

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

ROMMELL GUIMARÃES CAIXETA

**Uma abordagem computacional para
predição de mortalidade em UTIs
baseada em agrupamento de Processos
Gaussianos**

Goiânia
2016

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1 **1. Identificação do material bibliográfico:** **Dissertação** **Tese**

1 **2. Identificação da Tese ou Dissertação**

2

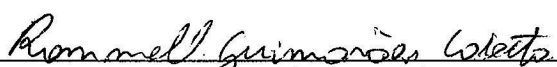
Nome completo do autor: Rommell Guimarães Caixeta

Título do trabalho: **Uma Abordagem Computacional para Predição de Mortalidade em UTIs Baseada em Agrupamento de Processos Gaussianos**

3. Informações de acesso ao documento:

Concorda com a liberação total do documento SIM NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.


Assinatura do (a) autor (a)

Data: 27 / 09 / 2016

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

ROMMELL GUIMARÃES CAIXETA

Uma abordagem computacional para predição de mortalidade em UTIs baseada em agrupamento de Processos Gaussianos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Ciência Da Computação.

Área de concentração: Ciência da Computação.

Orientador: Prof. Dr. Renato de Freitas Bulcão Neto

Coorientador: Prof. Dr. Anderson da Silva Soares

Goiânia
2016

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Caixeta, Rommell Guimarães
Uma Abordagem Computacional para Predição de Mortalidade em UTIs Baseada em Agrupamento de Processos Gaussianos [manuscrito] / Rommell Guimarães Caixeta. - 2016.
lix, 59 f.: il.

Orientador: Prof. Dr. Renato de Freitas Bulcão Neto; co-orientador Dr. Anderson da Silva Soares.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Goiânia, 2016.

Bibliografia.

Inclui siglas, algoritmos, lista de figuras, lista de tabelas.

1. mineração de dados. 2. sinais vitais. 3. séries temporais. 4. classificação. 5. processos gaussianos. I. Bulcão Neto, Renato de Freitas, orient. II. Título.

CDU 004



ATA Nº 15/2016

ATA DA SESSÃO DE JULGAMENTO DA DISSERTAÇÃO
DE MESTRADO DE ROMMELL GUIMARÃES CAIXETA

Aos nove dias do mês de setembro de dois mil e dezesseis, às catorze horas, na sala 150 do Instituto de Informática da Universidade Federal de Goiás, Campus Samambaia, reuniu-se a banca examinadora designada na forma regimental pela Coordenação do Curso para julgar a dissertação de mestrado intitulada "**Uma abordagem computacional para predição de mortalidade em UTIs baseada em agrupamento de Processos Gaussianos**", apresentada pelo aluno Rommell Guimarães Caixeta como parte dos requisitos necessários à obtenção do grau de Mestre em Ciência da Computação, área de concentração Ciência da Computação. A banca examinadora foi presidida pelo orientador do trabalho de dissertação, Professor Doutor Renato de Freitas Bulcão Neto (INF/UFG), tendo como membros o Professor Doutor Anderson da Silva Soares (coorientador – INF/UFG), o Professor Doutor Gustavo Teodoro Laureano (INF/UFG) e o Professor Doutor Marco Antônio Assfalk de Oliveira (EMC/UFG). Aberta a sessão, o candidato expôs seu trabalho. Em seguida, o aluno foi arguido pelos membros da banca e:

tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de sua dissertação, a banca concluiu pela **aprovação** do candidato, sem restrições.

tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de sua dissertação, a banca concluiu pela **aprovação** do candidato, condicionado a satisfazer as exigências listadas na Folha de Modificação de Dissertação de Mestrado anexa à presente ata, no prazo máximo de 60 dias, a contar da presente data, ficando o professor orientador responsável por atestar o cumprimento dessas exigências.

não tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de sua dissertação, a banca concluiu pela **reprovação** do candidato.

Os trabalhos foram encerrados às 17 horas. Nos termos do Regulamento Geral dos Cursos de Pós-Graduação desta Universidade, lavrou-se a presente ata que, lida e julgada conforme, segue assinada pelos membros da banca examinadora.

Prof. Dr. Renato de Freitas Bulcão Neto Renato Bulcão Neto
Prof. Dr. Anderson da Silva Soares Anderson da Silva Soares
Prof. Dr. Gustavo Teodoro Laureano Gustavo Teodoro Laureano
Prof. Dr. Marco Antônio Assfalk de Oliveira Marco Antônio Assfalk de Oliveira

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Rommell Guimarães Caixeta

Bacharel em Ciência da Computação pela Universidade de Brasília e Mestrando em Ciência da Computação pela Universidade Federal de Goiás. Sua pesquisa teve foco no uso de métodos computacionais na área da saúde. Atua como desenvolvedor de software no Tribunal Regional do Trabalho da 18ª Região. Suas áreas de interesse incluem aprendizagem de máquina e mineração de dados.

Dedico este trabalho à toda a minha família. De modo especial à minha mãe Jacirene Guimarães Nascimento(*in memoriam*), que me ensinou a nunca desistir de um sonho e ao meu pai José Tarcísio Caixeta, que desde muito cedo me proveu meios e incentivos para seguir este caminho.

Agradecimentos

Agradeço à Marcelo Tristão de Macedo, Rosana Tristão, Evani Tristão e Livia Tristão, por todo o amor, apoio e compreensão.

Ao meu orientador Prof. Dr. Renato de Freitas Bulcão Neto. Por muitas coisas, mas sobretudo por acreditar no meu trabalho mesmo quando havia grandes obstáculos a serem superados. Serei eternamente grato pela oportunidade concedida.

Ao meu coorientador Prof. Dr. Anderson da Silva Soares, pela ajuda na compreensão de problemas importantes durante o trabalho.

Ao Dr. David Clifton e toda a equipe de seu laboratório no Instituto de Engenharia Biomédica da Universidade de Oxford, que foi uma das fontes de inspiração do presente trabalho tanto pela gentileza e cordialidade na disseminação científica quanto no uso de métodos probabilísticos.

Aos colegas do Instituto de Informática da Universidade Federal de Goiás e do Tribunal Regional do Trabalho da 18ª Região pelo intercâmbio de idéias e experiências que contribuíram com este trabalho.

Resumo

Caixeta, Rommell Guimarães. **Uma Abordagem Computacional para Predição de Mortalidade em UTIs Baseada em Agrupamento de Processos Gaussianos**. Goiânia, 2016. 59p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

A análise das variáveis fisiológicas de um paciente pode melhorar a classificação do risco de óbito de um paciente em uma Unidade de Terapia Intensiva(UTI) e auxiliar na tomada de decisões e alocação de recursos disponíveis. Este trabalho propõe uma abordagem computacional de análise de variáveis fisiológicas para previsão de óbito de pacientes em UTI. Variáveis fisiológicas que compõem séries temporais(e.g., pressão arterial) são representadas como Processos Gaussianos Dependentes(PGDs). Variáveis que não representam séries temporais(e.g., idade) são utilizadas para agrupar os PGDs com Árvores de Decisão. A classificação é feita de acordo com uma medida de distância que combina Deformação Temporal Dinâmica e divergência Kullback-Leibler. O resultado desta abordagem quanto ao desempenho de classificação é superior ao método padronizado SAPS-I já utilizado em UTI no conjunto de dados considerado para testes. O resultado é similar à outros métodos computacionais publicados pela comunidade de pesquisa. Os resultados comparando variações da abordagem proposta também mostram que há vantagem em utilizar o agrupamento de PGDs descrito.

Palavras-chave

Mineração de Dados, sinais vitais, séries temporais, classificação, Processo Gaussianos

Abstract

Caixeta, Rommell Guimarães. **A Gaussian Process Clustering Based Approach to Mortality Prediction in ICUs**. Goiânia, 2016. 59p. MSc.Dissertation. Instituto de Informática, Universidade Federal de Goiás.

The analysis of physiological variables of a patient can improve the death risk classification in Intensive Care Units(ICU) and help decision making and resource management. This work proposes a computational approach to death prediction through physiological variables analysis in ICU. Physiological variables that compounds time-series(e.g., blood pressure) are represented as Dependent Gaussian Processes(DGP). Variables that do not represent time-series (e.g., age) are used to cluster DGPs with Decision Trees. Classification is made according to a distance measure that combines Dynamic Time Warping and Kullback-Leibler divergence. The results of this approach are superior to other method already used, SAPS-I, on the considered test dataset. The results are similar to other computational methods published by the research community. The results comparing variations of the proposed method show that there is advantage in using the proposed clustering of DGPs.

Keywords

Data Mining, vital signs, time series, classification, Gaussian Process

Sumário

Lista de Figuras	10
Lista de Tabelas	11
Lista de Algoritmos	12
Acrônimos	13
1 Introdução	14
1.1 Motivação	14
1.2 Problema	16
1.3 Objetivos	16
1.4 Metodologia	16
1.5 Estrutura da dissertação	17
2 Fundamentação Teórica	18
2.1 Regressão e Variável Aleatória	18
2.2 Distribuição Gaussiana Multivariada	19
2.3 Processos Gaussianos	20
2.4 Processos Gaussianos Dependentes	23
2.5 Divergência KL	26
2.6 Deformação Temporal Dinâmica	28
2.7 Árvores de Decisão	29
2.8 Descrição dos dados	30
2.9 Considerações Finais	33
3 Metodologia e Resultados	34
3.1 Representação de Processos Gaussianos Dependentes	34
3.2 O cálculo da distância entre Processos Gaussianos Dependentes	41
3.3 Agrupamento com Árvore de Decisão	41
3.4 Classificação baseada em distância	43
3.5 Combinação e classificação	44
3.6 Resultados	45
3.7 Considerações Finais	45
4 Trabalho Relacionados	47
4.1 Trabalhos relacionados ao <i>PhysioNet Challenge 2012</i>	47
4.2 Processos Gaussianos Multi-saída	49
4.3 Considerações Finais	51

5	Conclusões	52
5.1	Limitações	52
5.2	Trabalhos futuros	53
5.3	Considerações Finais	54
	Referências Bibliográficas	55

Lista de Figuras

1.1	Causas de admissão na UTI do HRAN.	15
2.1	Trajatórias de Processos Gaussianos.	22
2.2	Amostras de um Processo Gaussiano com dados observados.	23
2.3	Distribuição posterior de um Processo Gaussiano.	24
2.4	Proc. Gaussiano <i>versus</i> Proc. Gaussiano Dependente.	27
2.5	Deformação Temporal Dinâmica.	29
2.6	Deformação Temporal Dinâmica com ruído.	30
3.1	Posterior da frequência cardíaca para um paciente.	37
3.2	Posterior da pressão arterial sistólica para um paciente.	37
3.3	Posterior da temperatura para um paciente.	38
3.4	Posterior do volume de urina para um paciente.	38
3.5	Posterior do bicarbonato sérico para um paciente.	39
3.6	Posterior da escala de Glasgow para um paciente.	39
3.7	Posterior do nitrogênio uréico no sangue para um paciente.	40
3.8	Posterior da contagem de glóbulos brancos para um paciente.	40
3.9	Árvore de decisão.	43

Lista de Tabelas

2.1	Séries temporais do <i>Challenge 2012</i> .	32
3.1	Resultados para o escore 1	45
4.1	Resultados do <i>Challenge 2012</i>	50

Lista de Algoritmos

3.1 *KNN*($pg, Grupo_PGs, classes$)

44

Acrônimos

AD	Árvore de Decisão.
BMI	<i>Body Mass Index.</i>
BUN	<i>Blood Urea Nitrogen.</i>
DTD	Deformação Temporal Dinâmica.
GCS	<i>Glasgow Coma Score.</i>
HR	<i>Heart Rate.</i>
KNN	<i>K-Nearest Neighbors.</i>
MIMIC-II	<i>Multiparameter Intelligent Monitoring in Intensive Care II.</i>
PG	Processo Gaussiano.
PGD	Processo Gaussiano Dependente.
SAPS	<i>Simplified Acute Physiology Score.</i>
SysABP	<i>Systolic Arterial Blood Pressure.</i>
UTI	Unidade de Terapia Intensiva.
WBC	<i>White Blood Cell.</i>

Introdução

A Unidade de Terapia Intensiva (UTI) é uma seção presente em hospitais para o cuidado de pacientes em condições críticas. Para auxiliar na avaliação do risco de óbito de um paciente, variáveis fisiológicas (e.g., frequência cardíaca, pressão arterial) são monitoradas e registradas. Entretanto, o monitoramento pode não ser suficiente para detectar alterações significativas no risco de óbito de um paciente, devido à esparsidade de leituras feitas pela equipe de enfermagem e de possível falha em identificar as alterações[14]. A análise das variáveis fisiológicas de um paciente pode melhorar a avaliação do risco de óbito e auxiliar na tomada de decisões e alocação de recursos disponíveis[22]. Este trabalho propõe uma abordagem computacional de análise de variáveis fisiológicas para previsão de óbito de pacientes em UTIs.

1.1 Motivação

Transferências não planejadas para UTIs, paradas cardíacas e mortes inesperadas são eventos a serem evitados no cuidado de pacientes em hospitais. Em artigo de Vieira[46], é reportado que 29,7% das internações em uma UTI de um hospital brasileiro são devidas à instabilidades hemodinâmicas, 43,3% devidas a insuficiência respiratória e 17,9% são internações de urgência ou emergência. Em contraste, 9% são por pós-operatório eletivo ou hemodiálise. Neste último caso as internações são planejadas. Esses números sugerem que grande parte das internações não são planejadas, sendo reações à situações de risco para o paciente. A necessidade de métodos que avaliem riscos para o paciente de maneira mais precisa é evidente. A razão para isso é que, conforme ilustrado na Figura 1.1, situações não planejadas (pela incerteza do risco para o paciente) no ambiente clínico estão mais frequentemente associadas à óbitos.

Banaee et al.[2] reportou que nos últimos anos a área de pesquisa de sistemas em monitoramento de pacientes mudou sua ênfase do raciocínio simples sobre leituras instantâneas(i.e. limites e pontuações de dados) para a utilização de técnicas de mineração de dados e aprendizagem de máquina[13]. Em artigo de Esfandiari et al. [12], é reportado

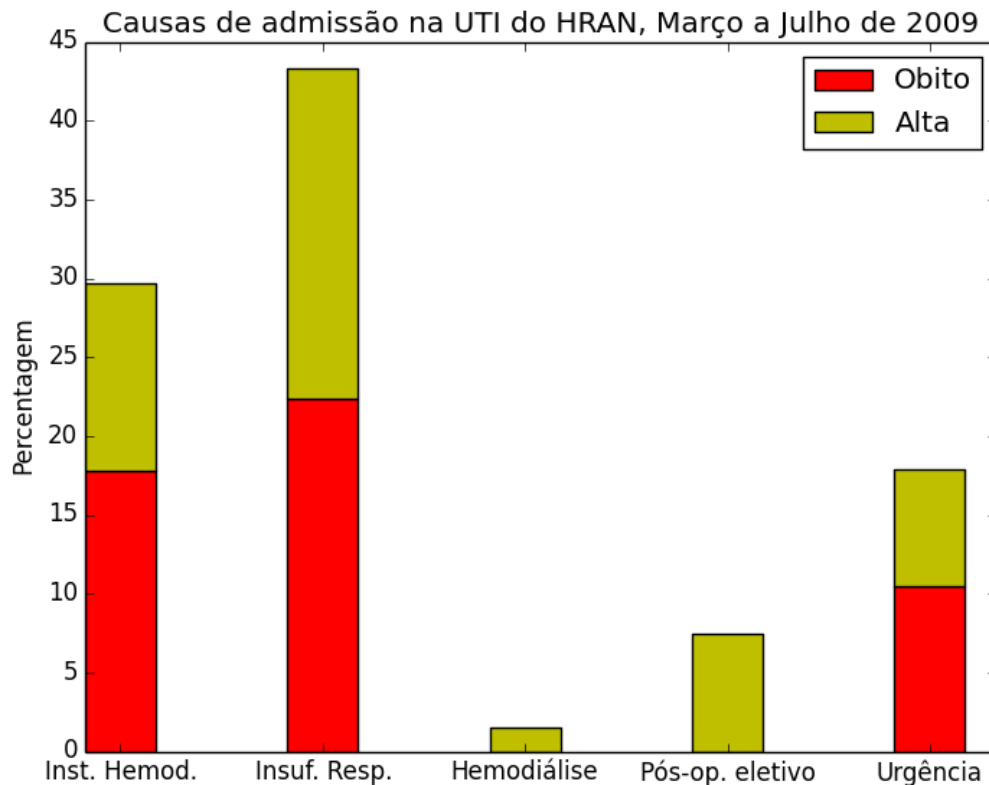


Figura 1.1: *Causas de admissão na UTI do Hospital Regional da Asa Norte(HRAN). A legenda mostra a proporção de admissões que resultaram em óbito ou sobrevivência.*

que a pesquisa em mineração de dados e aprendizagem de máquina na medicina cresce rapidamente.

