos	25
4/2003	2003000050	4545	Campina	17	Via Pública	João Paulo II, Avenida	68	Furto Sem Arrombamen...	6
4/2003	2003000050	4560	Guamá	17	Via Pública	Augusto Corrêa, Rua	229	Extravio de Documentos	25
4/2003	2003000051	4557	Curió-Utinga	7	Café,Bar,Etc	João Paulo II, Avenida	68	Furto Sem Arrombamen...	6

Exportar Fazer

Public Safety

Correlação

Menu ▾

- Bairro**

Estado	Probabilidade	Inferência
São Brás	0 %	...
Umarizal	100 %	...
Bengui	0 %	...
- Crime**

Estado	Probabilidade	Inferência
Manter casa de prostituição	0.01%	...
Jogo de Azar	0.00%	...
Art.311 - Adulteração de sinal	0.01%	...
- Unidade Policial**

Estado	Probabilidade	Inferência
AURÁ UP	0.01%	...
DCCPe-SUSB	3.58%	...
DCCPa - SUSB	4.83%	...
- Unidade Local Ocorrência**

Estado	Probabilidade	Inferência
Café,Bar,Etc	1.57%	...
Hotel, Pensão, Etc	0.23%	...
Habitação Coletiva	0.32%	...

Atributo: Selecionar...

11.37% - Art.157 - Roubo

Sistema Web de Suporte à Decisão para Gestão do Estado do Pará

Correlação

Início Opções

Inférfia Bayesiana

- MÊS
- BAIRRO
- CLASSE CRIME
- CRIME
- UNIDADE POLICIAL
- LOCAL DA OCORRÊNCIA
- LOGRADOURO

Rede Bayesiana

```

graph TD
    CC([CLASSE CRIME]) --> U([UNIDADE POLICIAL])
    CC --> M([MÊS])
    CC --> LO([LOCAL DA OCORRÊNCIA])
    U --> M
    U --> LO
    M --> LO
    C([CRIME]) --> LO
    
```

Análise Temporal Markoviana

Análise de Cenários

Mapa

Crime

- Manter casa de prostituição
- Art.311 - Adulteração de sinal identificador de veículo
- Art.168 - Apropriação indébita
- Outros crimes contra pessoa
- Denunciação caluniosa
- Assistente Social - Conflitos Conjugais
- Perturbações da Tranqüilidade
- Vias de Fato
- Extravio, Furto de Placa de Veículo
- Furto em Veículo
- Ato obsceno
- Art.50, A - Desmatar, explorar economicamente/deg
- Fuga do Lar
- Sequestro Reilmpago
- Art.251- Explosão
- Furto Sem Arrombamento
- Roubo em Veículo
- Art.301 - Falsidade - material de prova/condição

Gerar

Mapa | **Satélite** | **Híbrido**

Mapa | Satélite | Híbrido

Google

1000 pés
500 m

Mapa | **Satélite** | **Híbrido**

Google

1000 pés
500 m

Dados cartográficos ©2011 MapLink

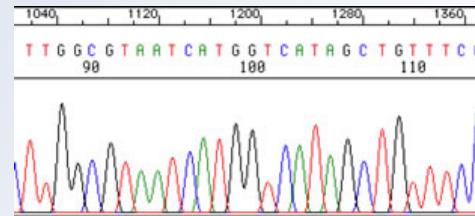
Logotipo de Usuário

Related application...

- ◆ Big data (over 1Mi) and data stream
 - Optimize time/tractability for parameter and structure learning with Bayesian networks in large datasets using MapReduce and non-relational databases

Bioinformatics [5]

- ◆ Over 5Gb of text data, per genetic analysis
- ◆ Objectives
 - Increase the reliability of data generated by next generation sequencer – Solid
 - Implementation of a pipeline to evaluate the quality and classify the data generated into biological characteristic types, such as: tRNA, cDNA, mRNA



```
>seq1  
gcagtcagtcacacatgtca...  
>seq2  
cgccatgcgcgtactctat...  
>seq3  
tcgagcatcatcagtcgtca...  
>seq4  
tatgcttatagcgagtcat...  
.....
```

Health (Gov.) [6]

- ◆ Analysis of governmental data in the area of health
- ◆ Create portrayal of health investments in the State of Pará
- ◆ Inference on possible scenarios for investments and projects
 - Find best projections to improve quality of life for the population

Livestock [7]

