

**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

Anderson Hiroshi de Siqueira

**Inferência em mistura dinâmica de gases utilizando redes
bayesianas**

São Carlos

2019

Anderson Hiroshi de Siqueira

Inferência em mistura dinâmica de gases utilizando redes bayesianas

Monografia apresentada ao Curso de Engenharia Elétrica com Ênfase em Eletrônica, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Engenheiro Eletricista.

Orientador: Prof. Dr. Carlos Dias Maciel

**São Carlos
2019**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

S545i Siqueira, Anderson Hiroshi
 Inferência em mistura dinâmica de gases
 utilizando redes bayesianas / Anderson Hiroshi
 Siqueira; orientador Carlos Dias Maciel. São Carlos,
 2019.

 Monografia (Graduação em Engenharia Elétrica com
 ênfase em Eletrônica) -- Escola de Engenharia de São
 Carlos da Universidade de São Paulo, 2019.

 1. Redes Bayesianas. 2. Redes Bayesianas Gaussianas
 . 3. Mistura de Gases. 4. Série Temporal. I. Título.

FOLHA DE APROVAÇÃO

Nome: Anderson Hiroshi de Siqueira

Título: “Inferência em mistura dinâmica de gases utilizando redes bayesianas”

Trabalho de Conclusão de Curso defendido e aprovado

em 28/11/2019,

com NOTA 10,0 (dez, zero), pela Comissão Julgadora:

Prof. Associado Carlos Dias Maciel - Orientador - SEL/EESC/USP

Mestre Talysson Manoel de Oliveira Santos - Doutorando - SEL/EESC/USP

Mestre Jordão Natal de Oliveira Júnior - Doutorando - SEL/EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Prof. Associado Rogério Andrade Flauzino

Este trabalho é dedicado ao meu avô, Masafumi Nagata.

AGRADECIMENTOS

Aos meus pais, meus avós e irmão pelo suporte incondicional durante todos esses anos de estudo e pelo incentivo na busca pelas minhas paixões.

A minha namorada que sempre me apoiou, nunca permitiu que eu desistisse em momento algum e que me ajudou a sobreviver as perdas e os momentos mais sombrios que já passei.

Aos amigos que encontrei em São Carlos, com quem nos últimos cinco anos pude dividir todas as conquistas e frustrações, companheiros que nunca me abandonaram e me ajudaram a passar por essa etapa da minha vida.

Aos mentores e colegas do grupo Warthog Robotics, com quem aprendi muito, e continuo a aprender, e com quem me diverti em quantidade inversamente proporcional a quantidade de sono que tive durante os anos de trabalho com o grupo.

Ao Prof. Dr. Carlos Dias Maciel pela orientação.

“O homem não é nada além daquilo que a educação faz dele.”

Immanuel Kant

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Objetivos	20
1.2	Organização	20
2	TEORIA	21
2.1	Teoria Probabilística Básica	21
2.2	Teoria Básica de Grafos	23
2.3	Redes Bayesianas	24
2.4	Inferência em Rede com Variáveis Contínuas	26
2.5	Aprendizado de Parâmetros de Rede com Variáveis Contínuas	30
2.5.1	Variáveis Normalmente Distribuídas Multivariadas	31
2.5.2	Conversão para Rede Bayesiana Gaussiana	32
2.5.3	Aprendendo os Parâmetros de uma Rede Gaussiana Bayesiana	33
2.6	Desvio Médio Absoluto	34
3	MATERIAIS E MÉTODOS	35
3.1	Conjunto de Dados: Mistura de Gases	35
3.2	Software	37
3.2.1	Python	37
3.2.2	C++	37
3.3	Rede Bayesiana	38
4	RESULTADOS	43
4.1	Aprendizado	43
4.2	Estrutura A	45
4.3	Estrutura B	47
5	CONCLUSÕES	51
	REFERÊNCIAS	53

LISTA DE FIGURAS

Figura 1 – Exemplo de DAG	24
Figura 2 – Uma rede bayesiana	25
Figura 3 – Uma rede bayesiana gaussiana	26
Figura 4 – Uma rede bayesiana gaussiana	30
Figura 5 – Algoritmo para obtenção de matriz de precisão	33
Figura 6 – Da esquerda para direita: sensores TGS-2600, TGS-2602, TGS-2610, TGS-2620 (imagem adaptada da ficha de especificação dos componentes)	35
Figura 7 – Concentração de CO e Etileno na primeira hora	36
Figura 8 – Concentração de CO e Etileno na segunda hora	36
Figura 9 – Rede bayesiana gaussiana para inferência de concentração em mistura de gases: Estrutura A	39
Figura 10 – Rede bayesiana gaussiana para inferência de concentração em mistura de gases: Estrutura B	40
Figura 11 – Inferência e <i>set point</i> para CO com Estrutura A em conjunto de treino	45
Figura 12 – Inferência e <i>set point</i> para Etileno com Estrutura A em conjunto de treino	46
Figura 13 – Inferência e <i>set point</i> para CO com Estrutura A em conjunto de teste .	46
Figura 14 – Inferência e <i>set point</i> para Etileno com Estrutura A em conjunto de teste	47
Figura 15 – Inferência e <i>set point</i> para CO com Estrutura B em conjunto de treino	48
Figura 16 – Inferência e <i>set point</i> para Etileno com Estrutura B em conjunto de treino	48
Figura 17 – Inferência e <i>set point</i> para CO com Estrutura B em conjunto de teste .	49
Figura 18 – Inferência e <i>set point</i> para Etileno com Estrutura B em conjunto de teste	49