Existem vários métodos de aprendizagem de máquina que são utilizados em aplicações na área da saúde, sendo alguns dos principais: Máquinas de Vetores de Suporte[10], Modelos de Mistura de Gaussianas[47], Modelos Baseados em Regras[41], Redes Neurais Artificiais[31], Modelos Ocultos de Markov[35] e Processos Gaussianos[36]. Entretanto, ainda não há consenso estabelecido na literatura sobre quais métodos são melhores em cada situação. De importância significativa na área da saúde é que o método utilizado seja interpretável[30], ou seja, que mostre relacionamentos entre os dados de entrada (no caso, variáveis fisiológicas) e os dados de saída (previsões). É evidente a necessidade de avaliação e comparação de vários métodos para cada aplicação.

Este trabalho é motivado pela crescente necessidade de métodos computacionais de previsão de óbito em UTIs assim como a necessidade de comparação de diferentes abordagens para cada aplicação.

1.2 Problema

Este trabalho propõe uma abordagem computacional de análise de variáveis fisiológicas para previsão de óbito de pacientes em UTIs. Este problema foi estruturado como tema de um desafio(*challenge*) feito pelo sítio *PhysioNet*[15] no ano de 2012. Foi um desafio aberto para a comunidade de pesquisa, onde objetivou-se a predição de morte de pacientes. Para isso, foram fornecidos como dados as séries temporais dos sinais vitais e outros dados fisiológicos das primeiras 48 horas de internação de pacientes em UTIs.

O desafio foi dividido em dois problemas. O primeiro problema consistiu em fazer uma classificação binária de cada paciente, rotulando cada um com a previsão de sobrevivência ou óbito. O segundo problema consistiu em quantificar a incerteza entre sobrevivência e óbito, atribuindo a cada paciente um valor para o risco de mortalidade. Neste trabalho, apenas o primeiro problema será abordado.

1.3 Objetivos

Este trabalho propõe uma abordagem para predição de óbito de pacientes a partir de características e séries temporais fisiológicas de pacientes integrando agrupamento de registros com Processos Gaussianos. Particularmente, propõe-se o agrupamento com Árvores de Decisão[7] e modelagem de séries temporais com Processos Gaussianos Dependentes[6]. Para realizar o agrupamento dos Processos Gaussianos, este trabalho propõe uma medida de distância baseada na divergência de distribuições probabilísticas entre pontos de dois Processos Gaussianos.

Os objetivos específicos compreendem:

- propor uma abordagem para representação de séries temporais baseada em Processos Gaussianos Dependentes aplicada à classificação de séries temporais fisiológicas.
- analisar e comparar a abordagem proposta a outros métodos existentes na literatura.

1.4 Metodologia

Esta seção apresenta as atividades realizadas, bem como a metodologia utilizada para alcançar os objetivos deste trabalho.

1. **Revisão Bibliográfica:** é constante e durante todas as etapas do trabalho, sendo que, ao início da pesquisa foi planejada e conduzida uma revisão sistemática abrangendo aprendizagem de máquina e modelos preditivos usando sinais vitais[8]. Esta revisão

identificou na literatura vários contextos de aplicação na área da saúde e métodos usados em cada contexto para mitigar dificuldades inerentes às tarefas de interesse.

2. **Definição do contexto:** à partir dos contextos identificados na revisão bibliográfica, foi escolhido como contexto do trabalho o monitoramento de pacientes em UTI por sua grande relevância. Especificamente, métodos de predição de óbito tendo como dados as primeiras 48 horas de internação na UTI. Conforme explicado, este contexto foi explorado pela comunidade científica na forma de um desafio, sendo que os métodos usados podem ser avaliados e comparados de forma objetiva. Esta definição também determinou o conjunto de dados a ser considerado para avaliação, também fornecido pelo desafio.
3. **Escolha da abordagem:** a revisão bibliográfica mostrou a utilização de vários modelos para representação de séries temporais de variáveis fisiológicas. Entre eles o modelo de Processos Gaussianos, que oferece possibilidades de análise de trajetórias, conforme descrito em Pimentel et al.[33]. Com a leitura de trabalhos relacionados ao desafio considerado, identificou-se que nenhum trabalho representava séries temporais como Processo Gaussianos. Como forma de explorar a aplicação do modelo de Processos Gaussianos no contexto considerado, escolheu-se uma abordagem que fizesse uso deste modelo de representação. A abordagem elaborada também define uma medida de distância derivada do modelo para fins de comparação, agrupamento e classificação de séries temporais.
4. **Implementação da abordagem:** para implementar o modelo e tarefas necessárias foi utilizada a linguagem de programação Python[45], assim como bibliotecas para computação científica disponíveis nesta linguagem.
5. **Teste e avaliação:** o referido desafio definiu a avaliação do desempenho da abordagem. Os resultados obtidos por outros métodos também podem ser verificados nas publicações relacionadas ao desafio. Especificamente para obter o escore utilizado, foi adotado procedimento de validação cruzada e repetições de testes com escolha aleatória de dados, de forma semelhante à reportada em trabalhos apresentados no desafio.

1.5 Estrutura da dissertação

Esta dissertação de qualificação está assim organizada: o Capítulo 2 compreende a fundamentação teórica sobre Processos Gaussianos, Processos Gaussianos Dependentes e demais assuntos relevantes, incluindo a descrição dos dados. O Capítulo 3 descreve métodos, experimentos e resultados. O Capítulo 4. descreve trabalhos relacionados e o Capítulo 5 apresenta conclusões, limitações e trabalhos futuros.

Fundamentação Teórica

Este capítulo descreve os conceitos para a elaboração da abordagem proposta. Este trabalho usa regressão para estimar pontos de séries temporais e a partir destas regressões calcular distâncias entre séries temporais. Estas distâncias podem ser usadas para agrupar e classificar conjuntos de séries temporais. São apresentadas a abordagem de Processos Gaussianos para regressão assim como a formulação multi-saída para modelagem de várias séries temporais e demais definições necessárias.

2.1 Regressão e Variável Aleatória

No contexto deste trabalho, um problema de regressão é um problema em que busca-se aprender um mapeamento entre entradas e saídas (de valores contínuos), dado um conjunto de treinamento de pares de entrada-saída. Considera-se n entradas de treinamento $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$ em um espaço de entradas \mathcal{X} (contínuo ou discreto). No caso da modelagem de séries temporais, considera-se o espaço \mathcal{X} como sendo de dimensão 1 e representando instantes de tempo. A i -ésima entrada de treinamento \mathbf{x}_i está associada com uma saída de treinamento y_i , que no caso mais simples é um valor escalar real. Assume-se que os valores de saída surgem como função f (desconhecida) das entradas em \mathbf{X} , e podem também sofrer interferência de ruído (também desconhecido). Os valores de saída são combinados em um vetor $\mathbf{y} = [y_1 \dots y_n]$. Assume-se então que cada observação y_i depende de uma variável aleatória (uma variável cujo valor é indeterminado) f_i , como a seguir:

$$y_i = f_i + \varepsilon_i, \quad (2-1)$$

onde $f_i = f(\mathbf{x}_i)$ e $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ ¹ é assumido como ruído Gaussiano independente e identicamente distribuído(i.i.d.).

¹A notação $P \sim Q$ denota a que a variável aleatória P tem distribuição Q . O símbolo \mathcal{N} denota distribuição Gaussiana.

Dessa forma, considera-se n variáveis aleatórias em um vetor $\mathbf{f} = [f_1 \dots f_n]^\top$ correspondentes ao vetor de entrada \mathbf{X} . Podem-se considerar hipóteses sobre a distribuição probabilística conjunta deste vetor de variáveis aleatórias. Uma hipótese considerada neste trabalho, é a de que as séries temporais de interesse podem ser consideradas como estacionárias, i.e., suas propriedades estatísticas (média, variância, etc) permanecem constantes ao longo do tempo. Por meio dessa hipótese, supõe-se que a função desconhecida f representa séries temporais estacionárias. Outra hipótese a ser feita é a de que qualquer subconjunto do vetor \mathbf{f} tem distribuição (aproximadamente) Gaussiana multivariada. Esta distribuição é explicada a seguir.

2.2 Distribuição Gaussiana Multivariada

Em Probabilidade e Estatística, a distribuição (conjunta) Gaussiana multivariada é uma generalização da distribuição Gaussiana univariada (unidimensional). Um vetor aleatório $\mathbf{f} \in \mathbb{R}^n$ com n componentes tem distribuição Gaussiana multivariada se qualquer combinação linear de seus n componentes tem uma distribuição Gaussiana univariada². A função $p(\mathbf{f})$ de densidade de probabilidade é dada por:

$$p(\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{f} - \boldsymbol{\mu})\right), \quad (2-2)$$

onde $\boldsymbol{\mu}$ é o vetor de média (de tamanho n) e Σ é a matriz de covariância (de tamanho $n \times n$). Aqui a notação $|\Sigma|$ denota o determinante da matriz Σ e a notação \top denota transposição. Usa-se a notação semelhante às distribuições Gaussianas univariadas para dizer que o vetor \mathbf{f} tem distribuição Gaussiana multivariada:

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma). \quad (2-3)$$

De acordo com as duas hipóteses sobre as séries temporais consideradas na seção anterior, assume-se *a priori* que a distribuição de \mathbf{f} é Gaussiana multivariada para qualquer n e $\mathbf{x}_i \in \mathcal{X}$. Quando valores de \mathbf{X} e \mathbf{y} são observados, deseja-se realizar previsões *a posteriori* (i.e., após observar dados) sobre valores não observados. O modelo de Processos Gaussianos oferece uma maneira de realizar tais previsões. Este modelo é explicado a seguir.

²Nota-se que a condição da combinação linear arbitrária de componentes ser Gaussiana univariada é necessária, i.e., mesmo que todos os componentes tenham distribuição marginal Gaussiana univariada, a distribuição conjunta não necessariamente é Gaussiana multivariada.

2.3 Processos Gaussianos

Processos Gaussianos (PG) podem ser entendidos como distribuições probabilísticas sobre valores de funções. Um PG é definido como uma coleção de variáveis aleatórias. Em um PG, para cada valor (determinístico) de entrada x_i está associada uma variável aleatória que descreve o valor de f_i . Qualquer conjunto finito dessas variáveis aleatórias f_i tem uma distribuição conjunta Gaussiana. Formalmente:

Definição 2.1 *Um processo Gaussiano é uma coleção de variáveis aleatórias, sendo qualquer subconjunto finito delas tendo uma distribuição Gaussiana multivariada.*

Dessa forma, assume-se uma distribuição Gaussiana multivariada³ sobre \mathbf{f} :

$$\mathbf{f} \mid \mathbf{X}, \theta \sim \mathcal{N}(0, \mathbf{K}) \quad (2-4)$$

onde 0 é o vetor zero⁴ de tamanho n e \mathbf{K} é uma matriz $n \times n$ de covariância dependente de \mathbf{X} e determinados parâmetros θ . A entrada (i, j) da matriz \mathbf{K} é igual a $k(x_i, x_j)$ onde $k(\cdot, \cdot)$ é uma função positiva definida parametrizada por θ . Neste contexto, $k(\cdot, \cdot)$ é chamada de função de covariância. Assume-se portanto que, *a priori*, \mathbf{f} tem distribuição conjunta Gaussiana multivariada.

Dadas observações e uma função de covariância, pode-se fazer previsões usando o modelo de Processo Gaussiano. Para fazer isso considera-se um ponto de teste \mathbf{x}_* e o valor latente associado f_* . De acordo com o modelo de Processos Gaussianos, a distribuição conjunta do vetor \mathbf{f} e da variável aleatória f_* também é Gaussiana multivariada e é obtida aumentando a Equação 2-4 com a nova variável aleatória f_* :

$$\begin{bmatrix} \mathbf{f} \\ f_* \end{bmatrix} \mid \mathbf{X}, \theta \sim \mathcal{N} \left(0, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^\top & \kappa \end{bmatrix} \right) \quad (2-5)$$

onde $\mathbf{k} = [k(\mathbf{x}_*, \mathbf{x}_1) \dots k(\mathbf{x}_*, \mathbf{x}_n)]^\top$ é o vetor $n \times 1$ formado da covariância entre \mathbf{x}_* e as entradas de treinamento em \mathbf{X} . O valor escalar κ é igual a $k(\mathbf{x}_*, \mathbf{x}_*)$.

Dada a suposição de ruído Gaussiano na Equação 2-1, pode-se expressar a distribuição conjunta sobre os valores observados \mathbf{y} e os valores (de teste) não observados y_* :

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \mid \mathbf{X}, \theta \sim \mathcal{N} \left(0, \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I} & \mathbf{k} \\ \mathbf{k}^\top & \kappa + \sigma^2 \end{bmatrix} \right) \quad (2-6)$$

onde \mathbf{I} é a matriz Identidade de tamanho n .

³Na fórmula, o símbolo \mid denota probabilidade condicional

⁴Pode-se considerar um vetor constante diferente do vetor zero, sem prejuízo para generalização.

Sendo a distribuição conjunta Gaussiana, pode-se condicionar em \mathbf{y} (detalhes em Rasmussen e Williams[36], seção A.2, págs.200-201) para obter:

$$y_* | \mathbf{y}, \mathbf{X}, \theta, \sigma^2 \sim \mathcal{N}(m(\mathbf{x}_*), v(\mathbf{x}_*)) \quad (2-7)$$

onde a média e a variância das predições é dada por

$$m(\mathbf{x}_*) = \mathbf{k}^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \quad (2-8)$$

$$v(\mathbf{x}_*) = \kappa + \sigma^2 - \mathbf{k}^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k} \quad (2-9)$$

Dessa forma, dada uma função de covariância definida por parâmetros θ , pode-se calcular uma distribuição preditiva Gaussiana para qualquer ponto de teste \mathbf{x}_* . De maneira mais geral, pode-se calcular uma distribuição preditiva Gaussiana multivariável pra um conjunto de quaisquer m pontos $\mathbf{X}_* = [\mathbf{x}_{1*} \dots \mathbf{x}_{m*}]$ da seguinte maneira:

$$m(\mathbf{X}_*) = \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \quad (2-10)$$

$$v(\mathbf{X}_*) = \mathbf{K}_{**} + \sigma^2 \mathbf{I} - \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{K}_* \quad (2-11)$$

onde \mathbf{K}_* é uma matriz $n \times m$ de covariâncias entre as entradas de treinamento \mathbf{X} e os pontos de teste \mathbf{X}_* . A matriz \mathbf{K}_{**} de $m \times m$ consiste das covariâncias entre os pontos de teste.

Escrevemos que uma função f é descrita por um Processo Gaussiano usando a seguinte notação, que especifica completamente o processo:

$$f(\mathbf{x}) \sim PG(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (2-12)$$

onde $m(\mathbf{x})$ é a função de média e $k(\mathbf{x}, \mathbf{x}')$ é função de covariância, também chamada de cerne (em inglês, *kernel*). É um requisito que o cerne seja definido positivo para que a Equação 2-4 seja normalizável e portanto uma distribuição probabilística. Por simplicidade, assume-se a média como sendo zero.

Uma função de covariância comumente utilizada é a Exponencial Quadrada decrescente (EQ), ou curva de Gauss. A função de covariância especifica a covariância entre pares de variáveis aleatórias

$$\text{cov}(f(\mathbf{x}_p), f(\mathbf{x}_q)) = k(\mathbf{x}_p, \mathbf{x}_q) = \exp(-\frac{1}{2} |\mathbf{x}_p - \mathbf{x}_q|^2). \quad (2-13)$$

Assumindo-se *a priori* que uma função f é descrita por um Processo Gaussiano, a especificação das funções de média e covariância com dados observados implicam

uma distribuição *a posteriori* sobre f . Para ilustrar, considera-se cada variável aleatória f_i como descrição para valor de uma função de \mathbf{x}_i e retira-se amostras destas distribuições. Podem ser geradas várias trajetórias possíveis deste processo estocástico (seqüências de variáveis aleatórias indexadas por instantes de tempo). A figura 2.1 ilustra 5 trajetórias, que mostram curvas suaves devido às características do cerne exponencial quadrático decrescente. Nesta figura não há dados observados, as trajetórias amostradas são arbitrárias. Cada trajetória corresponde a um vetor \mathbf{y} de tamanho 100 e respectivos valores de \mathbf{x}_* uniformemente espaçados.