- ◆ Analysis of livestock data to find best mating patterns among animals for the Nelore Breeding Program
- ◆ Consider the cattle's features :
 - Fertility
 - Maternal ability
 - Growth
- ◆ Discover levels of genetic influence of the cow's attributes to the progeny (calves)

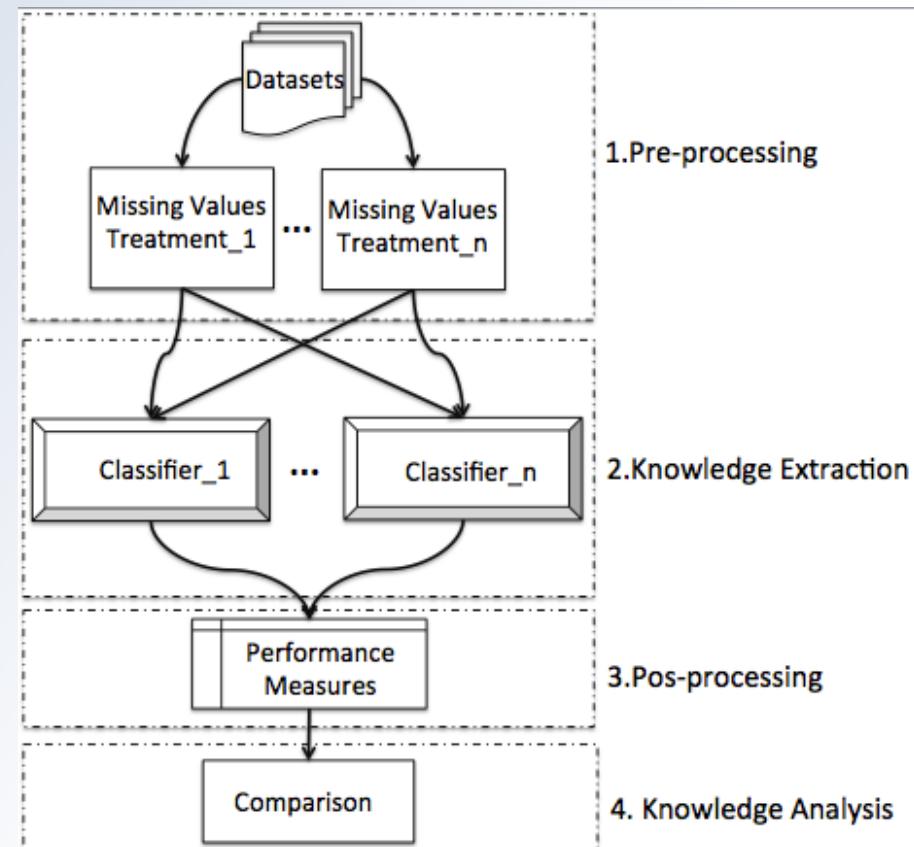
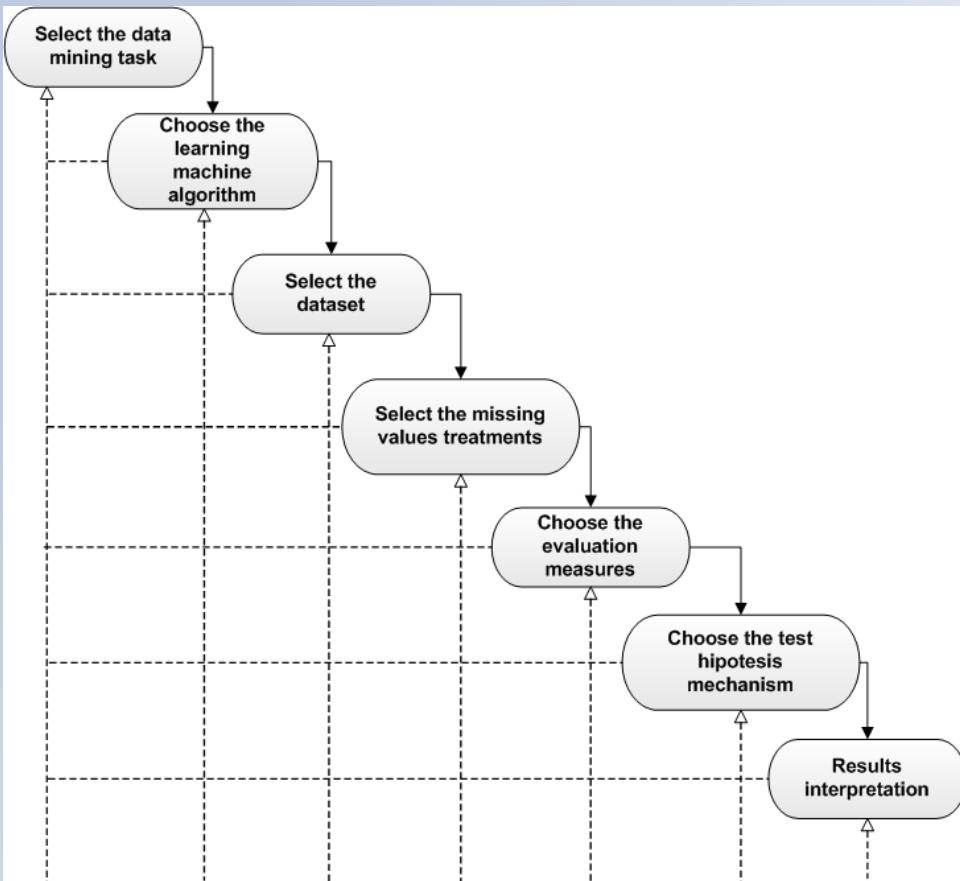
Telecommunications [8]