LISTA DE TABELAS

Tabela 1 – Descrição das variáveis aleatórias da figura 2	25
Tabela 2 – Relação entre variável e índice para algoritmo de aprendizado de parâmetros	39
Tabela 3 – Relação entre variável e índice para Estrutura B	40
Tabela 4 – Médias μ_i^* e variâncias σ_i^{2*} atualizadas	43
Tabela 5 – Coeficientes de correlação atualizados b_{ij}^* (colunas 1 à 6)	43
Tabela 6 – Coeficientes de correlação atualizados b_{ij}^* (colunas 7 à 12)	44
Tabela 7 – Coeficientes de correlação atualizados b_{ij}^* (colunas 13 à 18)	44
Tabela 8 – Comparação do desvio médio absoluto do conjunto de treino com inferência e média dos dados (Estrutura A)	46
Tabela 9 – Comparação do desvio médio absoluto do conjunto de teste com inferência e média dos dados (Estrutura A)	47
Tabela 10 – Comparação do desvio médio absoluto do conjunto de treino com inferência e média dos dados (Estrutura B)	49
Tabela 11 – Comparação do desvio médio absoluto do conjunto de teste com inferência e média dos dados (Estrutura B)	50

RESUMO

Siqueira, A. H. **Inferência em mistura dinâmica de gases utilizando redes bayesianas**. 2019. 54p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

A detecção e medição de gases é uma preocupação que surgiu há mais de um século. Do uso de pequenos animais para detecção de gases nocivos no século 20, até utilização de sensores semicondutores atualmente, o problema ainda é bastante relevante em diversas áreas como diagnóstico médico, análise de alimentos e monitoramento de ambientes industriais. Apesar dos avanços na tecnologia, os sensores atuais sofrem com problemas como baixa seletividade, respostas não lineares e instabilidade, dificultando seu uso para medição de concentração de gases em misturas de gases. Neste trabalho é proposto o uso de rede bayesiana gaussiana para inferência das concentrações individuais de dois gases (CO e C_2H_2) em uma mistura de gases no tempo.

Palavras-chave: Redes Bayesianas, Redes Bayesianas Gaussianas, Mistura de Gases, Série Temporal.

ABSTRACT

Siqueira, A. H. **Inference in dynamic gas mixtures using bayesian networks.** 2019. 54p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

Gas detection and measurement is a concern that has emerged over a century ago. From the use of small animals to detect harmful gases in the 20th century, to the use of semiconductor sensors today, the problem is still quite relevant in many areas such as medical diagnosis, food analysis and industrial safety. Despite advances in technology, today's sensors suffer from problems such as low selectivity, nonlinear responses, and instability, making it difficult to use for gas concentration measurement in gas mixtures. This work proposes the use of Gaussian Bayesian Networks to infer the individual concentrations of two gases (CO and C_2H_2) in a mixture of gases over time.

Keywords: Bayesian Network, Gaussian Bayesian Network, Gas Mixture, Time Series.

1 INTRODUÇÃO

A detecção e medição de gases é uma preocupação que surgiu há mais de um século, com o uso de animais pequenos como canários, coelhos e ratos em minas de carvão para detecção de gases nocivos aos trabalhadores sendo comum no início do século 20 (BURRELL; SEIBERT, 1914). Desde então, com o avanço da tecnologia, foram desenvolvidos outros métodos como a lanterna de segurança e, mais tarde, sensores eletrônicos para detecção de gases. Atualmente, os sensores de gases utilizados são, predominantemente, transdutores que convertem informação química em alguma forma de sinal elétrico, sendo classificados de acordo com seu princípio de operação (YUNUSA et al., 2014). Assim, com o aumento da variedade e qualidade de sensores disponíveis, nas últimas décadas diversos estudos e aplicações foram desenvolvidas em áreas como diagnóstico médico, análise de alimentos e monitoramento de ambientes industriais (LIU; ZHOU; LEI, 2015).

Contudo, sensores de gás do tipo metal-óxido semicondutor sofrem com uma baixa seletividade, respostas não lineares e instabilidade. Apesar da melhora desses sensores nos últimos anos, ainda é impossível identificar diversos gases com apenas um sensor (LIU; ZHOU; LEI, 2015).

Uma forma de abordar esse problema é o uso de um vetor com vários sensores de gases de modelos diferentes, assim obtendo mais informações sobre o gás. Em teoria, é possível desenvolver uma expressão matemática baseada no conhecimento da forma de funcionamento do sensor e nos dados obtidos, mas na prática alguns dos parâmetros para expressão precisam ser estabelecidos de forma empírica. Com o aumento no número de componentes de gases a expressão matemática para os sensores se torna mais complexa, já que a detecção de uma molécula pode ser influenciada pela presença de outro tipo de molécula (SUNDGREN et al., 1991).

Outra abordagem proposta para esse problema é o uso de técnicas de inteligência artificial para reconhecimento de padrões através de conjunto de dados extraídos do vetor de sensores através de experimentos, eliminando a necessidade de um modelo criado a partir da teoria de operação de cada modelo de sensor (Bermak; Belhouari, 2006).