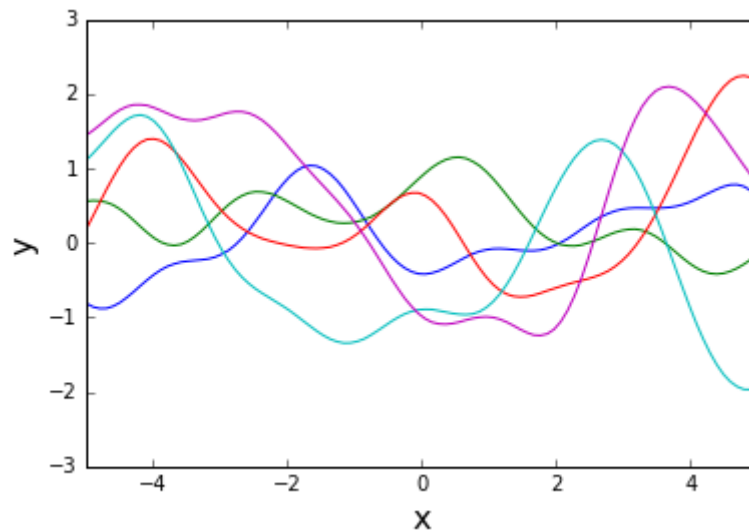


Figura 2.1: Trajetórias de Processos Gaussianos. As trajetórias aleatórias são geradas a partir de Processos Gaussianos.

Quando dados são observados, pode-se gerar trajetórias usando distribuições condicionadas a esses dados (equações 2-8 e 2-9). As trajetórias são obtidas de maneira semelhante à distribuição *a priori*, mas alterando a média e diminuindo a variância perto dos pontos observados, conforme ilustra a Figura 2.2.

A figura 2.3 mostra a distribuição preditiva com 10 pontos observados. Observa-se que a média realiza uma interpolação dos valores conhecidos da função, enquanto a variância permite estimar a incerteza em relação às estimativas: quando a estimativa é de um ponto perto de um outro ponto observado a incerteza é pequena, mas quando não há pontos observados perto do ponto estimado a variância aumenta, e com ela a incerteza da estimativa. Para cada ponto \mathbf{x}_* , é obtida uma distribuição Gaussiana univariada em que a média é dada por $m(\mathbf{x}_*)$ (Equação 2-8) e a variância é dada por $v(\mathbf{x}_*)$ (Equação 2-9).

Evidentemente, a função de covariância influencia no ajuste da interpolação da média e da variância. Modificando a função de covariância e introduzindo parâmetros, pode-se alterar o ajuste para que ele seja o mais adequado possível aos dados. Por

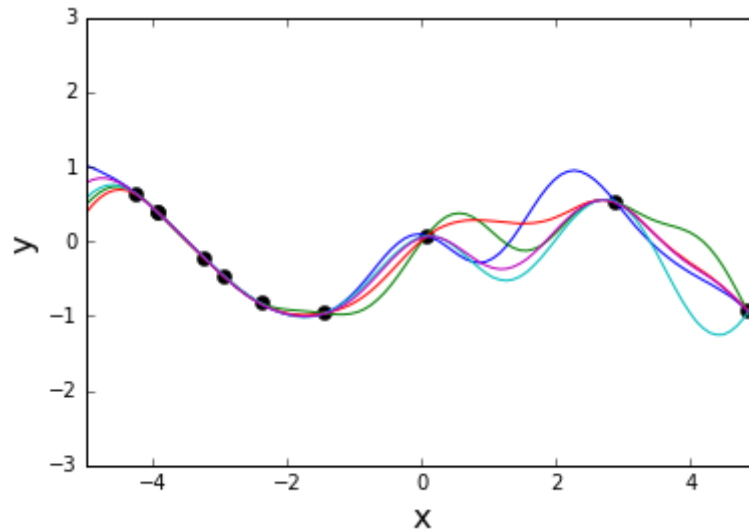


Figura 2.2: Amostras da distribuição a posteriori de um Processo Gaussiano.

exemplo, pode-se introduzir os parâmetros σ_f e l :

$$\text{cov}(f(\mathbf{x}_p), f(\mathbf{x}_q)) = k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |\mathbf{x}_p - \mathbf{x}_q|^2\right). \quad (2-14)$$

Métodos para conseguir um bom ajuste desses parâmetros são variados. Pode-se fazer isso, por exemplo, maximizando (em função dos parâmetros introduzidos na função de covariância) a *verossimilhança logarítmica*, dada por

$$\log p(\mathbf{f}|X) = -\frac{1}{2} \mathbf{f}^T K^{-1} \mathbf{f} - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi, \quad (2-15)$$

onde n é o número de pontos observados.

Dessa forma, realiza-se a aprendizagem de parâmetros em função dos dados observados, com o objetivo de se generalizar o modelo e fazer melhores previsões com dados não observados. Com esse ajuste, pode-se realizar estimativas e também obter informação sobre a incerteza dessa estimativa. Entretanto este modelo considera apenas uma única saída, i.e., regressões sobre apenas uma série temporal. Existem aplicações que consideram várias séries temporais potencialmente relacionadas. Um modelo que considera estas possíveis relações é o modelo de Processos Gaussianos Dependentes, explicado a seguir.

2.4 Processos Gaussianos Dependentes

O modelo de Processos Gaussianos Dependentes foi proposto por Boyle e Frean[6]. A descrição desse modelo feita aqui segue a apresentação feita pelos referidos

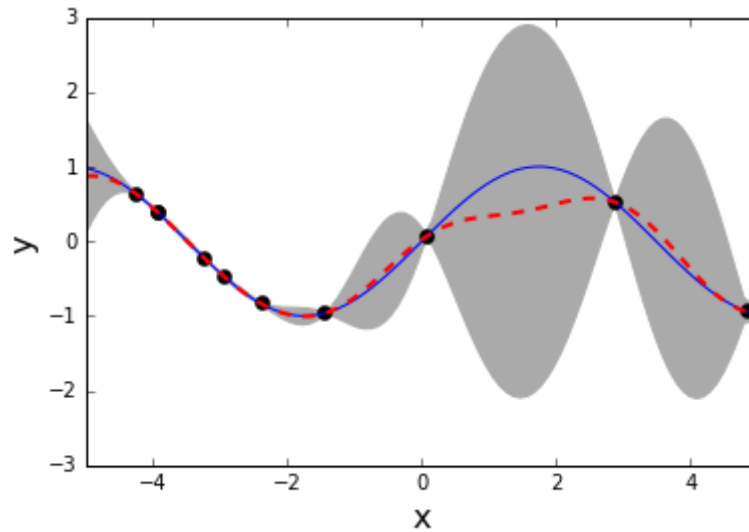


Figura 2.3: Distribuição posterior de um Processo Gaussiano. Dados observados são pontos pretos. A linha azul é a função original, a linha vermelha pontilhada é a média da distribuição probabilística posterior (i.e., condicionada aos pontos observados) e a região cinza é o intervalo de 3 desvios-padrão da posterior.

autores.

Nesta abordagem, os Processos Gaussianos são tratados como saídas de filtros lineares estáveis (Haykin [18] apud Boyle[5], p. 19). A saída de um filtro linear em resposta a uma entrada $w(\mathbf{x})$ é

$$\mathbf{y}(\mathbf{x}) = w(\mathbf{x}) * k(\mathbf{x}) = \int_{-\infty}^{\infty} w(\mathbf{x} - \tau)k(\tau)d\tau \quad (2-16)$$

onde $k(\tau)$ define a resposta ao impulso do filtro e $*$ denota operação de convolução. Supondo que o filtro linear é estável e $w(\mathbf{x})$ é ruído branco Gaussiano, o processo de saída $\mathbf{y}(\mathbf{x})$ é necessariamente um processo Gaussiano (Haykin [18] apud Boyle[5], p. 19). O ruído branco Gaussiano é um Processo Gaussiano particular no qual a covariância entre dois pontos quaisquer \mathbf{x}_i e \mathbf{x}_j é $\delta_{ij}\sigma^2$, onde δ_{ij} é a função delta de Kronecker e σ^2 é a variância do ruído. Este ruído modela um processo que é latente (não observado), mas que influencia a saída \mathbf{y} pela convolução com k . É possível também caracterizar filtros lineares estáveis com M -entradas (neste caso, ruídos) e N -saídas (\mathbf{y}) por um conjunto de $M \times N$ respostas ao impulso. As N saídas resultantes são processos Gaussianos Dependentes. A dependência se caracteriza pelo fato das saídas serem influenciadas pelos mesmos processos latentes, representados pelos ruídos brancos Gaussianos. Pode-se então, modelar múltiplas saídas dependentes pela parametrização do conjunto de respostas ao impulso para um filtro linear de múltiplas saídas, e inferir os valores dos parâmetros por dados observados. Ao invés de se parametrizar funções de covariância definidas positivas,

especifica-se e parametriza-se respostas ao impulso. A única restrição é que o filtro seja linear e estável, e isso é atingido pelo requisito de que as respostas ao impulso sejam absolutamente integráveis. Esta abordagem é equivalente a construir Processos Gaussianos pela convolução de cernes.

Considera-se como conjunto de dados $\mathcal{D} = \{\mathcal{D}_1 \dots \mathcal{D}_N\}$, onde $\mathcal{D}_i = \{(x_{i1}, y_{i1}) \dots (x_{in_i}, y_{in_i})\}$, sendo n_i o número de entradas da saída i .

Para modelar múltiplas saídas, são assumidos múltiplos processos latentes. Algumas saídas podem compartilhar processos latentes comuns e outras podem depender apenas de processos que afetam somente uma única saída. Dessa forma, pode-se definir N saídas, cada uma definida em um espaço de entradas de dimensão Dim . Em geral, para $x \in \mathbb{R}^{Dim}$, pode-se definir um modelo que assume M processos de ruído branco Gaussiano independentes denotados por $w_1(x) \dots w_M(x)$, N saídas denotadas por $y_1(x) \dots y_N(x)$ e $M \times N$ cernes. O cerne k_{mn} define a conexão da entrada m com a saída n . As funções de variância ($i = j$) e a covariância entre processos i e j são dadas por:

$$cov_{ij}(\mathbf{d}) = \sum_{m=1}^M \int_{\mathbb{R}^{Dim}} k_{mi}(x) k_{mj}(x + \mathbf{d}) d^{Dim} x, \quad (2-17)$$

onde \mathbf{d} denota $x_{ia} - x_{jb}$, a diferença entre entradas arbitrárias de índice a e b das saídas i e j , respectivamente.

Com isso, pode-se construir uma matriz de covariância \mathbf{C}_{ij} entre saídas i e j , da seguinte maneira:

$$\mathbf{C}_{ij} = \begin{bmatrix} cov_{ij}(x_{i1} - x_{j1}) & \dots & cov_{ij}(x_{i1} - x_{jn_j}) \\ \vdots & \ddots & \vdots \\ cov_{ij}(x_{in_i} - x_{j1}) & \dots & cov_{ij}(x_{in_i} - x_{jn_j}) \end{bmatrix} \quad (2-18)$$

Combinando as N saídas, define-se a matriz simétrica definida positiva \mathbf{C} :

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \dots & \mathbf{C}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{N1} & \dots & \mathbf{C}_{NN} \end{bmatrix}. \quad (2-19)$$

A verossimilhança logarítmica é dada por

$$\mathcal{L} = -\frac{1}{2} \log |\mathbf{C}| - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} - \frac{n}{2} \log 2\pi, \quad (2-20)$$

onde $\mathbf{y}^T = [(y_{11} \dots y_{1n_1}) \dots (y_{i1} \dots y_{in_i}) \dots (y_{N1} \dots y_{Nn_N})]$ e n é a soma da quantidade de entradas para cada saída n_i . Pode-se então, entre outras alternativas, encontrar os parâmetros das funções de covariância que maximizam a função de verossimilhança logarítmica para treinar o modelo, conforme o modelo de Processos Gaussianos de única saída.

A distribuição das predições para um ponto de teste x_* na i -ésima saída é Gaussiana com média $m(x_*)$ e variância $v(x_*)$ dadas por

$$m(x_*) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y} \quad (2-21)$$

e

$$v(x_*) = \kappa - \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k} \quad (2-22)$$

sendo $\kappa = cov_{ii}(0)$, $\mathbf{k}^T = [\mathbf{k}_1^T \dots \mathbf{k}_j^T \dots \mathbf{k}_N^T]$ e $\mathbf{k}_j^T = [cov_{ij}(x_* - x_{j1}) \dots cov_{ij}(x_* - x_{jn_j})]$.

A Figura 2.4 ilustra diferenças entre Processos Gaussianos de única saída e Processos Gaussianos Dependentes, mostrando também a capacidade dos últimos de capturar dependências entre múltiplas saídas. Quando há intervalos em que não há dados em uma das saídas, o modelo de única saída não tem informações suficientes para realizar uma boa estimativa (imagem superior direita). Já o modelo de Processos Gaussianos Dependentes consegue realizar estimativas mais próximas da função original por causa da dependência com dados de outra saída (imagem inferior direita).

Entretanto, este modelo não considera explicitamente comparar ou estabelecer uma medida de similaridade entre conjuntos de séries temporais, tarefa importante em várias aplicações com séries temporais. Uma forma de comparar séries temporais envolve primeiramente comparar pontos entre dois conjuntos de séries temporais pareadas (de mesmo tipo ou categoria). No contexto de comparação de séries temporais fisiológicas, um exemplo seria comparar as séries temporais de frequência de pulso de dois pacientes. O modelo de Processos Gaussianos Dependentes fornece distribuições Gaussianas em cada um desses pontos. Uma maneira de comparar distribuições probabilísticas é estabelecida pela Divergência-KL, explicada a seguir.

2.5 Divergência KL

Em Probabilidade e Teoria da Informação, a Divergência-KL (Kullback-Leibler, também chamada de discriminação de informação ou divergência de informação) é uma medida de diferença entre duas distribuições probabilísticas.

Para distribuições P e Q de uma variável aleatória contínua x , a Divergência-KL é definida pela integral:

$$KL(P||Q) = \int_{-\infty}^{\infty} \log \frac{p(x)}{q(x)} dx \quad (2-23)$$

onde p e q denotam as densidades de P e Q .

É possível mostrar que a Divergência-KL entre duas distribuições Gaussianas

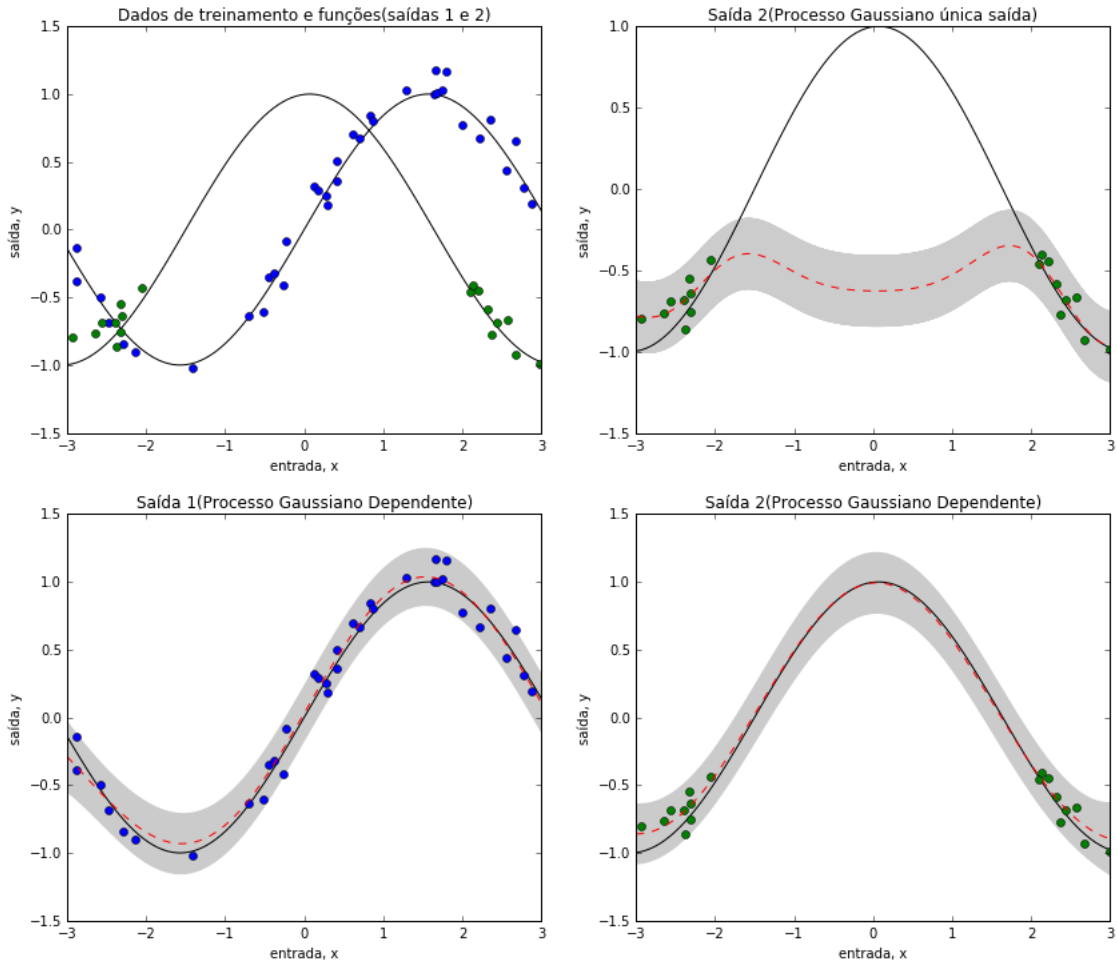


Figura 2.4: *Diferenças entre Processos Gaussianos e Processos Gaussianos Dependentes. Os dados foram amostrados das funções denotadas pelas linhas pretas. A linha pontilhada vermelha denota a função de média obtida com o modelo de Processos Gaussianos. A região cinza é o intervalo de 3 desvios-padrão da posterior.*

$P \sim \mathcal{N}(\mu_1, \sigma_1)$ e $Q \sim \mathcal{N}(\mu_2, \sigma_2)$ de uma única variável é dada por:

$$KL(P\|Q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \quad (2-24)$$

A Divergência-KL não é simétrica, portanto consideramos a média entre $KL(P\|Q)$ e $KL(Q\|P)$ como o valor para essa medida. Dessa forma, a Divergência-KL provê uma maneira de comparar dois pontos de séries temporais diferentes que tem distribuições probabilísticas associadas à esses pontos. Entretanto, é necessário determinar quais pontos entre duas séries temporais devem ser comparados. Uma abordagem utilizada para determinar esses pontos é a Deformação Temporal Dinâmica, explicada a seguir.