- ◆ Data analysis to:
 - Identify disturbances that may affect the level of speech quality and signaling in a cellular system
 - Identify channels or sources responsible for dropped calls
 - Guarantee minimum QoS (Quality of Service) and QoE (Quality of Experience) levels in DSL networks for Triple Play (voice, video and data)

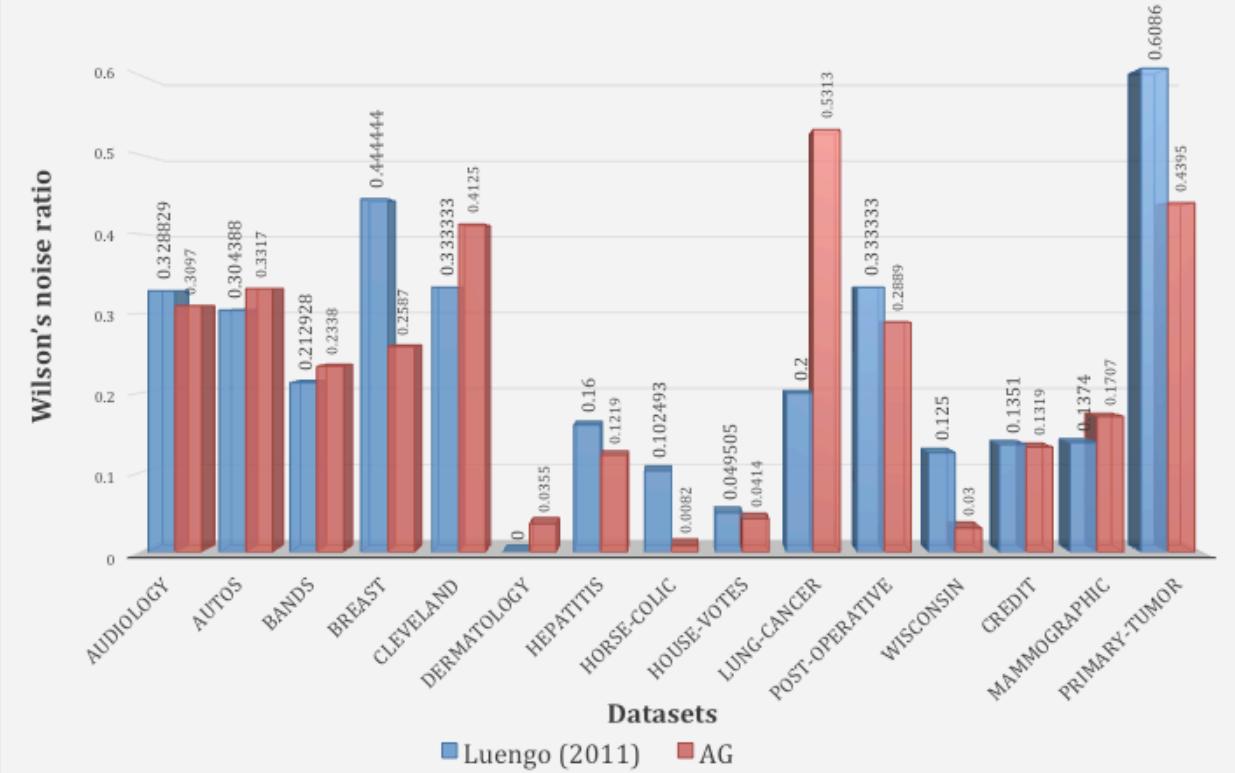
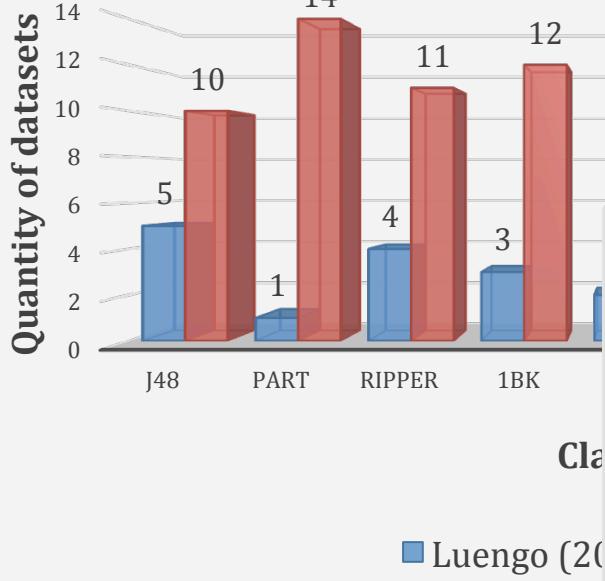
Missing Values Treatments