Dentre os modelos de inteligência artificial propostos na literatura, as Redes Bayesianas se destacam por descreverem probabilidade condicional entre vértices de um grafo. Essa característica trás implicações interessante, neste modelo não existe vértices de entrada ou saída, a partir de uma evidencia em qualquer nó pode-se inferir o valor mais provável dos outros vértices, assim este modelo trabalha bem quando existem dados faltantes, como é muito comum na área médica, por exemplo. Este modelo também captura a relação de causa e efeito entre as variáveis (NEAPOLITAN, 2004).

Neste trabalho, é proposto o uso de Redes Bayesianas com variáveis contínuas para inferência da concentração de uma mistura de gases, a partir de um vetor de sensores de gases metal-óxido semicondutores.

1.1 Objetivos

O objetivo geral deste trabalho é produzir uma **rede bayesiana para inferência das concentrações individuais de dois gases em uma mistura que varia no tempo**. Mais especificamente, pode-se dividir nos seguintes objetivos secundários:

- Estudo de modelos bayesianos com variável contínua;
- Implementação do algoritmo de aprendizado de parâmetros do modelo;
- Construção de um ou mais modelos a partir dos parâmetros obtidos;
- Implementação do algoritmo de inferência dos gases nos modelos;
- Comparação dos resultados.

1.2 Organização

O trabalho está organizado em 5 capítulos. No capítulo 2 é feito um breve estudo sobre teoria probabilística básica, teoria básica de grafos e redes bayesianas gaussianas. No capítulo 3 são apresentadas as linguagens de programação, bibliotecas utilizadas e o método de implementação e teste dos modelos. No capítulo 4 são apresentados os resultados obtidos com os modelos. Finalmente, o capítulo 5 apresenta discussão dos resultados e as considerações finais.

2 TEORIA

Neste capítulo serão discutidos os princípios teóricos necessários para contextualizar este trabalho, tratando dos fundamentos da teoria probabilística, fundamentos de teoria de grafos e por fim, redes Bayesianas.

2.1 Teoria Probabilística Básica

Neste capítulo serão apresentados alguns conceitos necessários para o desenvolvimento das ideias presentes nas próximas seções.

Um dos conceitos mais importantes da teoria probabilística e fundamental para este trabalho é a probabilidade condicional.

WALPOLE (2006) define que, dado dois eventos A e B, tal que a probabilidade do evento B seja diferente de zero. Então a probabilidade de ocorrer o evento A dado que o evento B foi observado, isto é, a **probabilidade condicional**, é dada por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ dado que } P(B) > 0. \quad (2.1)$$

Na equação 2.1, o termo $P(A \cap B)$ representa a probabilidade da intersecção dos eventos A e B. No caso deste termo ser igual a zero, diz-se que os eventos A e B são mutuamente exclusivos.

Segundo WALPOLE (2006), os eventos A e B são considerados independentes se pelo menos uma das condições a seguir for verdadeira:

1. $P(A|B) = P(A)$ e $P(B) \neq 0$.
2. $P(A) = 0$ ou $P(B) = 0$.

Então, a probabilidade da intersecção dos eventos A e B, no caso dos eventos serem independentes, é dado por:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B). \quad (2.2)$$

O Teorema de Bayes permite a inferência da probabilidade de determinado evento, dado o conhecimento da ocorrência de algum evento conhecido. O teorema é definido a seguir.

Dado dois eventos A e B, de forma que a probabilidades de ambos não seja nula, tem-se

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.3)$$

Ainda, esse resultado pode ser expandido. Supondo uma série de eventos mutuamente exclusivos e exaustivos A_1, A_2, A_k , de forma que a $P(A_i) \neq 0$ para $i = 1, 2, \dots, k$, e algum evento B tal que $P(B) \neq 0$, então:

$$P(A_k|B) = \frac{P(A_k \cap B)}{\sum_{i=1}^k P(A_i \cap B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^k P(A_i)P(B|A_i)}. \quad (2.4)$$

A prova pode ser encontrada em WALPOLE (2006, p. 71).

Quando não se tem informações sobre a probabilidade condicional de algum evento, e se tem as informações do lado direito da equação 2.3 ou 2.4, é possível fazer uma inferência bayesiana.

Segundo WALPOLE (2006, p. 172) “A distribuição contínua mais importante em todo campo da estatística é a distribuição normal”. Essa distribuição é caracterizada pelo seu gráfico em formato de sino, e é extensamente utilizada em diversas áreas do conhecimento. É descrita com a seguinte formula:

$$p(x) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (2.5)$$

Na equação 2.5, diz-se que a variável aleatória X tem uma distribuição normal com média $E(X) = \mu$ e variância $V(X) = \sigma^2$.

Uma distribuição normal possui média zero e variância infinita representa completa incerteza, enquanto um distribuição com média contante e variância zero representa completa certeza sobre o estado da variável.

Uma distribuição normal multivariada é uma generalização da distribuição normal unidimensional (NEAPOLITAN, 2004). Neste caso, uma um vetor \mathbf{X} que possui distribuição normal multivariada é representada da seguinte forma:

$$\rho_{\mathbf{X}}(\mathbf{x}) = N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\psi}) \quad (2.6)$$

onde o vetor \mathbf{X} é um vetor n-dimensional de variáveis aleatórias, denotado na

equação 2.7.