2.6 Deformação Temporal Dinâmica

A Deformação Temporal Dinâmica, DTD (em inglês, *Dynamic Time Warping*), é uma técnica que visa encontrar um alinhamento ótimo entre duas séries temporais se uma das séries pode estar deformada de maneira não linear ao longo do eixo temporal. Dadas duas séries temporais $\mathbf{Y} = [y_1 \dots y_{ny}]$ e $\mathbf{Z} = [z_1 \dots z_{nz}]$ de tamanhos ny e nz respectivamente, constrói-se um caminho de deformação $\mathbf{W} = [w_1 \dots w_{nw}]$ de tamanho nw tal que

$$\max(ny, nz) \leq nw < ny + nz. \quad (2-25)$$

O k -ésimo elemento de \mathbf{W} é uma tupla ordenada dada por

$$w_k = (i, j) \quad (2-26)$$

onde i é um índice da série temporal \mathbf{Y} e j é um índice da série \mathbf{Z} . O caminho de deformação deve iniciar com $w_1 = (1, 1)$ e terminar com $w_{nw} = (ny, nz)$. São impostas também as seguintes restrições:

$$w_k = (i, j), w_{k+1} = (i', j') \quad (2-27)$$

$$i \leq i' \leq i + 1, j \leq j' \leq j + 1 \quad (2-28)$$

O caminho de deformação ótimo é o caminho de distância mínima, dado por:

$$\text{Dist}(\mathbf{W}) = \sum_{k=1}^{k=nw} \text{Dist}(w_{ki}, w_{kj}) \quad (2-29)$$

onde $\text{Dist}(w_{ki}, w_{kj})$ é a distância entre dois pontos de dados (um em \mathbf{Y} e outro em \mathbf{Z}) no k -ésimo elemento de \mathbf{W} . A escolha mais simples para $\text{Dist}(w_{ki}, w_{kj})$ é o valor absoluto da diferença entre os pontos, i.e., $|w_{ki} - w_{kj}|$. A Figura 2.5 ilustra um pareamento possível para duas séries temporais aplicando a DTD.

Neste trabalho, será utilizada a Divergência-KL para calcular a distância $\text{Dist}(w_{ki}, w_{kj})$, já que o modelo de PGs fornece distribuições probabilísticas em pontos arbitrários de séries temporais. Os pontos sendo comparados correspondem ao valor da média do modelo de PGD avaliada nos pontos de treinamento, como ilustra a Figura 2.6.

Com isso, pode-se realizar classificação baseada em distância de séries temporais com base na comparação de distribuições probabilísticas obtidas em cada ponto de treinamento. Entretanto, podem existir variáveis compondo o conjunto de dados sendo analisado que não sejam séries temporais. Essas variáveis podem ser usadas, por exemplo, para agrupar séries temporais, tendo potencial para ganho na diminuição da comple-

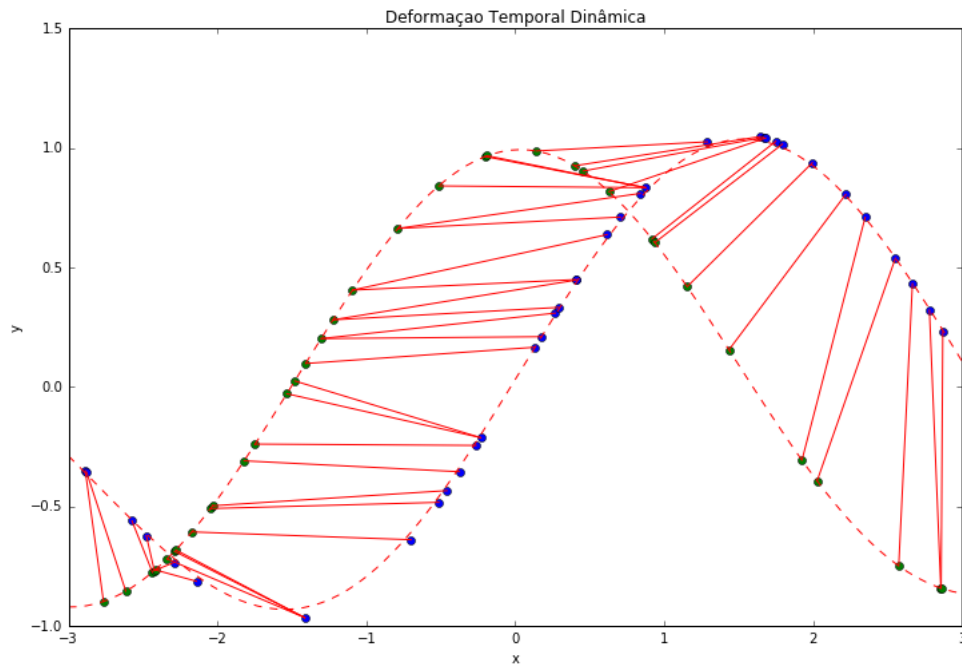


Figura 2.5: *Deformação Temporal Dinâmica. Pontos da mesma cor representam pontos da mesma série temporal (o tempo é representado pela coordenada x), delineados por linhas vermelhas pontilhadas. Neste caso os dados não são distorcidos por ruído. As linhas vermelhas sólidas ligam pontos pareados pela DTD.*

xidade computacional ou na performance de classificação. Para tratar essas variáveis, será utilizada a técnica de Árvore de Decisão, explicada a seguir.

2.7 Árvores de Decisão

Árvore de Decisão, AD[7], é uma técnica geral de aprendizagem formada por uma coleção de regras baseadas em variáveis do conjunto de dados. Neste trabalho, serão tratadas especificamente as Árvores de Classificação. As Árvores de Classificação montam uma árvore binária à partir de dados fornecidos, chamados de dados de treinamento. Cada nó da árvore corresponde a uma regra a ser aplicada para classificar. As folhas da árvore determinam qual classe um exemplo será classificado. Para classificar um novo exemplo em uma categoria, são aplicadas regras iniciando no nodo raiz e seguindo até encontrar uma folha, que determina a classificação do exemplo em questão.

Para construir uma árvore, regras baseadas nos valores das variáveis são selecionadas para obter a divisão que melhor categoriza o conjunto de dados considerado. Uma vez que uma regra é selecionada para compor a árvore e criar dois novos ramos, o mesmo processo é aplicado recursivamente em cada ramo descendente. A divisão para quando

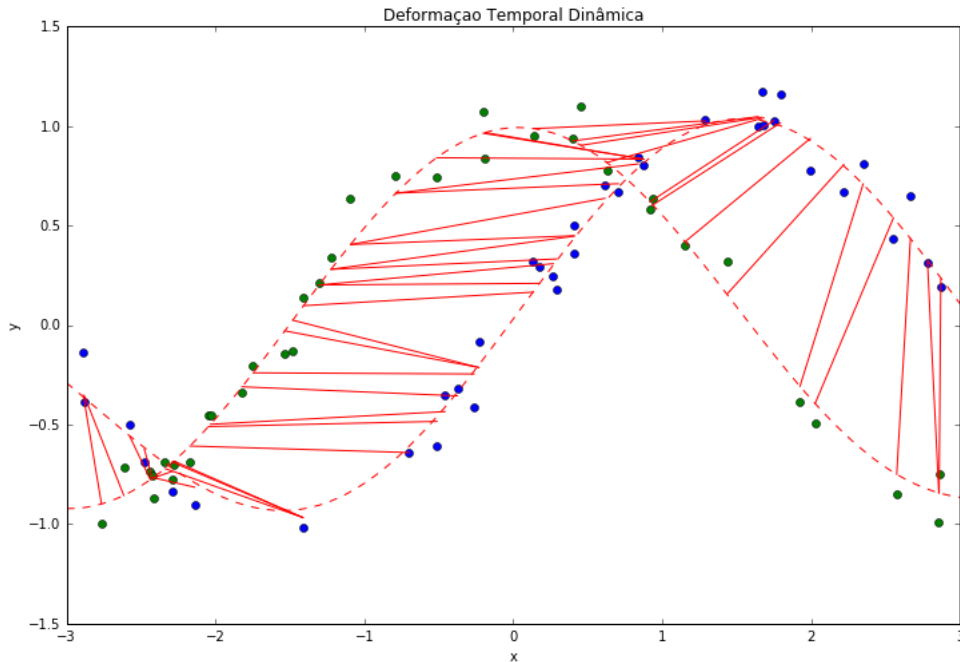


Figura 2.6: Pontos da mesma cor representam pontos da mesma série temporal (o tempo é representado pela coordenada x), corrompidos por ruído. As linhas vermelhas sólidas ligam pontos pareados pela DTD. Aqui, as linhas pontilhadas denotam a função média do modelo de PGDs.

é detectado que ganhos não podem mais serem obtidos ou de acordo com critérios de parada pré-estabelecidos. Cada ramo da árvore acaba em um nodo terminal.

2.8 Descrição dos dados

Os dados de UTI usados para o *PhysioNet Challenge 2012* foram extraídos do banco de dados clínicos MIMIC II, versão 2.6[37]. Foram selecionados 12.000 casos aleatoriamente de indivíduos cuja idade na ocasião da internação era maior do que 15 anos e cuja internação inicial na UTI foi de pelo menos 48 horas, não sendo aplicado qualquer outro critério de exclusão. Foram registradas um total de 41 variáveis, de forma que cada variável aparece em pelo menos um paciente ao menos uma vez ao longo das primeiras 48 horas depois da internação dos pacientes. Nem todas as variáveis estão disponíveis para todos os casos.

Há variáveis que são descritores gerais, como idade, gênero, altura, tipo de UTI e peso inicial. Conforme Silva et al.[40], as médias (e desvios padrão) da idade, altura, e peso são respectivamente 64.5(17.1) anos, 169.5(17.1) centímetros, e 81.2(23.8) kilogramas. Na variável gênero, 43.8% são do gênero feminino e 56.1% são do gênero masculino. Um maior número de pacientes foram internados na UTI médica (35.8%),

seguido pelas UTIs cirúrgica (28.4%), de recuperação de cirurgia cardíaca (21.1%), e coronária (21.1%). O resto das 36 variáveis são séries temporais, descritas na Tabela 2.1, para as quais múltiplas observações podem estar disponíveis. Cada observação tem um rótulo temporal indicando o tempo passado entre o instante da observação e o início da internação na UTI em horas e minutos.

Para o problema de classificação binária a avaliação é em função de positivos verdadeiros(PV), negativos falsos(NF) e positivos falsos(PF). A sensibilidade e a preditividade positiva são dadas por:

$$Se = \frac{PV}{PV + NF} \quad (2-30)$$

$$P^+ = \frac{PV}{PV + PF} \quad (2-31)$$

e é definido como *score* do problema 1 a menor destas medidas:

$$Score1 = \min(Se, P^+). \quad (2-32)$$

Para o problema de risco de morte, a avaliação é feita utilizando uma versão modificada da estatística Hosmer-Lemeshow[20], H . O cálculo da estatística H para uma entrada candidata ao desafio é feito primeiro ordenando o risco estimado calculado para todas as instâncias de um conjunto de *teste* (4000 para este desafio) e então particionando os registros correspondentes em decis designados por $g = 1, 2, 3, \dots, 10$. A estatística H e o *score* 2 são então calculados:

$$H = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{N_g \pi_g (1 - \pi_g) + 0.001} \quad (2-33)$$

$$Score2 = \frac{H}{\pi_{10} - \pi_1} \quad (2-34)$$

onde para cada decil g tem-se: o número observado de mortes no hospital O_g , o número predito de mortes E_g , o número de registros N_g (400 para este desafio), e a média do risco estimado do decil π_g . O score final é calculado normalizando a estatística H . O objetivo do problema 2 é minimizar o *score* 2.

Neste trabalho, apenas o escore 1 será abordado. Os resultados dos candidatos ao desafio são descritos em [40].

Tabela 2.1: Séries temporais do Challenge 2012.

Variáveis das séries temporais do desafio e percentagem de pacientes em que pelo menos uma medida foi coletada. Tabela disponível em [40] e adaptada para a nomenclatura em português.

Medida	%	Unidades
ABP(Pressão sanguínea arterial)		
Invasiva(diastólica, média, sistólica)	98.4	mmHg
Não-Invasiva(diastólica)	87.3	mmHg
Não-Invasiva(média)	87.2	mmHg
Não-Invasiva(sistólica)	87.6	mmHg
Albumina	40.5	g/dL
ALP(fosfatase alcalina)	42.4	IU/L
ALT(alanina transaminase)	43.4	IU/L
AST(aspartato transaminase)	43.4	IU/L
Bilirubina	43.4	mg/dL
BUN(nitrogênio uréico no sangue)	98.4	mg/dL
Colesterol	7.9	mg/dL
Creatinina	98.4	mg/dL
FiO2(fração inspirada de oxigênio)	67.6	[0-1]
GCS(score de Glasgow)	98.4	[3-15]
Glucose	97.5	mg/dL
HCO3(bicarbonato sérico)	98.2	mmol/L
HCT(hematócrito)	98.4	%
HR(batimento cardíaco)	98.4	bpm
K(potássio sérico)	97.9	mEq/L
Lactato	54.8	mmol/dL
Mg(magnésio sérico)	97.5	mmol/L
Mech. Vent.(ventilação mecânica)	63.1	[sim/não]
Na(sódio sérico)	98.2	mEq/L
PaCO2	75.4	mmHg
PaO2	75.4	mmHg
pH	75.9	[0-14]
Plaquetas	98.3	células/nL
Respiração	27.7	bpm
SaO2	44.7	%
Temperatura	98.4	Celsius
Troponina-I	4.7	ug/L
Troponina-T	21.9	ug/L
Urina	97.4	mL
WBC(contagem de cél. brancas)	98.2	células/nL
Weight(Peso)	67.7	kg

2.9 Considerações Finais

Neste capítulo foram descritos os fundamentos da abordagem proposta. Especial ênfase foi dada aos modelos de Processos Gaussianos e Processos Gaussianos Dependentes. O uso destes modelos é a principal diferença entre a abordagem proposta e outras abordagens do desafio. Este uso será descrito com mais detalhes no Capítulo [3](#).

Metodologia e Resultados

O objetivo deste trabalho visa melhorar a predição de mortalidade em UTIs. A predição proposta neste trabalho é embasada na criação de modelos computacionais preditivos à partir de dados observados. Os dados processados são compostos por dados de 4000 pacientes internados em UTIs. O processamento para realizar a predição foi dividido em etapas. A primeira etapa foi representar cada um dos registros como um modelo de PGs e obter a distância entre eles. A segunda etapa consistiu em usar uma rotina de AD para agrupar os registros. Na terceira etapa um algoritmo baseado em distância foi utilizado para obter a probabilidade de óbito de registros agrupados na etapa anterior. O valor dessa probabilidade é então combinado com a probabilidade fornecida pela mesma AD usada para agrupar os registros para fornecer a classificação final de cada registro. Para avaliar o desempenho foram utilizados os critérios do *Physionet Challenge 2012* e também a validação cruzada. Este capítulo descreve com mais detalhes as etapas citadas.