- ◆ Problems observed
 - Different datasets (and performance measures) used in the missing values treatment → *unable to compare recent researches*
 - Impact of MV in data mining application
- ◆ Provide ways to make experiments more replicable [9]
- ◆ Improve accuracy of data imputation with evolutionary algorithms [10]

Protocol

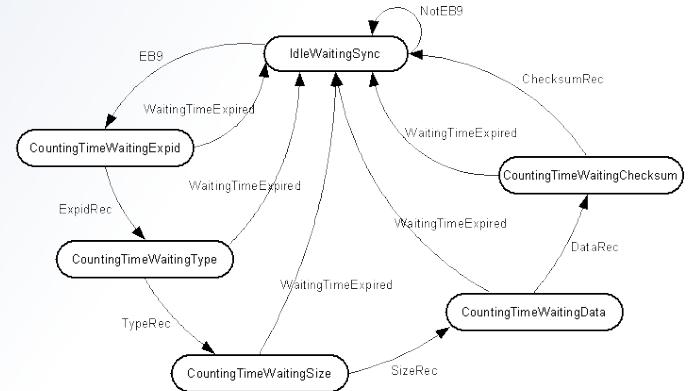
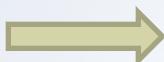
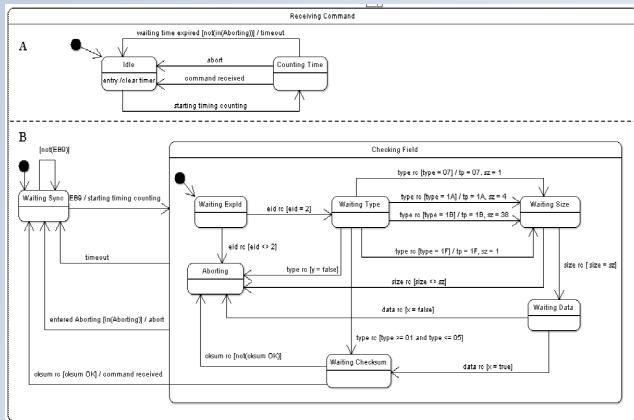