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \quad (2.7)$$

A matriz de covariância ψ e o vetor μ de médias são denotados, respectivamente, como nas equações 2.8 e 2.9:

$$\psi = \begin{pmatrix} V(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & V(X_n) \end{pmatrix} \quad (2.8)$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad (2.9)$$

A matriz de covariância ψ também é comumente representada pela equação 2.10, esta segunda notação é a adotada nas seções seguintes por simplicidade.

$$\psi = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} \quad (2.10)$$

Ainda, NEAPOLITAN (2004, p. 418), define a matriz de precisão \mathbf{T} como sendo a matriz inversa de ψ .

$$T = \psi^{-1} \quad (2.11)$$

2.2 Teoria Básica de Grafos

A teoria de grafos estuda estruturas chamadas de grafo, que representam a relação entre objetos. Como estabelecido por WEST (2000), um **grafo** \mathbf{G} é definido por um conjunto não vazio de vértices $\mathbf{V}(\mathbf{G})$, um conjunto de arestas $\mathbf{E}(\mathbf{G})$, e a relação que associa cada aresta a dois vértices.

A relação entre os vértices, neste caso é simétrica. Mas existem casos que se deseja modelar relações assimétricas entre os vértices. Nestes casos usa-se um grafo direcionado.

Um **grafo direcionado** $G(V,E)$, como estabelecido por WEST (2000), é definido por um conjunto não vazio de vértices V , um conjunto de arestas E , e a relação que associa cada aresta a um par ordenado de vértices. Neste caso, se $(X,Y) \in E$, diz-se que existe uma aresta de X para Y , porém, no contexto desse trabalho, X é denominado pai de Y , e Y é denominado descendente de X .

Dado um grafo $G(V,E)$, supondo que existam os vértices $V(G) = [X_1, X_2, \dots, X_n]$ para $n \geq 2$, e que existam as arestas $(X_i, X_{i+1}) \in E(G)$ para $1 \leq i \leq n-1$. O conjunto de n arestas que conectam os vértices de X_1 até X_n é chamado de **caminho** de X_1 a X_n . Ainda, o caminho de um nó até ele mesmo é denominado **ciclo direto**. Mas, no caso de não ocorrer a repetição de nenhum vértice no caminho ele é chamado de **caminho simples**.

Supondo, então, um grafo direcionado G que não possua caminho que seja um ciclo direto. Este grafo é chamado de **grafo acíclico dirigido**, referido daqui em diante como **DAG** (*Directed Acyclic Graph*) (NEAPOLITAN, 2004, p. 31). Um exemplo de um grafo acíclico dirigido com quatro vértices é ilustrado na figura 1.

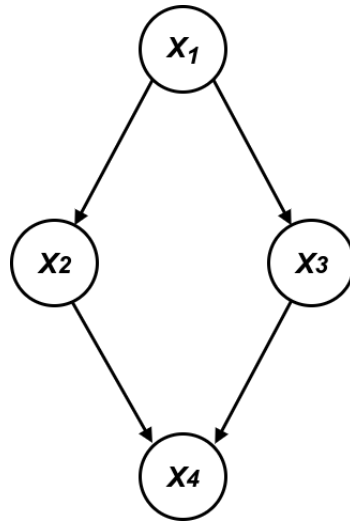


Figura 1: Exemplo de DAG

2.3 Redes Bayesianas

Os assuntos abordadas nas seções 2.1 e 2.2 são fundamentais para tratar das Redes Bayesianas. Contudo, existe mais uma definição importante a ser abordada: a condição de Markov.

Supondo que exista distribuição de probabilidade conjunta P de um conjunto de variáveis aleatórias V e um DAG $G(V,E)$. Diz-se que (G,P) satisfaz a **condição de Markov** se para cada variável $X \in V$, X é condicionalmente independente do conjunto de todos os seus não-descendentes dado o conjunto de todos os seus pais. Denota-se os

não-descendentes de X como ND_X e os pais de X como PA_X , dessa forma, indica-se essa relação como:

$$I_p(\{X\}, ND_X | PA_X). \quad (2.12)$$

Caso o conjunto (G, P) satisfaça a condição de Markov, diz-se então, que \mathbf{G} e \mathbf{P} satisfazem a condição de Markov entre si (NEAPOLITAN, 2004, p. 31).

Finalmente, dado uma distribuição de probabilidade conjunta \mathbf{P} de um conjunto de variáveis aleatórias \mathbf{V} e um DAG $\mathbf{G}(\mathbf{V}, \mathbf{E})$, chama-se (\mathbf{G}, \mathbf{P}) uma **rede bayesiana** se G e P satisfazem a condição de Markov entre si (NEAPOLITAN, 2004, p. 40).

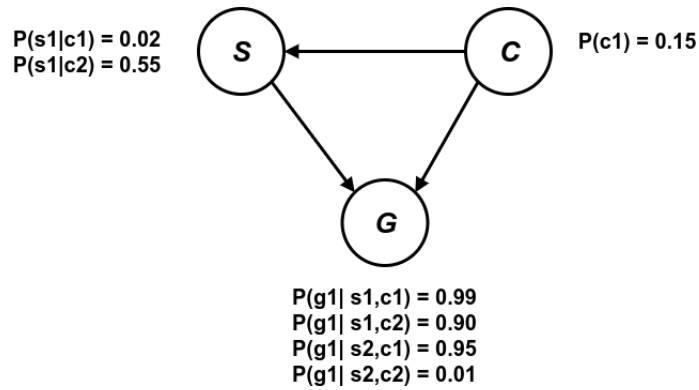


Figura 2: Uma rede bayesiana

A figura 2 mostra uma rede bayesiana com três variáveis aleatórias. A descrição de cada uma das variáveis está detalhada na tabela 1.