3.1 Representação de Processos Gaussianos Dependentes

Todos os registros foram primeiramente modelados como PGDs. Para representar os registros de séries temporais de pacientes foram consideradas 8 séries temporais, cujos identificadores são:

- HR
- SysABP
- Temp
- Urine
- HCO3
- GCS
- BUN
- WBC

Os valores de frequência de pulso (HR), pressão arterial sistólica invasiva (SysABP) e temperatura (Temp) foram escolhidos pelo uso clínico comum e uso em esco-

res como o SAPS[24]. Os valores de urina (Urine), bicarbonato sérico (HCO_3), escore de coma de Glasgow (GCS), nitrogênio uréico no sangue (BUN) e a contagem de células brancas (WBC) foram escolhidos por comporem características apontadas por Xia et al.[48] como tendo maior relação com a situação de morte no hospital. É importante observar que um dado registro de paciente pode não ter todos esses sinais disponíveis, e essa situação deve ser tratada para efeito de comparação da distância entre dois registros. Neste trabalho, quando dois registros são comparados e em pelo menos um deles os dados de uma série temporal estão faltando, é assumida como distância a média das outras distâncias (de outras séries temporais pareadas) presentes entre os dois registros.

O modelo de PGDs também requer a especificação do número de processos latentes e dos cerne que determinam as dependências entre processos latentes e saídas. Tais escolhas não são óbvias. Existem várias alternativas possíveis para determinar o número de processos latentes, incluindo o uso de Processo de Banquete Indiano como no trabalho de Guarnizo et al.[16]. Por simplicidade, neste trabalho foi escolhido usar como número de processos latentes o mesmo número de saídas (sinais) de um determinado registro de paciente, mas com máximo de 5 processos latentes. Essa escolha foi feita com a hipótese de que saídas compartilham processos latentes. O cerne escolhido todas as saídas e processos latentes foi o exponencial quadrático decrescente, por ter uso mais conhecido e comum. A função de covariância resultante é mostrada em Boyle et al.[5], Apêndice A.

A implementação do modelo de PGDs foi utilizando a linguagem de programação Python[45]. Esta implementação foi baseada na modificação de uma biblioteca pré-existente para PGs chamada PyGP[19]. Foram criadas rotinas para tratar PGDs com múltiplas saídas. A implementação permite definir modelos, treinar os parâmetros dos cerne com dados de treinamento e calcular a distribuição posterior nos pontos de interesse para realizar regressões.

Cálculos de funções exponenciais em computadores com aritmética finita podem sofrer com a transbordação (em inglês *overflow*) de valores representáveis pelo computador. Resultados significativamente errados podem ocorrer e é necessário o uso de técnicas para mitigar este efeito. Por essa razão, os instantes de tempo das leituras das séries temporais foram normalizados com a normalização min-max para o intervalo $[-1, 1]$. Pelo mesmo motivo, os valores correspondentes das séries temporais de cada sinal foram normalizados com escore padrão (ou *z-score*). Essa escolha é consistente com as suposições de um modelo de PGs. Como as comparações entre séries temporais foram feitas em termos de distribuições probabilísticas, os valores específicos utilizados na normalização não interferem no cálculo de distâncias e conseqüentemente nos resultados de classificação.

Com as séries temporais carregadas e definidos os números de saídas (sinais) e processos latentes do modelo de PGDependente, é preciso ajustar os parâmetros desse

modelo para realizar regressões (estimativas do valor da saída considerada). A abordagem adotada neste trabalho é adotar um valor pontual destes parâmetros que, aproximadamente, maximiza a função de verossimilhança logarítmica do modelo. Para isso, foi utilizada a rotina `f_min_l_bfgs_b` codificada na biblioteca de rotinas para computação científica `Scipy`[32], escrita na linguagem de programação `Python`.

A função de verossimilhança pode ser uma função multimodal (i.e., com múltiplos pontos extremos locais). Por essa razão, a rotina `f_min_l_bfgs_b` pode atribuir valores demasiadamente baixos ou altos para a variância associada a uma dada saída. Dessa forma, o valor dos parâmetros de variância foi limitado para o mínimo de 0.05 e máximo de 2.0 após a otimização.

Na rotina de otimização de parâmetros `f_min_l_bfgs_b` podem surgir imprecisões numéricas inerentes à aritmética de computadores com precisão finita. Essas imprecisões causam falhas em cálculos que compõem a função de verossimilhança de um PG, especialmente o cálculo da decomposição Choleski de uma matrix de covariância. Por esse motivo, foi utilizada a estratégia de inicialização aleatória dos parâmetros. Caso haja falha na rotina, é realizada outra inicialização aleatória até que haja sucesso na rotina ou que tenham sido realizadas 20 tentativas. Nos eventos em que todas as 20 tentativas falharam, foi assumido o estado de parâmetros imediatamente anterior ao último estado inválido. Esta estratégia produziu ajustes satisfatórios para os parâmetros.

As Figuras seguintes de 3.1 a 3.8 mostram as distribuições posteriores correspondentes às séries temporais de um paciente obtidas com o processo descrito acima. Os pontos em azul ligados em sequência são os pontos da série temporal original, a linha verde corresponde à média da distribuição em cada instante de tempo e a área cinza delimita o intervalo de 95% de confiança. As Figuras ilustram a capacidade do modelo de estabelecer distribuições probabilísticas em cada instante de tempo, fornecendo intervalos de incerteza. As curvas observadas denotam tendências gerais de variação, que podem ser usadas para calcular a distância entre duas séries temporais de forma mais robusta quanto à leituras errôneas.

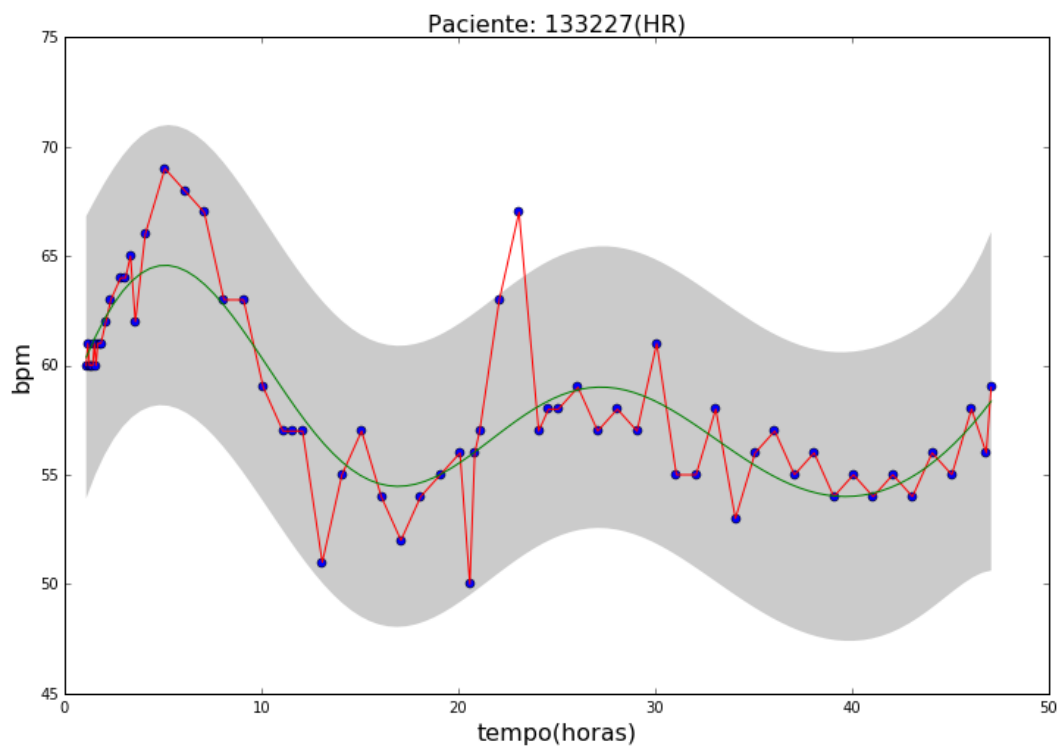


Figura 3.1: Distribuição de probabilidade posterior correspondente à série temporal da frequência cardíaca de um paciente.

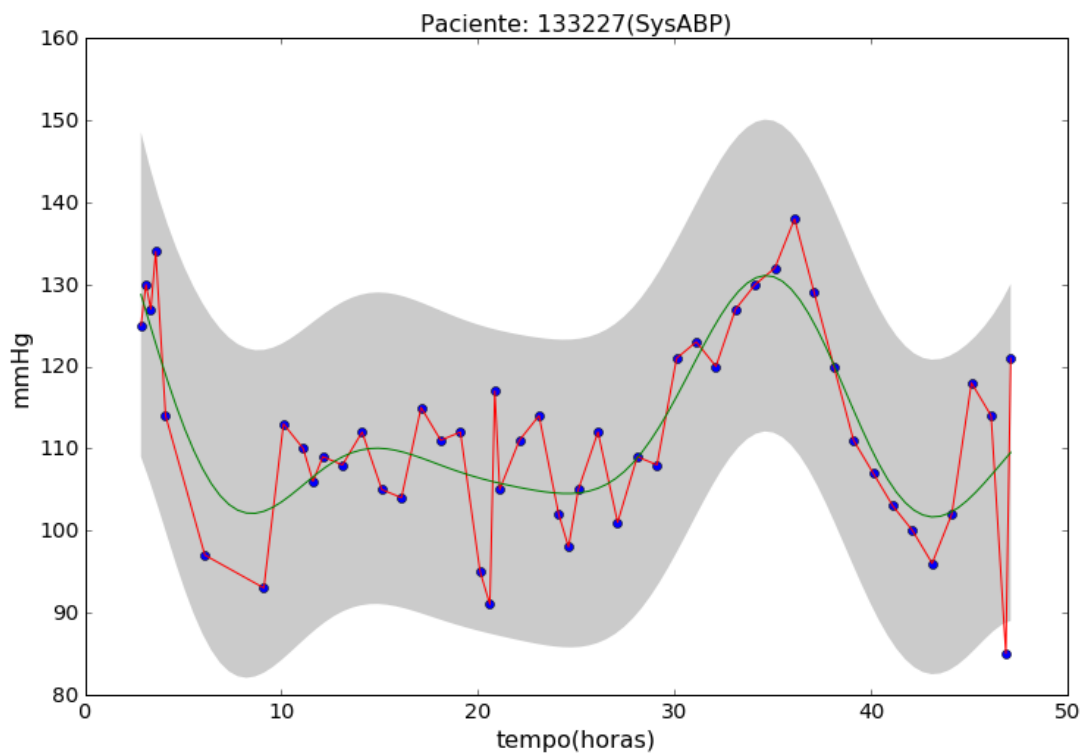


Figura 3.2: Distribuição de probabilidade posterior correspondente à série temporal da pressão arterial sistólica de um paciente.

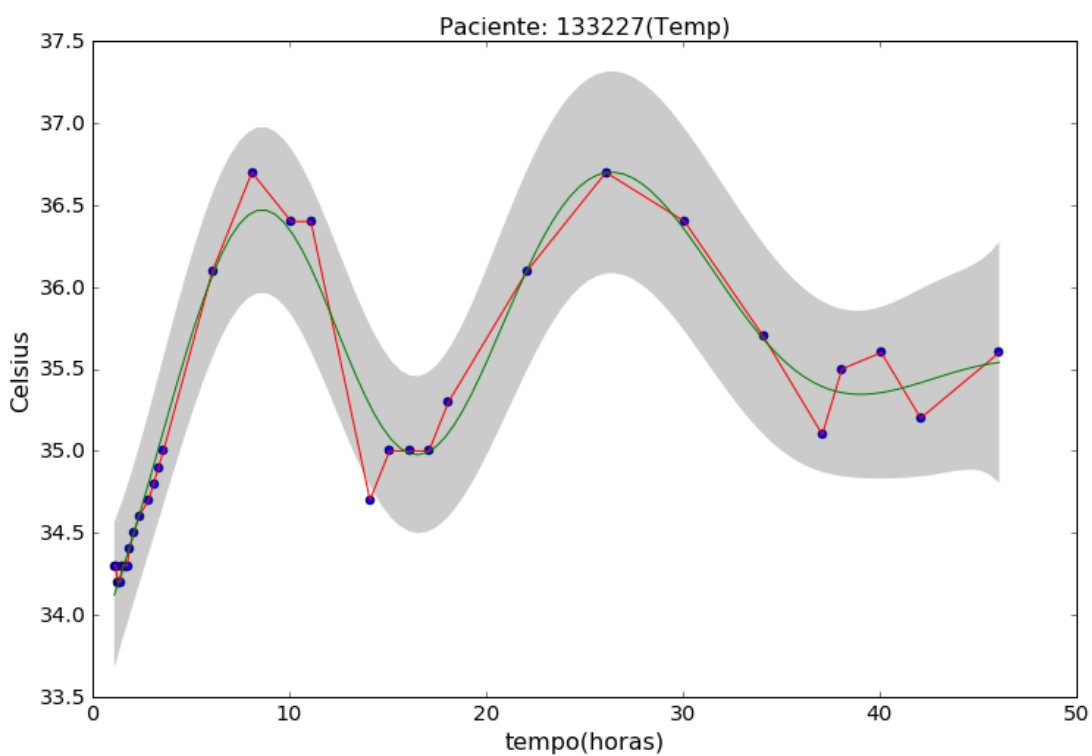


Figura 3.3: Distribuição de probabilidade posterior correspondente à série temporal da temperatura de um paciente.

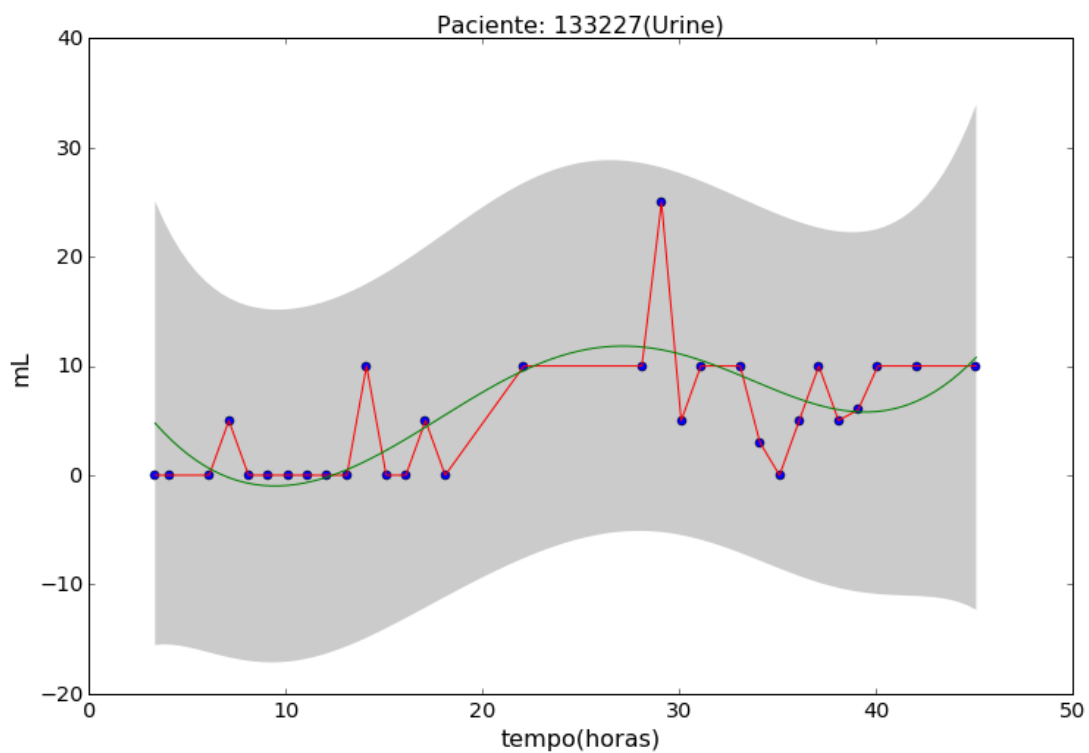


Figura 3.4: Distribuição de probabilidade posterior correspondente à série temporal do volume de urina de um paciente.