Promising results with GAs

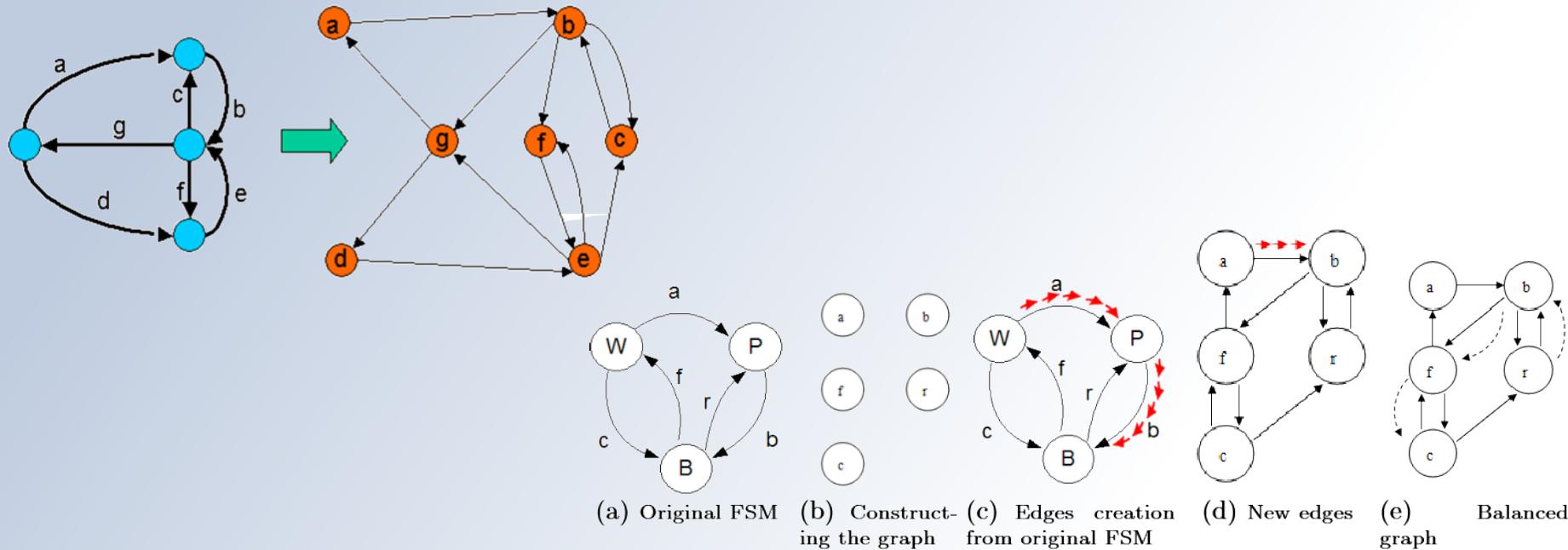


Automatic software testing

- ◆ Automatic generation of test sequences for black box (software) models
 - Critic/parallel systems modeled with statecharts
- ◆ CI optimization over Switch Cover
 - Graph balancing and optimization with PSO



Automatic software testing



- ♦ Reduced time to generate sequences: $\sim 8.4x$ faster
- ♦ Reducing the number of test steps/cases

Related Publications

- ◆ [1] Decision support system for load forecasting and inference: a new undertaking for Brazilian power suppliers. *Int. Journal of Electrical Power & Energy Systems*, 2012.
- ◆ [2] Strategies for Improving the Modeling and Interpretability of Bayesian Networks. *Data & Knowledge Engineering*, 2007.
- ◆ [3] Densification trick using mean shift to allow demand forecasting in special days with scarce data. In: 17th Int. Conf. on Intelligent System Applications to Power Systems, 2013.
- ◆ [4] Web tool for decision support in public safety management. In: Proceedings of the X Brazilian Congress on Computational, 2011.
- ◆ [5] Probabilistic approach for characterizing the marking system of multiplex sequencing in ABI SOLiD platform. In: Int. Conf. on Bioinformatics and Computational Biology, 2011.
- ◆ [6] Health Monitoring in the Amazon - Decision Support System for Public Health Managers. In: Proceedings of the Environmental and Health World Congress, 2006.
- ◆ [7] Bayesian Networks for Knowledge Discovery in a Database from the Program for Genetic Improvement of the Nelore Breed. In: 2nd Int. Conf. on Data Mining, 2000.
- ◆ [8] Bandwidth Scheduler for Triple Play Services based on Fuzzy Theory. In: 12th IFIP/IEEE Int. Symposium on Integrated Network Management, 2011
- ◆ [9] Testbed for the Experiments Performed in Missing Value Treatments. In: International Conference on Information Society, 2013
- ◆ [10] Runtime Performance Evaluation of GPU and CPU using a Genetic Algorithm Based on Neighborhood Model. In: Int. Conf. on Parallel and Distributed Processing Techniques and Applications, 2013

Development Tools*



PostgreSC



Redis



mongoDB



Cassandra



Riak

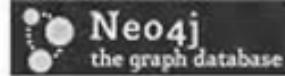
HTML



CSS



relax



Tokyo Cabinet



/* the Business Intelligence Free Platform */

JASPER SOFT

Laboratory of Computational Intelligence and Operational Research – LINC/UFPA



Dr. Adamo Santana (coordinator)

adamo@ufpa.br

+55 91 3201-8112 / +55 91 8412-9598