Tabela 1: Descrição das variáveis aleatórias da figura 2

Variavel	Valor	Descrição
S	s1	Sprinkler foi utilizado
	s2	Sprinkler não utilizado
C	c1	Choveu
	c2	Não choveu
G	g1	Grama está molhada
	g2	Grama não está molhada

Dado um conjunto de variáveis aleatórias V , se para cada $X, Y \in V$ for definida uma aresta de X para Y , e se, e somente se X é uma causa direta de Y , chama-se a DAG resultante de DAG causal NEAPOLITAN (2004).

O conceito de causalidade entre X e Y , neste contexto, pode ser entendido como a fato de uma mudança nos valores de X causar uma variação, algumas vezes, no valor de

Y. NEAPOLITAN (2004) e COOPER (1999) fazem uma discussão mais aprofundada do significado de causalidade e de como uma DAG causal em geral satisfaz a condição de Markov. Contudo, a definição mais simples é suficiente para este trabalho.

Uma rede bayesiana gaussiana é uma rede bayesiana onde seus vértices são exclusivamente variáveis aleatórias com distribuição normal, como ilustra a figura 3.

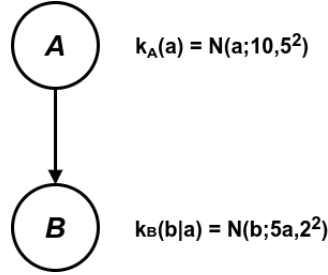


Figura 3: Uma rede bayesiana gaussiana

2.4 Inferência em Rede com Variáveis Contínuas

O algoritmo descrito nesta seção é proposto por NEAPOLITAN (2004, p. 185-190) para inferência em redes bayesianas gaussianas individualmente conectadas.

Suponha uma rede bayesiana gaussiana onde cada uma das variáveis aleatória tem sua distribuição definida por uma combinação linear de seus pais. Assim, sendo PA_X os pais de X , pode-se escrever que:

$$x = w_X + \sum_{Z \in PA_X} b_{xz} z \quad (2.13)$$

onde w_X tem distribuição $N(w_X; 0, \sigma_{w_X}^2)$ e modela a incerteza em relação ao valor de X , e b_{XZ} representa a influencia de Z em X . Dessa forma a distribuição da variável X é dada por:

$$\rho(x|pa_x) = N(x; \sum_{Z \in PA_x} b_{xz} z, \sigma_{w_X}^2) \quad (2.14)$$

Note que o vértice X não tem média própria, sua média é uma combinação linear de seus pais. Contudo, a variância é própria do vértice. Porém, caso de um vértice raiz, isto é, um vértice que não tem nenhum pai, sua distribuição é dada simplesmente por:

$$\rho(x) = N(x; \mu_x, \sigma_X^2). \quad (2.15)$$

A seguir, são apresentadas as fórmulas utilizadas no algoritmo de inferência. E, neste caso, para proporcionar uma leitura mais fácil, a variância é representada por σ , ao invés de σ^2 .

Valor λ é:

$$\sigma_X^\lambda = \left[\sum_{U \in CH_X} \frac{1}{\sigma_{UX}^\lambda} \right]^{-1} \quad (2.16)$$

$$\mu_X^\lambda = \sigma_X^\lambda \sum_{U \in CH_X} \frac{\mu_{UX}^\lambda}{\sigma_{UX}^\lambda} \quad (2.17)$$

Valor π é:

$$\sigma_X^\pi = \sigma_{W_X} \sum_{Z \in PA_X} b_{XZ}^2 \sigma_{XZ}^\pi \quad (2.18)$$

$$\mu_X^\pi = \sum_{Z \in PA_X} b_{XZ} \mu_{XZ}^\pi \quad (2.19)$$

A média e variância esperada para uma variável X é calculada:

$$\sigma_X = \frac{\sigma_X^\pi \sigma_X^\lambda}{\sigma_X^\pi + \sigma_X^\lambda} \quad (2.20)$$

$$\mu_X = \frac{\sigma_X^\pi \mu_X^\pi + \sigma_X^\lambda \mu_X^\lambda}{\sigma_X^\pi + \sigma_X^\lambda} \quad (2.21)$$

A mensagem π de Z para um filho X:

$$\sigma_{XZ}^\pi = \frac{1}{\sigma_Z^\pi} + \sum_{Y \in CH_Z - \{X\}} \frac{1}{\sigma_{YZ}^\lambda} \quad (2.22)$$

$$\mu_{XZ}^\pi = \frac{\frac{\mu_Z^\pi}{\sigma_Z^\pi} + \sum_{Y \in CH_Z - \{X\}} \frac{\mu_{YZ}^\lambda}{\sigma_{YZ}^\lambda}}{\frac{1}{\sigma_Z^\pi} + \sum_{Y \in CH_Z - \{X\}} \frac{1}{\sigma_{YZ}^\lambda}} \quad (2.23)$$

A mensagem λ de X para um pai Y:

$$\sigma_{YX}^\lambda = \frac{1}{b_{YX}^2} \left[\sigma_Y^\lambda + \sigma_{W_Y}^\lambda + \sum_{Z \in PA_Y} b_{YZ}^2 \sigma_{YZ}^\pi \right] \quad (2.24)$$

$$\mu_{YX}^\lambda = \frac{1}{b_{YX}} \left[\mu_Y^\lambda + \sum_{Z \in PA_Y} b_{YZ} \mu_{YZ}^\pi \right] \quad (2.25)$$