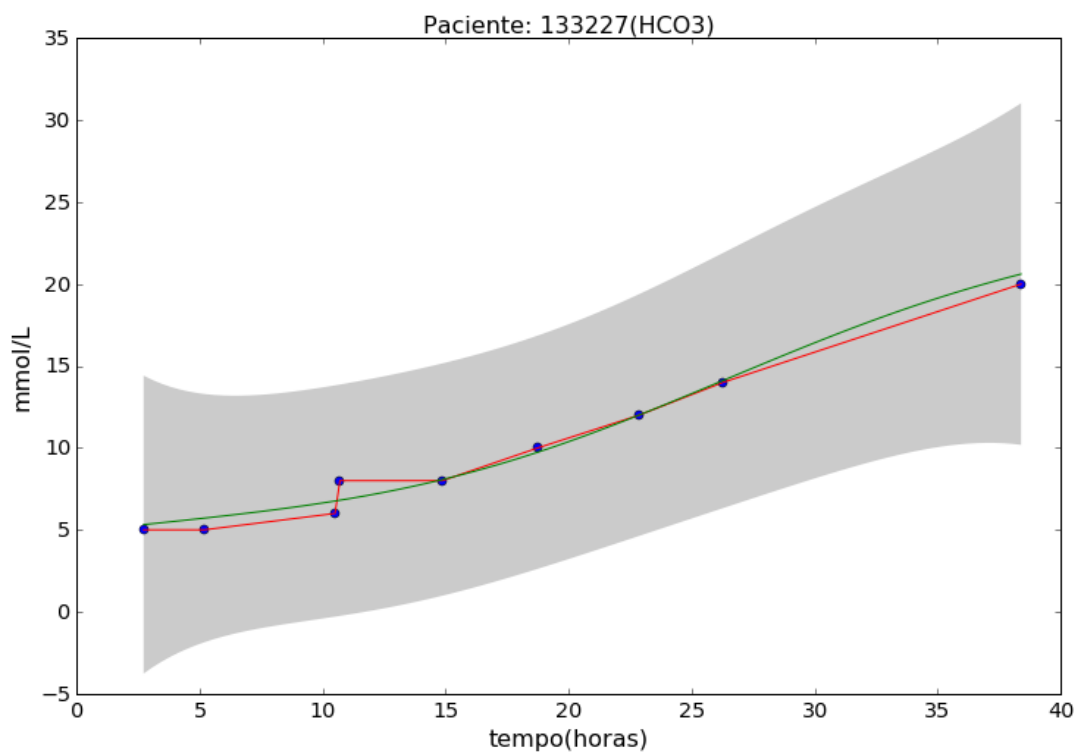


Figura 3.5: Distribuição de probabilidade posterior correspondente à série temporal do bicarbonato sérico de um paciente.

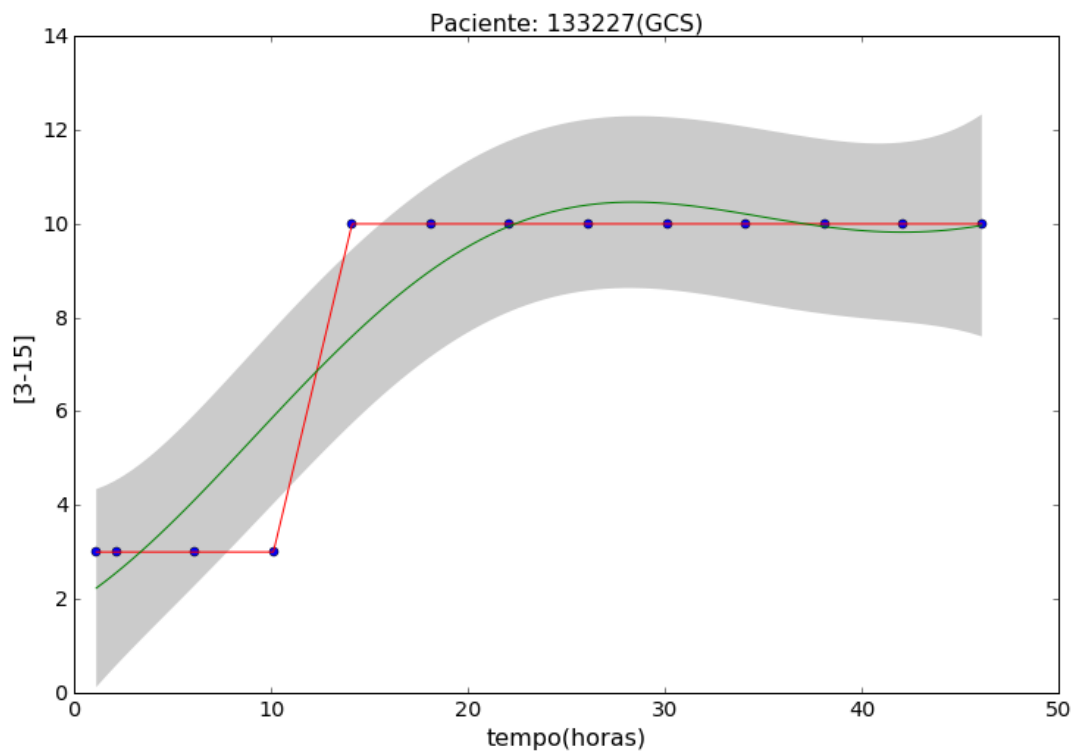


Figura 3.6: Distribuição de probabilidade posterior correspondente à série temporal da escala de Glasgow de um paciente.

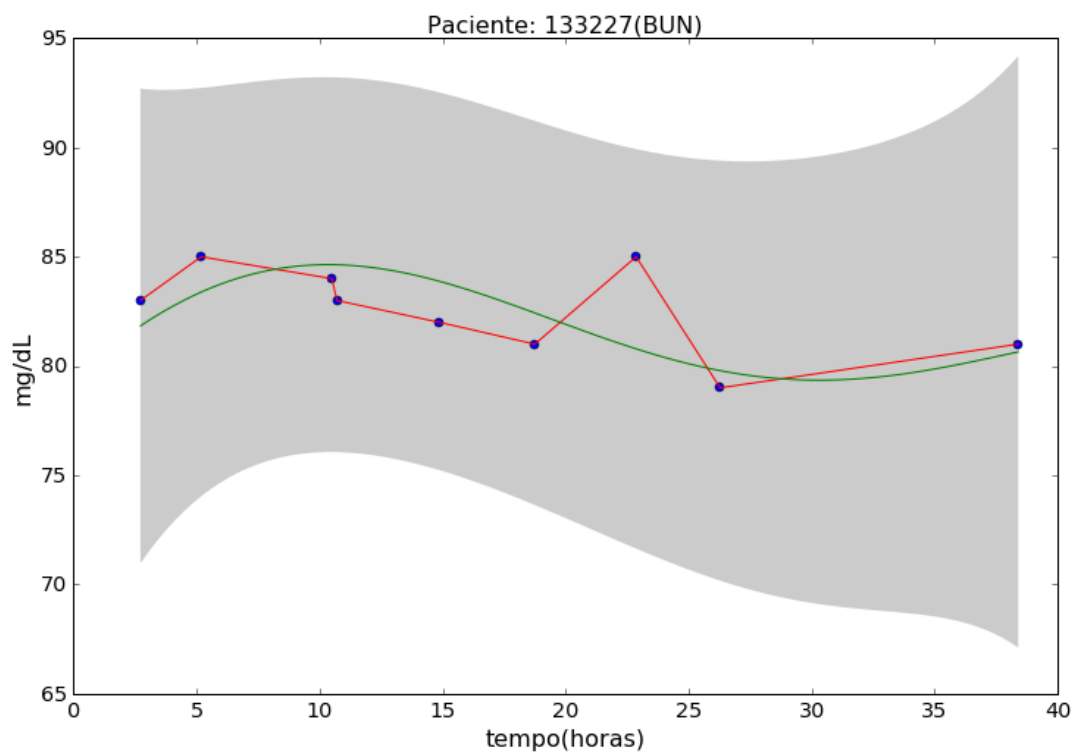


Figura 3.7: Distribuição de probabilidade posterior correspondente à série temporal do nitrogênio uréico no sangue de um paciente.

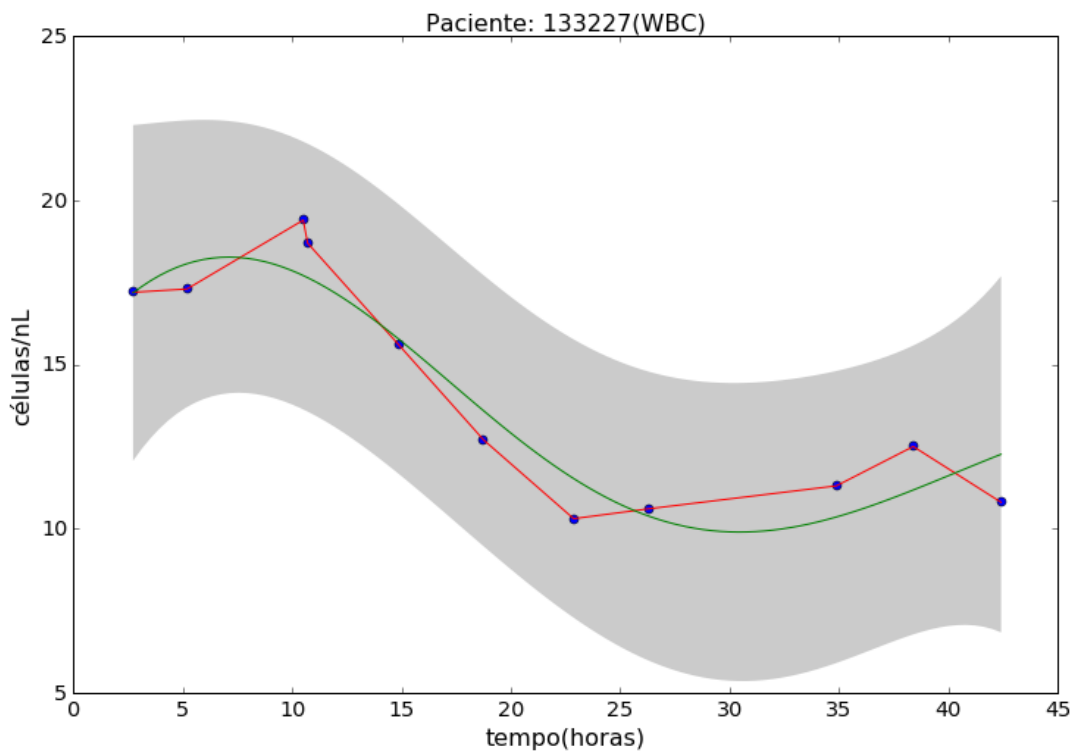


Figura 3.8: Distribuição de probabilidade posterior correspondente à série temporal da contagem de glóbulos brancos.

3.2 O cálculo da distância entre Processos Gaussianos Dependentes

Para calcular a distância entre dois PGDs, foi utilizada a média das distâncias entre cada série temporal considerada pareada (pelo tipo de sinal, e.g., HR, GCS) entre os dois processos. A distância entre séries pareadas (i.e, correspondentes ao mesmo sinal) é calculada utilizando combinando a DTD e as divergências-KL (eq. 2-24) entre os pontos de ambas as séries, de maneira semelhante a Sun e Zhao[42]. Este cálculo é feito da seguinte forma:

1. Verifica-se a ausência de valores. Se os valores de séries temporais estiverem ausentes, a distância ficará marcada com um valor negativo para ser substituída pela média de outras distâncias válidas posteriormente.
2. Para obter a distância entre as séries, usa-se um algoritmo de Deformação Temporal Dinâmica denominado *FastDTW*[38] com a implementação em linguagem Python disponibilizada em [43]. A distância entre dois pontos é obtida pela Divergência-KL das distribuições posteriores correspondentes a cada ponto, obtidas pelo modelo de PGDs.
3. Para que uma série que tenha mais pontos não produza distâncias maiores, é considerada a média da distância entre os pontos
4. As distâncias inválidas (representadas por números negativos) são então substituídas pela média das distâncias válidas.

Ao fim desta etapa, pode-se obter o valor da distância entre registros quaisquer. Esses valores serão utilizados na etapa de classificação baseada em distância.

3.3 Agrupamento com Árvore de Decisão

Uma limitação da representação de registros de pacientes como PGs é que parte importante dos dados presentes nos registros de pacientes não são séries temporais. Características como idade, peso e altura podem influenciar nos resultados quando se compara variáveis fisiológicas entre 2 pacientes. Preferencialmente, deve-se comparar variáveis fisiológicas de séries temporais entre pacientes que tenham idade e outras características parecidas. Neste trabalho, foi adotada a estratégia de agrupar pacientes com características similares (que não sejam séries temporais) para só então comparar séries temporais de pacientes dentro de um mesmo grupo.

Para realizar este agrupamento, foi utilizada a rotina de AD da biblioteca Scipy. Foram utilizados as seguintes características, extraídas de cada registro de paciente:

- Age
- ICUType
- Weight
- Height
- Gender
- BMI
- LastGCS

Os valores de idade (Age), tipo de UTI (ICUType), peso (Weight), altura (Height) e gênero (Gender) compõe as características fornecidas de cada paciente nos dados considerados. Foram adicionados os valores de índice de massa corporal (BMI) por sua importância clínica e última leitura do escore de Glasgow (LastGCS) por ser uma característica que foi apontada no trabalho de Xia et al.[48] como a mais relevante para previsão de óbito.

A Figura 3.9 mostra uma árvore de decisão gerada para agrupar e classificar os registros. Neste trabalho, as árvores de decisão foram limitadas a terem no máximo 10 folhas. As folhas de uma Árvore de Decisão tendem a ter dados com características mais parecidas. Dessa forma, a ideia principal nesta estratégia é formar grupos mais homogêneos de pacientes, fornecidos pela Árvore de Decisão. Isso pode ser constatado nos valores de probabilidade das classes em cada folha. Nota-se que em algumas folhas, o valor da probabilidade de óbito é próximo de 0.5, entretanto na maioria das folhas esse valor é mais próximo de 0.0 ou de 1.0.

As ADs dessa etapa podem ser usadas para agrupar registros ou para fornecer um valor de probabilidade para combinar à outros classificadores ou ser utilizado isoladamente com um limite para classificação binária. Na abordagem proposta, é utilizado tanto o agrupamento quanto o valor de probabilidade para combinação com outros classificadores.

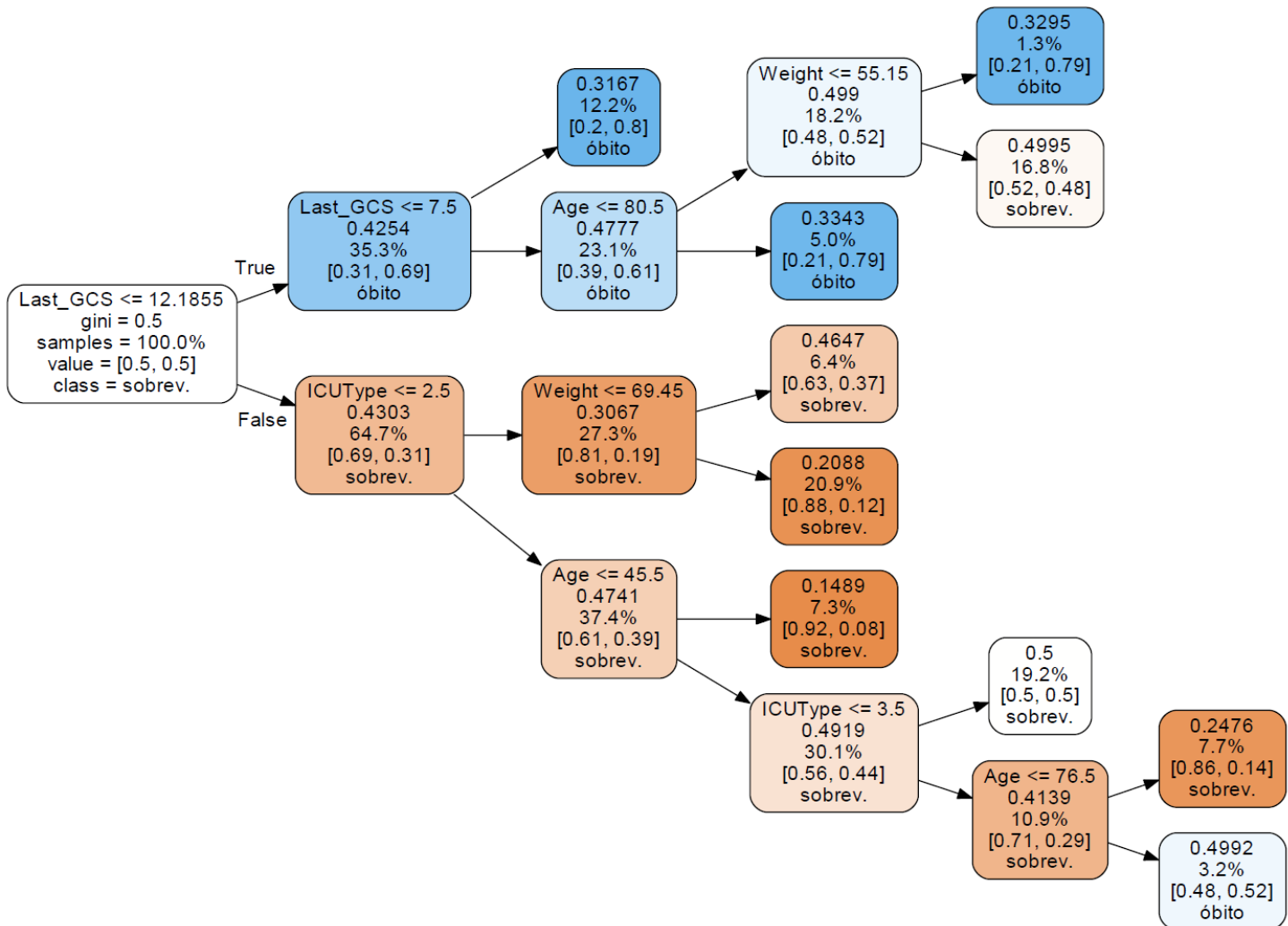


Figura 3.9: *Árvore de decisão. Em cada nó estão indicados (de cima para baixo) qual atributo divide o nó, o índice de impureza, a proporção de amostras, a probabilidade das classes e a classe mais provável.*

3.4 Classificação baseada em distância

A classificação baseada em distância foi feita com um algoritmo baseado na técnica K-Vizinhos Mais Próximos (em inglês, *K-Nearest Neighbors*, KNN)[1], da seguinte forma:

1. Para cada registro a ser classificado é obtido qual agrupamento (folha da árvore de decisão) ele pertence
2. Dentro do agrupamento que o registro a ser classificado foi incluído, ele é classificado combinando o algoritmo KNN com a probabilidade atribuída àquela folha pelo algoritmo de árvore de decisão.