Ao se instanciar uma variável V com valor \hat{v} , faz-se:

$$\sigma_V^\pi = \sigma_V^\lambda = \sigma_V = 0. \quad (2.26)$$

$$\mu_V^\pi = \mu_V^\lambda = \mu_V = \hat{v}. \quad (2.27)$$

A ignorância total em relação ao valor de determinada variável é representado pela distribuição normal:

$$N(x, \mu = 0, \sigma = \infty). \quad (2.28)$$

O algoritmo proposto para cálculo de expectativa das variáveis aleatórias na rede possui quatro funções principais: `iniciarRede`, `atualizarRede`, `enviarMensagemPi`, `enviarMensagemLambda`. Para facilitar a leitura, os valores λ e π de um vértice A serão representados val_A^λ e val_A^π , respectivamente. Ainda, as mensagens λ e π do vértice A para o B serão representadas men_{AB}^λ e men_{BA}^π , respectivamente.

Note também que os valores e as mensagens λ e π representam uma distribuição normal, e nos algoritmos apresentados os valores e mensagens serão denotados como tal. Finalmente, os algoritmos para as funções são apresentados a seguir:

Algoritmo 1 Inicialização da Rede

```

1: função INICIARREDE(RedeBayesiana &  $B = (G(E, V), P)$ )
2: para cada  $X \in V$  faça
3:    $val_X^\lambda = N(x; 0, \infty);$  ▷ Ignorância total
4:   para cada  $Z$  pai de  $X$  faça
5:      $men_{XZ}^\lambda = N(x; 0, \infty);$  ▷ Inicializa mensagem  $\lambda$ 
6:   para cada  $Y$  filho de  $X$  faça
7:      $men_{YX}^\pi = N(x; 0, \infty);$  ▷ Inicializa mensagem  $\pi$ 
8:
9: para cada vértice raiz  $R$  faça
10:    $expectativa_R = N(r; \mu_R, \sigma_R);$  ▷ Inicializa valor da expectativa
11:    $val_R^\pi = N(r; \mu_R, \sigma_R);$  ▷ Inicializa valor  $\pi$ 
12:   para cada  $X$  filho de  $R$  faça
13:     enviarMensagemPi(R, X); ▷ Atualiza filhos em relação a expectativa
14:
15: fim função

```

Algoritmo 2 Atualização da Rede

```

1: função ATUALIZARREDE(RedeBayesiana & B = ( $\mathbf{G}(E, V)$ ,  $\mathbf{P}$ ), Evidencias & A, Vértice V, Valor v)
2:  $A = A \cup V$ ; ▷ Adiciona variável a lista de evidencias
3:  $val_V^\pi = val_V^\lambda = expectativa_V = N(v; v, 0)$ ; ▷ Certeza em relação a V
4:
5: para cada Z pai de V tal que  $Z \notin A$  faça
6:     enviarMensagemLambda(V, Z); ▷ Atualiza pais em relação a expectativa
7: para cada X filho de V faça
8:     enviarMensagemPi(V, X); ▷ Atualiza filhos em relação a expectativa
9:
10: fim função
  
```

Algoritmo 3 Mensagem Lamda de Y para X

```

1: função ENVIARMENSAGEMLAMBDA(Vértice Y, Vértice X)
2:  $men_{YX}^\lambda = calcular\_mensagem\_lambda()$ ; ▷ eq. 2.24 e 2.24
3:  $val_X^\lambda = calcular\_valor\_lambda()$ ; ▷ eq. 2.16 e 2.17
4:  $expectativa_X = calcular\_expectativa()$ ; ▷ eq. 2.20 e 2.21
5:
6: para cada Z pai de X tal que  $Z \notin A$  faça
7:     enviarMensagemLambda(X, Z); ▷ Atualiza pais em relação a expectativa
8: para cada W filho de X tal que  $W \neq Y$  faça
9:     enviarMensagemPi(X, W); ▷ Atualiza filhos em relação a expectativa
10:
11: fim função
  
```

Algoritmo 4 Mensagem Pi de Z para X

```

1: função ENVIARMENSAGEMPI(Vértice Z, Vértice X)
2:  $men_{XZ}^\pi = calcular\_mensagem\_pi()$ ; ▷ eq. 2.22 e 2.22
3:
4: se  $X \notin A$  então
5:      $val_X^\pi = calcular\_valor\_pi()$ ; ▷ eq. 2.18 e 2.19
6:      $expectativa_X = calcular\_expectativa()$ ; ▷ eq. 2.20 e 2.21
7:     para cada Y filho de X faça
8:         enviarMensagemPi(X, Y); ▷ Atualiza filhos em relação a expectativa
9:     fim se
10:
11: se  $\sigma_X^\pi \neq \infty$  então ▷ Se estiver inicializado
12:     para cada W pai de X tal que  $W \neq Z$  e  $W \notin A$  faça
13:         enviarMensagemLambda(X, W);
14:
15: fim função
  
```

2.5 Aprendizado de Parâmetros de Rede com Variáveis Contínuas

Segundo NEAPOLITAN (2004), "Uma rede bayesiana gaussiana determina uma distribuição normal multivariada não singular, e vice e versa". Dessa forma, é possível transformar uma rede bayesiana em uma distribuição normal multivariada, realizar a aprendizagem de parâmetros e por fim retornar o modelo para forma de rede bayesiana gaussiana.

Nesta seção é apresentado como é realizado a aprendizagem de parâmetros em uma distribuição normal multivariada e a conversão do modelo para rede bayesiana gaussiana.

Contudo, antes dessa discussão, é necessário estabelecer uma definição para os vértices da rede um pouco diferente da proposta na equação 2.13. Nesta equação, W_X tinha distribuição $N(w_X; 0; \sigma_{W_X}^2)$, mas nesta seção é proposto que W_X não tenha mais que, necessariamente, ter média zero e passando a ser representado com uma distribuição $N(w_X; E(W_X); \sigma_{W_X}^2)$.

Com essa nova definição, a probabilidade condicional de um vértice X_i de uma rede gaussiana bayesiana é:

$$\rho_X(x_i|pa_i) = N(x_i; \mu_i + \sum_{X_j \in PA_i} b_{ij}(x_j - \mu_j), \sigma_i^2) \quad (2.29)$$

Note que essa notação incorpora a média do vértice que aparece na definição da distribuição normal multivariada, onde cada variável no vetor tem média e variância própria, além dos valores de covariância entre as variáveis.

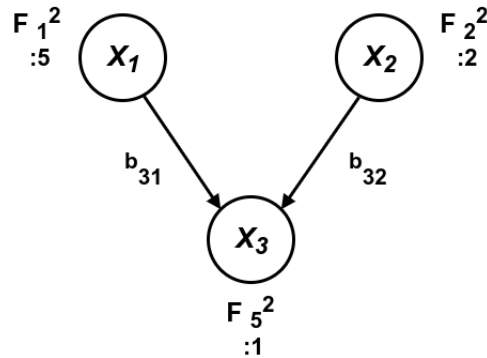


Figura 4: Uma rede bayesiana gaussiana

Outra observação importante é que essa nova definição da distribuição de W_X pode ser levada para notação anterior criando um vértice auxiliar Z , pai de X_i , de tal forma que Z tenha média μ_i , variância zero e $b_{XZ} = 1$.

2.5.1 Variáveis Normalmente Distribuídas Multivariadas

Seja um vetor \mathbf{X} com distribuição normal multivariada, supondo uma amostra normal multivariada $\mathbf{D}=[X^{(1)}, X^{(2)}, \dots, X^{(M)}]$ com os parâmetros $\{\mathbf{A}, \mathbf{R}\}$, onde \mathbf{A} é um vetor que representa a convicção existente a respeito do valor da média de \mathbf{X} e \mathbf{R} é uma matriz que representa a convicção existente a respeito do valor da matriz de precisão de \mathbf{X} , pode-se calcular:

$$\bar{\mathbf{x}} = \frac{\sum_{h=1}^M \mathbf{x}^{(h)}}{M} \quad (2.30)$$

$$s = \sum_{h=1}^M (\mathbf{x}^{(h)} - \bar{\mathbf{x}})(\mathbf{x}^{(h)} - \bar{\mathbf{x}})^T \quad (2.31)$$

E então, a função de densidade a posteriori de \mathbf{R} é dada por:

$$\rho_R(r|d) = \text{Wishart}(r; \alpha^*, \beta^*) \quad (2.32)$$

onde:

$$\beta^* = \beta + s + \frac{vM}{v+M}(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \quad (2.33)$$

$$\alpha^* = \alpha + M \quad (2.34)$$

e a função de densidade condicional a posteriori de \mathbf{A} é dada por:

$$\rho_A(a|r, d) = N(a; \boldsymbol{\mu}^*, (v^* \mathbf{r})^{-1}) \quad (2.35)$$

onde:

$$\boldsymbol{\mu}^* = \frac{v\boldsymbol{\mu} + M\bar{\mathbf{x}}}{v+M} \quad (2.36)$$

$$v^* = v + M \quad (2.37)$$

Neste contexto, os valores $\boldsymbol{\mu}$, v e β representam, respectivamente, a média, o tamanho e a matriz de covariância da amostra hipotética onde foi baseada a convicção a priori do valor de \mathbf{A} . Para modelar ignorância total sobre o modelo estudado a priori, basta admitir $\beta = 0$, $v = 0$ e $\alpha = -1$ nas expressões 2.33, 2.34, 2.36 e 2.37.

Finalmente, a nossa convicção a posteriori para o valor de \mathbf{X}^{M+1} é dado pela distribuição:

$$\rho_{\mathbf{X}^{(M+1)}}(\mathbf{x}^{(M+1)}|d) = t(\mathbf{x}^{(M+1)}; \alpha^* - n + 1, \boldsymbol{\mu}^*, \frac{v^*(\alpha^* - n + 1)}{v^* + 1}\beta^*) \quad (2.38)$$

O desenvolvimento das equações é feito por NEAPOLITAN (2004) e a prova pode ser encontrada em (DEGROOT, 2005).

2.5.2 Conversão para Rede Bayesiana Gaussiana

O método de obtenção do vetor de média e matriz de precisão de uma distribuição de uma distribuição normal multivariada a partir de uma rede bayesiana gaussiana, abordado nesta seção, foi proposto em (SHACHTER; KENLEY, 1989).