A rotina utilizada é mostrada no algoritmo 3.1. Observa-se a utilização de peso para mitigar o desbalanceamento entre as classes. O algoritmo retorna a probabilidade

do paciente ser da classe de óbito. Essa probabilidade pode ser usada na combinação com outro classificador (abordagem proposta) ou pode ser utilizada isoladamente com um limite para classificação binária.

Algoritmo 3.1: $KNN(pg, Grupo_PGs, classes)$

Entrada: pg : processo gaussiano de teste pg ,
 $Grupo_PGs$: lista de procs. gaussianos,
 $classes$: classes dos proc. da lista $classes$.

Saída: $prob$: probabilidade de óbito

```

1  $n \leftarrow$  número de elementos em  $Grupo\_PGs$  :.
2  $k \leftarrow \lceil \sqrt{n} \rceil$ 
3  $distancias \leftarrow$  lista de distâncias entre  $pg$  e cada proc. em  $Grupo\_PGs$ 
4 substitua valores inválidos pela média das distâncias válidas em  $distancias$ 
5  $k\_vizinhos \leftarrow$  lista dos  $k$  procs. mais próximos na lista  $distancias$ 
6  $peso1 \leftarrow 1 / (\text{proporção de óbitos em } Grupo\_PGs)$ 
7  $peso0 \leftarrow 1 / (1 - peso1)$ 
8  $pesos = peso1$  se  $classe = \text{óbito}$ , senão  $peso0$  para cada elemento em  $k\_vizinhos$ 
9  $prob \leftarrow$  média de  $classes$  correspondentes à  $k\_vizinhos$  ponderada por  $pesos$ 

```

3.5 Combinação e classificação

A combinação do algoritmo 3.1 com a Árvore de Decisão é feita por seriação da seguinte forma:

1. O valor de probabilidade predito é calculado por:

$$predição = \alpha(pred_{DT} + pred_{KNN}) \quad (3-1)$$

onde valor $pred_{DT}$ é o valor predito pela árvore de decisão e o valor $pred_{KNN}$ é o valor predito pelo algoritmo KNN.

2. Se o valor $predição$ for maior do que 0.5, o registro é classificado como óbituário. Caso contrário, é classificado como sobrevivente.

O valor α é uma constante determinada empiricamente. Os melhores resultados verificados foram obtidos com o valor sendo aproximadamente 0.415. O fato deste valor não ser exatamente 0.5 (que coincidiria com a média aritmética entre as predições) reflete a falta de balanceamento entre as classes e o efeito causado em falsos positivos de ambos

Tabela 3.1: Resultados para o escore 1

Abordagem	Média	Desvio Padrão
1. Árvore de Decisão(AD)	0.2778	0.0352
2. KNN(agrup.)	0.1923	0.0203
3. KNN	0.2331	0.0197
4. KNN(agrup.)+AD	0.3796	0.0372
5. KNN+AD	0.3796	0.0348

os algoritmos que foram combinados. Devido ao desbalanceamento das classes, casos limítrofes tendem a ter maior taxa de acerto na classificação quando são classificados como sobreviventes do que baseados estritamente em um limite (neste caso, 0.5).

3.6 Resultados

De forma semelhante à Johnson et al.[21], foram feitas 32 repetições de montagem de modelos de árvores de decisão para agrupamento e classificação, com permutação aleatória de registros para cada repetição. Para cada repetição, foi realizada a validação cruzada com 4 partições, sendo uma partição para teste e as demais para treinamento de maneira sucessiva. Para avaliar o desempenho de classificação, nenhum registro de teste foi utilizado simultaneamente em treinamento. O conjunto de treinamento foi utilizado para montar uma árvore de decisão. Para cada registro de teste foram feitas classificações com diferentes abordagens, sendo todas variações da abordagem descrita até aqui:

1. Apenas árvore de decisão
2. KNN em registros agrupados por árvore de decisão
3. Apenas KNN em todos os registros de treinamento
4. KNN em registros agrupados por árvore de decisão combinada com classificação por árvore de decisão(abordagem proposta e descrita neste capítulo)
5. KNN em todos os registros de treinamento combinada com classificação por árvore de decisão

A Tabela 3.1 mostra os resultados das diferentes abordagens.

3.7 Considerações Finais

A abordagem 4 combinando Árvores de Decisão e KNN por serialização obteve média do escore 1 do desafio *Physionet Challenge 2012* de 0.38 e desvio padrão de 0.04. Utilizando apenas a árvore de decisão gerada para agrupar os PGs (abordagem 1), o escore 1 teve média de 0.28 e desvio padrão de 0.04. Utilizando apenas a predição do algoritmo

KNN aplicado aos PGs agrupados pela árvore de decisão (abordagem 2), a média do score 1 foi 0.19 com desvio padrão de 0.02. Estes resultados confirmam que nesse caso (e em geral) a combinação das duas predições tem um desempenho de classificação melhor do que a utilização isolada. Dessa forma, a combinação dos classificadores de árvore de decisão e KNN com agrupamento é vantajosa. Este resultado indica que há capacidade de discriminação na abordagem proposta para o problema considerado, mostrando potencial para uso combinado a outros classificadores.

Observa-se que as abordagens que tiveram maior escore 1 são as abordagens 4 e 5. Estas abordagens tem desempenho do escore 1 bastante parecido. Entretanto, na abordagem 4 o número de registros considerados para o algoritmo KNN é significativamente reduzido por conta do agrupamento realizado pela árvore de decisão. Dessa forma, existe vantagem computacional em utilizar a abordagem com agrupamento (4) em vez da abordagem sem agrupamento (5). Isso se deve ao fato de que o algoritmo KNN precisa realizar menos comparações, pois o conjunto considerado para comparação de distâncias é apenas um grupo e não o conjunto inteiro de treinamento. Entretanto, é importante observar que há um aumento no desvio-padrão utilizando a abordagem 4 em comparação com a abordagem 5.

O desempenho da abordagem proposta supera um método já utilizado em UTI, o SAPS-I, cujo resultado teve média de 0.31 no mesmo conjunto de dados. O resultado deste método e de outros trabalhos é discutido no Cap. 4.

Trabalho Relacionados

Na literatura existe um número crescente de publicações envolvendo o tema da mineração de dados e aprendizagem de máquina aplicados na área da saúde com a descrição de variadas aplicações. Exemplos de tarefas médicas onde a mineração de dados é aplicada são o diagnóstico, o suporte à tomada de decisões, a predição(ou prognóstico), detecção de anomalias e monitoramento. Estas aplicações da mineração de dados nas tarefas médicas são descritas em artigos de Esfandiari et al.[12] e Banee et al.[2]. Neste capítulo, serão descritos trabalhos mais especificamente relacionados com o presente trabalho.

4.1 Trabalhos relacionados ao *PhysioNet Challenge 2012*

Especificamente sobre o desafio *PhysioNet Challenge 2012*, houveram várias publicações do resultado de variados métodos aplicados ao problema. Foram utilizados três conjuntos de dados de 4.000 registros cada para o desafio, totalizando 12.000 registros de pacientes. O conjunto de dados A foi fornecido com registros com séries temporais e outros dados de pacientes juntamente com um arquivo contendo o resultado(sobrevivência ou óbito) para cada paciente. O conjunto B foi fornecido com dados semelhantes ao conjunto A, mas sem resultados de sobreviventes. Finalmente o conjunto C foi usado para avaliar os participantes e não foi fornecido para os participantes.

Johnson et al.[21] aplicaram uma abordagem de combinação de classificadores bayesianos ao problema. O modelo consiste em uma soma de árvores aditivas regressivas bayesianas, e a inferência é feita por um algoritmo denominado *Backfitting Markov Chain Monte Carlo*. O modelo tem muitas vantagens, como tratamento de valores ausentes e utilização de atributos facilmente obtíveis a partir parâmetros clínicos regularmente coletados. Esta abordagem obteve a maior pontuação para o primeiro problema, a classificação de pacientes, e a quarta colocação na estimativa de risco de mortalidade.

Citi e Barbieri[9] usaram descritores gerais e variáveis agregadas dos dados como atributos de classificadores do tipo máquina de vetores de suporte quadráticas. Um modelo linear generalizado foi usado para prever a probabilidade de risco de morte

usando as saídas de seis máquinas de vetores de suporte usadas para a classificação. Esta abordagem obteve a segunda maior pontuação para o primeiro problema, a classificação de pacientes, e a primeira colocação na estimativa de risco de mortalidade.

Vairavan et al.[44] propõem um algoritmo baseado em regressão logística e modelos ocultos de Markov. A ideia básica dos modelos ocultos de Markov é a estimação de estados ocultos de um processo usando variáveis observáveis ao longo do tempo. Esta abordagem obteve a terceira maior pontuação para o primeiro problema.

Xia et al.[48] usaram redes neurais para tratar o problema. As redes neurais identificaram 26 características que melhor predizem o óbito de um paciente.

McMillan et al.[28] abordaram o problema como uma procura por padrões de séries temporais. Esta abordagem é baseada em fazer previsões de acordo com a ocorrência consistente de padrões em séries temporais entre os registros. As séries temporais são subdivididas e para cada subdivisão é atribuído um valor. Este valor pode ser baixo, médio ou alto. Dessa forma, cada série temporal é representada por uma sequência de valores (baixo, médio ou alto) que pode ser analisada e pareada com padrões associados com o resultado (óbito ou sobrevivência).

Xu et al.[49] usaram análise de agrupamento de dados de séries temporais multi-granulares. As características usadas foram o mínimo e o máximo de 16 segmentos de séries temporais do período disponível (48 horas). Para realizar a análise, 10 variáveis foram selecionadas, totalizando 20 características por paciente.

Lee et al.[25] usaram regressão logística. Foram geradas características a partir dos dados fornecidos combinados com dados demográficos. Entre as características extraídas, estão as últimas leituras para cada série temporal. Para tratar dados faltantes, foi utilizada a imputação de valores médios.

Krajnak et al.[23] usaram análise clínica de especialistas para melhorar um sistema baseado em regras difusas (*Fuzzy Rule Based System*). Um otimizador baseado em algoritmos genéticos foi usado para gerar os coeficientes para a solução final.

Severeyn et al.[39] usaram um classificador baseado em Análise de Correspondência Simples (*Simple Correspondence Analysis*). Este método é baseado na combinação de dados clínicos e laboratoriais com escores mais tradicionais como o APACHE-II e SAPS-II.

Macaš et al.[26] usaram classificador Bayesiano linear simples, selecionando características com um otimizador baseado em Teoria de Impacto Social. Foram usadas características baseadas em escores já estabelecidos assim como variáveis dos pacientes, totalizando 935 características.

Marco et al.[27] usaram regressão logística para classificar os registros. Para cada variável foi considerada como característica a média da verossimilhança de fazer parte do grupo de sobreviventes. Para selecionar as características usadas foi utilizado um critério

de seleção sequencial.

Bosnjak e Montilla[4] usaram máquinas de vetores de suporte combinadas a estatísticas (média, desvio padrão e intervalo de variação) de cada variável fisiológica considerada. Inicialmente foram consideradas as mesmas variáveis utilizadas no escore SAPS-I e posteriormente foram adicionadas mais variáveis pela recomendação de médicos para verificar melhora no desempenho.

Pollard et al.[34] exploraram a estabilidade de medidas tais como frequência cardíaca e pressão sanguínea com um método usado para estudar frequência e intensidade de erupções solares. A abordagem incorpora uma rede neural artificial treinada com características extraídas dos dados dos pacientes.

Bera e Nayak[3] usaram regressão logística com características incluindo média, mínimo e máximo de 30 variáveis. Ao todo 88 características foram extraídas, selecionadas conforme disponíveis nos dados dos pacientes.

Hamilton et al.[17] usaram regressão logística com características incluindo primeiro valor, último valor, média, mínimo e máximo de 37 séries temporais. A significância de cada característica foi determinada por regressão linear. Dados faltantes foram completados com a média dependendo se o resultado foi óbito ou sobrevivência.

Yi et al.[50] usaram análise de histograma das variáveis médicas em um modelo de aprendizado *Adaboost* em cascata. Nesta proposta, as séries temporais são interpoladas para que tenham uma leitura por minuto. Em seguida, o valor médio é obtido a cada 60 minutos para cada paciente, obtendo um vetor de características em 1776 dimensões. Uma métrica de diferenciação foi elaborada baseada em histogramas de características por hora. A classificação é então feita usando-se essa métrica de diferenciação com o modelo *Adaboost* em cascata.

A Tabela 4.1 resume os resultados obtidos pelos participantes para os dados do conjunto A. Participantes cujos resultados para o conjunto A não foram disponibilizados publicamente não foram incluídos na tabela. Os resultados usados pela organização do evento para eleger as melhores abordagens foram obtidos com o conjunto de dados C, para o qual não foram fornecidos descritores de resultado (sobrevivência ou óbito) para a comunidade. Dessa forma, os resultados para o conjunto A são mais importantes para o presente trabalho.

4.2 Processos Gaussianos Multi-saída

Este trabalho propõe uma abordagem baseada em Processos Gaussianos Dependentes, que são um tipo de Processo Gaussiano Multi-saída. Existem algumas possíveis abordagens multi-saída na literatura, que são citadas aqui.

Tabela 4.1: Resultados do Challenge 2012

Autores	Abordagem	Escore(evento 1)
Hamilton et al.	regressão logística	0.57
Yi et al.	análise de histograma e Adaboost	0.56
Marco et al.	regressão logística	0.55
Variavan et al.	regressão logística	0.52
Xia et al.	redes neurais artificiais	0.50
Johnson et al.	combinação de classif. bayesianos	0.48
Lee et al.	regressão logística	0.47
Macaš et al.	classificador Bayesiano linear	0.47
Severeyn et al.	análise de correspondência simples	0.44
Bera et al.	regressão logística	0.44
Krajnak et al.	regras difusas com análise clínica	0.39
Silva et al.	SAPS-I	0.31
Xu et al.	agrupamento multi-granular	0.23

Nguyen e Bonilla[29] introduziram a abordagem de Processos Gaussianos Multi-saída Colaborativos para aprendizagem de tarefas dependentes em grandes conjuntos de dados. No contexto deste trabalho, uma tarefa é uma série temporal. A abordagem modela correlação entre tarefas pela combinação linear de processos esparsos e pelo compartilhamento de múltiplos conjuntos de pontos indutivos. Isso facilita a aplicação de inferência que pode escalar a um grande número de observações. Uma possível desvantagem é que, conforme mencionado, a modelagem entre processos é feita como uma combinação linear, que pode não capturar dependências mais complexas.

Durichen et al.[11] utilizaram o modelo de Processos Gaussianos Multi-Tarefa(em inglês *Multi-Task Gaussian Process* - MTGP) para modelar séries temporais fisiológicas. O modelo MTGP foi aplicado à conjuntos de dados sintéticos, monitoramento e radioterapia. Os autores reportaram ter obtido resultados melhores do que modelos que desconsideram dependência entre séries temporais. Assim como o modelo de Processos Gaussianos Dependentes, o modelo MTGP também realiza a convolução de cernes. Entretanto, pode ser desafiador estabelecer funções de covariância que sejam positivas definidas para esse modelo, que é um requisito para funções de covariância de Processos Gaussianos. Em comparação, para os Processos Gaussianos Dependentes basta que o cerne tenha valor absoluto integrável.

Guarnizo et al.[16] estende significativamente a formulação originalmente proposta por Boyle e Frean para os Processos Gaussianos Dependentes. Os autores adotaram uma formulação mais genérica para esse tipo de processo, passando a denominá-los Processos Gaussianos Convolutos Multi-saída. Além disso, os autores propuseram o uso do Processo de Banquete Indiano para escolha do número de processos Gaussianos latentes. Os autores aplicaram essa formulação para estudar modelos de forças latentes. Por ter

uma formulação mais simples e interpretável, no presente trabalho escolheu-se o modelo dos Processos Gaussianos Dependentes como formulado por Boyle e Freaan[6] e descrito no capítulo de Fundamentação Teórica(Cap. 2).

4.3 Considerações Finais

A principal diferença entre a abordagem proposta e todos os outros trabalhos apresentados no desafio *PhysioNet Challenge 2012* é que nenhum outro trabalho modela as séries temporais como Processo Gaussianos. Estabelecer distribuições probabilísticas sobre séries temporais pode oferecer possibilidades de análise úteis para avaliação de risco de morte. Uma possibilidade é a análise de trajetórias médias de pacientes agrupados. Essa análise pode fornecer caracterização e visualização de tendências gerais em grupos de pacientes, como no trabalho de Pimentel et al.[33].

O desempenho do escore 1 da abordagem proposta(0.38) é intermediário entre os escores mais altos e mais baixos entre os trabalhos submetidos ao desafio. Isto indica o potencial da abordagem, considerando a identificação de possibilidades de melhoria, discutidas no Cap. 5.

Conclusões

Este trabalho tem como principal contribuição uma abordagem computacional de análise de variáveis fisiológicas para previsão de óbito de pacientes em UTI. Esta abordagem utilizou Processos Gaussianos Dependentes para representar séries temporais correspondentes à dados de pacientes em UTI. Para a previsão de óbito, foi definida uma medida de distância entre Processos Gaussianos. A medida de distância proposta é baseada na divergência de distribuições probabilísticas entre pontos de séries temporais pareadas pela Deformação Temporal Dinâmica. A classificação foi feita combinando classificação por Árvores de Decisão com o algoritmo de classificação baseada em distância em registros agrupados pela Árvore de Decisão. Esta abordagem obteve valor 0.38 para o escore 1 do desafio *Challenge 2012* do sítio *Physionet*. Este resultado indica que há capacidade de discriminação na abordagem proposta para o problema considerado, mostrando potencial para uso combinado a outros classificadores. Resultados utilizando técnicas isoladamente (sem combinação de classificadores) ou sem agrupamento indicam vantagem em utilizar a abordagem com agrupamento em vez da abordagem sem agrupamento para o caso do algoritmo de classificação baseada em distância.

Outra contribuição deste trabalho é a extensão da biblioteca PyGP para tratar PGDs.

5.1 Limitações

O autor deste trabalho reconhece como limitações :

- Os critérios de seleção de variáveis fisiológicas neste trabalho foram elaborados por constatações de outros trabalhos na literatura. Esta limitação se aplica tanto aos critérios usados para seleção das séries temporais usadas na modelagem dos Processos Gaussianos tanto quanto na seleção de atributos para as árvores de decisão e das séries temporais usadas no cálculo da distância.
- Alternativas para cálculo da distância entre Processos Gaussianos foram pouco exploradas, sendo escolhidas alternativas utilizadas em outros trabalhos, como por

exemplo a Deformação Temporal Dinâmica. Também, alternativas para agrupamento dos Processos Gaussianos foram pouco exploradas.

- O número de processos latentes no modelo de Processos Gaussianos Dependentes não é uma escolha óbvia, e foi determinado de maneira simplificada. Da mesma forma, os cernes.
- A determinação de constantes usadas na combinação de classificadores foi feita de maneira empírica e não sistemática.

5.2 Trabalhos futuros

O autor deste trabalho indica como possíveis trabalhos futuros:

- Adotar ou desenvolver critérios sistemáticos de seleção de variáveis fisiológicas possivelmente mais apropriados para a abordagem proposta e considerar variáveis que não são séries temporais no cálculo da distância.
- Explorar trajetórias médias entre Processos Gaussianos classificados na mesma categoria e próximos de acordo com a medida de distância desenvolvida. Este estudo pode evidenciar tendências gerais nas séries temporais de pacientes com risco de óbito.
- Explorar relações entre parâmetros do modelo de Processos Gaussianos Dependentes. Este estudo pode evidenciar possíveis dependências entre séries temporais fisiológicas de um paciente.
- Melhorar o cálculo da distância entre Processos Gaussianos. Uma escolha diferente dos pontos para pareamento e cálculo da Divergência-KL com Deformação Temporal Dinâmica pode ser proposta e resultados melhores podem ser obtidos.
- Melhorar o agrupamento de Processos Gaussianos. Um método de agrupamento diferente da Árvore de Decisão pode apresentar melhores resultados.
- Melhorar os critérios para definir o número de processos latentes no modelo de Processos Gaussianos Dependentes. Escolher um número possivelmente mais apropriado de processos dependentes pode levar à melhora no desempenho.
- Melhorar os critérios para definir a constante usada na combinação por seriação das probabilidades resultantes da árvore de decisão e do algoritmo KNN. A definição desta constante baseada em estimativas de predição pode melhorar o resultado de classificação.

5.3 Considerações Finais

Pelo que foi considerado no capítulo introdutório, na apresentação da abordagem e nos resultados obtidos, os objetivos propostos foram alcançados. A abordagem proposta foi testada com dados reais. Foram obtidos resultados comparáveis a outros métodos usados na literatura. Estas constatações indicam potencial para melhoria e desenvolvimento para uso como sistema auxiliar de avaliação de risco de órbita.

Referências Bibliográficas

- [1] ALTMAN, N. S. **An introduction to kernel and nearest-neighbor nonparametric regression.** *The American Statistician*, 46(3):175–185, 1992.
- [2] BANAEI, H.; AHMED, M.; LOUFI, A. **Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges.** *Sensors*, (13):17472–17500, 2013.
- [3] BERA, B.; NAYAK, M. **Mortality risk assessment for icu patients using logistic regression.** In: *2012 Computing in Cardiology*, p. 493–496, Sept 2012.
- [4] BOSNJAK, A.; MONTILLA, G. **Predicting mortality of icu patients using statistics of physiological variables and support vector machines.** In: *2012 Computing in Cardiology*, p. 481–484, Sept 2012.
- [5] BOYLE, P. **Gaussian Processes for Regression and Optimisation.** PhD thesis, Victoria University of Wellington, 2007.
- [6] BOYLE, P.; FREAN, M. **Dependent gaussian processes.** In: *In Advances in Neural Information Processing Systems 17*, p. 217–224. MIT Press, 2005.
- [7] BREIMAN, L.; FRIEDMAN, J.; OLSEN, R.; STONE, C. **Classification and Regression Trees.** Wadsworth and Brooks, Monterey, CA, 1984.
- [8] CAIXETA, R. **Uma revisão sistemática sobre o uso de aprendizagem de máquina no monitoramento de pacientes.** In: *III Escola Regional de Informática de Goiás*, 2015.
- [9] CITI, L.; BARBIERI, R. **Physionet 2012 challenge: Predicting mortality of icu patients using a cascaded svm-glm paradigm.** In: *Computing in Cardiology (CinC), 2012*, p. 257–260, Sept 2012.
- [10] CORTES, C.; VAPNIK, V. **Support-vector networks.** *Mach. Learn.*, 20(3):273–297, Sept. 1995.

- [11] DURICHEN, R.; PIMENTEL, M.; CLIFTON, L.; SCHWEIKARD, A.; CLIFTON, D. **Multitask gaussian processes for multivariate physiological time-series analysis**. *Biomedical Engineering, IEEE Transactions on*, 62(1):314–322, Jan 2015.
- [12] ESFANDIARI, N.; BABAVALLIAN, M.; MOGHADAM, A.; TABAR, V. **Knowledge discovery in medicine: Current issue and future trend**. *Expert Systems with Applications*, (41):4434–4463, 2014.
- [13] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From data mining to knowledge discovery in databases**. *AI Magazine*, (17):37–54, 1996.
- [14] FINDLAY, G.; SHOTTON, H.; KELLY, K.; MASON, M. **Cardiac arrest procedures: Time to intervene?** Technical report, National Confidential Enquiry into Patient Outcome and Death, UK, 2012.
- [15] GOLDBERGER, A. L.; AMARAL, L. A. N.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. **PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals**. *Circulation*, 101(23):e215–e220, 2000 (June 13). *Circulation Electronic Pages*: <http://circ.ahajournals.org/cgi/content/full/101/23/e215> PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.
- [16] GUARNIZO, C.; ÁLVAREZ, M. A.; OROZCO, A. A. **Indian Buffet Process for Model Selection in Latent Force Models**, p. 635–642. Springer International Publishing, Cham, 2015.
- [17] HAMILTON, S.; HAMILTON, J. **Predicting in-hospital-death and mortality percentage using logistic regression**. In: *2012 Computing in Cardiology*, p. 489–492, Sept 2012.
- [18] HAYKIN, S. **Communication Systems, 3rd ed.** Wiley.
- [19] HOFFMAN, M. **PyGP**. <https://github.com/mwhoffman/pygp> (acessado 3 de Fevereiro de 2016).
- [20] HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. 2000. "A Wiley-Interscience publication."
- [21] JOHNSON, A.; DUNKLEY, N.; MAYAUD, L.; TSANAS, A.; KRAMER, A.; CLIFFORD, G. **Patient specific predictions in the intensive care unit using a bayesian ensemble**. In: *Computing in Cardiology (CinC), 2012*, p. 249–252, Sept 2012.

- [22] KENWARD, T. J.; VLACKONIKOLIS, G. I.; PAYNE, S.; CASTLE, N.; CROUCH, R.; INESON, N.; SHAIKH, L. **Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital.** *Resuscitation*, 54(2):115–123, 2002.
- [23] KRAJNAK, M.; XUE, J.; KAISER, W.; BALLONI, W. **Combining machine learning and clinical rules to build an algorithm for predicting icu mortality risk.** In: *2012 Computing in Cardiology*, p. 401–404, Sept 2012.
- [24] LE GALL, J.; LOIRAT, P.; ALPEROVITCH, A.; GLASER, P.; GRANTHIL, C.; MATHIEU, D.; MERCIER, P.; THOMAS, R.; VILLERS, D. **A simplified acute physiology score for icu patients.** *Critical Care Medicine*, 12(11):975–7, Nov 1984.
- [25] LEE, C.; ARZENO, N.; HO, J.; VIKALO, H.; GHOSH, J. **An imputation-enhanced algorithm for icu mortality prediction.** In: *2012 Computing in Cardiology*, p. 253–256, Sept 2012.
- [26] MACAŠ, M.; KUZILEK, J.; ODSTRČILÍK, T.; HUPTYCH, M. **Linear bayes classification for mortality prediction.** In: *2012 Computing in Cardiology*, p. 476–473, Sept 2012.
- [27] MARCO, L.; BOJARNEJAD, M.; KING, S.; DUAN, W.; MARIA, C.; ZHENG, D.; MURRAY, A.; LANGLEY, P. **Robust prediction of patient mortality from 48 hour intensive care unit data.** In: *2012 Computing in Cardiology*, p. 477–480, Sept 2012.
- [28] McMILLAN, S.; CHIA, C.; VAN ESBROECK, A.; RUBINFELD, I.; SYED, Z. **Icu mortality prediction using time series motifs.** In: *2012 Computing in Cardiology*, p. 265–268, 2012.
- [29] NGUYEN, T. V.; BONILLA, E. V. **Collaborative multi-output gaussian processes.** In: *UAI*, 2014.
- [30] OTTE, C. **Safe and interpretable machine learning: A methodological review.** In: Moewes, C.; Nürnberger, A., editors, *Computational Intelligence in Intelligent Data Analysis*, volume 445 de **Studies in Computational Intelligence**, p. 111–122. Springer Berlin Heidelberg, 2013.
- [31] PALIWAL, M.; KUMAR, U. A. **Neural networks and statistical techniques: A review of applications.** *Expert Syst. Appl.*, p. 2–17, 2009.
- [32] PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISSEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY,

- E. **Scikit-learn: Machine learning in Python**. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [33] PIMENTEL, M. A. F.; CLIFTON, D. A.; TARASSENKO, L. **Gaussian process clustering for the functional characterisation of vital-sign trajectories**. In: *IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2013, Southampton, United Kingdom, September 22-25, 2013*, p. 1–6, 2013.
- [34] POLLARD, T.; HARRA, L.; WILLIAMS, D.; HARRIS, S.; MARTINEZ, D.; FONG, K. **2012 physionet challenge: An artificial neural network to predict mortality in icu patients and application of solar physics analysis methods**. In: *2012 Computing in Cardiology*, p. 485–488, Sept 2012.
- [35] RABINER, L. R.; JUANG, B. H. **An introduction to hidden markov models**. *IEEE ASSP Magazine*, 1986.
- [36] RASMUSSEN, C. E.; WILLIAMS, C. K. I. **Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)**. The MIT Press, 2005.
- [37] SAEED, M.; VILLARROEL, M.; REISNER, A. T.; CLIFFORD, G.; LEHMAN, L.-W.; MOODY, G.; HELDT, T.; KYAW, T. H.; MOODY, B.; MARK, R. G. **Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database**. *Critical Care Medicine*, 39:952–960, May 2011.
- [38] SALVADOR, S.; CHAN, P. **Toward accurate dynamic time warping in linear time and space**. *Intell. Data Anal.*, 11(5):561–580, Oct. 2007.
- [39] SEVEREYN, E.; ALTUVE, M.; NG, F.; LOLLETT, C.; WONG, S. **Towards the prediction of mortality in intensive care units patients: a simple correspondence analysis approach**. *Computing in Cardiology 2012*, (39):469–472, 2012.
- [40] SILVA, I.; MOODY, G.; SCOTT, D.; CELI, L.; MARK, R. **Predicting in-hospital mortality of patients in icu: The physionet/computing in cardiology challenge 2012**. *Computing in Cardiology 2012*, (39):245–248, 2012.
- [41] SINGLA, J.; GROVER, D.; BHANDARI, A. **Article: Medical expert systems for diagnosis of various diseases**. *International Journal of Computer Applications*, 93(7):36–43, May 2014.
- [42] SUN, X.; ZHAO, Y. **Integrated exemplar-based template matching and statistical modeling for continuous speech recognition**. *EURASIP Journal on Audio, speech, and Music Processing*, 4:1–16, 2014.

- [43] TANIDA, K. **fastdtw**. <https://github.com/slaypni/fastdtw> (acessado 3 de Junho de 2016).
- [44] VAIRAVAN, S.; ESHELMAN, L.; HAIDER, S.; FLOWER, A.; SEIVER, A. **Prediction of mortality in an intensive care unit using logistic regression and a hidden markov model**. In: *Computing in Cardiology (CinC)*, 2012, p. 393–396, Sept 2012.
- [45] VAN ROSSUM, G. **Python**, 2016. "<http://www.python.org> (acessado 3 de Junho de 2016)".
- [46] VIEIRA, M. S. **Perfil geográfico e clínico de pacientes admitidos na uti através da central de regulação de internações hospitalares**. *Com. Ciências Saúde*, 3(22):201–210, 2011.
- [47] WANG, W.; WANG, H.; HEMPEL, M.; PENG, D.; SHARIF, H.; CHEN, H.-H. **Secure stochastic ecg signals based on gaussian mixture model for e-healthcare systems**. *Systems Journal, IEEE*, 5(4):564–573, Dec 2011.
- [48] XIA, H.; DALEY, B. J.; PETRIE, A.; ZHAO, X. **A neural network model for mortality prediction in icu**. In: *2012 Computing in Cardiology*, p. 261–264, Sept 2012.
- [49] XU, J.; LI, D.; ZHANG, Y.; DJULOVIC, A.; LI, Y.; ZENG, Y. **Cinc challenge: Cluster analysis of multi-granular time-series data for mortality rate prediction**. In: *2012 Computing in Cardiology*, p. 497–500, 2012.
- [50] YI, C.; SUN, Y.; TIAN, Y. **Cinc challenge: Predicting in-hospital mortality in the intensive care unit by analyzing histograms of medical variables under cascaded adaboost model**. In: *2012 Computing in Cardiology*, p. 397–400, 2012.