Seja:

$$t_i = \frac{1}{\sigma_i^2} \quad (2.39)$$

e:

$$b_i = \begin{pmatrix} b_{i1} \\ \vdots \\ b_{i,i-1} \end{pmatrix} \quad (2.40)$$

e o vetor de média da distribuição normal multivariada relativa a rede bayesiana gaussiana:

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \quad (2.41)$$

O algoritmo apresentado na figura 5 cria a matriz de precisão, onde n é o número de vértices na rede bayesiana:

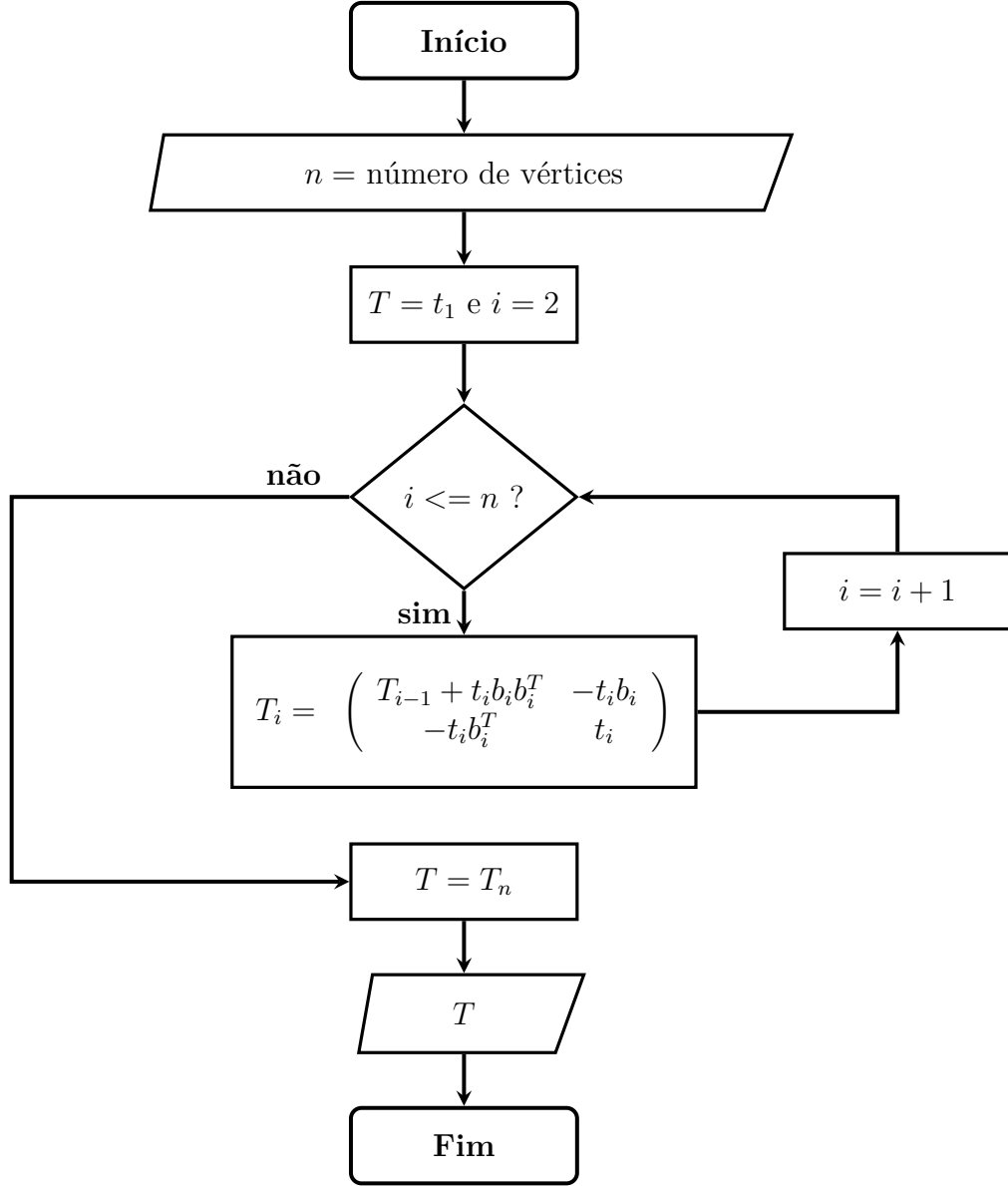


Figura 5: Algoritmo para obtenção de matriz de precisão

Um exemplo de dessa matriz para n igual a quatro é:

$$T_4 = \begin{pmatrix} \frac{b_{41}^2}{\sigma_4^2} + \frac{b_{31}^2}{\sigma_3^2} + \frac{b_{21}^2}{\sigma_2^2} + \frac{1}{\sigma_4^2} & \frac{b_{41}b_{42}}{\sigma_4^2} + \frac{b_{31}b_{32}}{\sigma_3^2} - \frac{b_{21}}{\sigma_2^2} & \frac{b_{41}b_{43}}{\sigma_4^2} - \frac{b_{31}}{\sigma_3^2} & -\frac{b_{41}}{\sigma_4^2} \\ \frac{b_{41}b_{42}}{\sigma_4^2} + \frac{b_{31}b_{32}}{\sigma_3^2} - \frac{b_{21}}{\sigma_2^2} & \frac{b_{42}^2}{\sigma_4^2} + \frac{b_{32}^2}{\sigma_3^2} + \frac{1}{\sigma_2^2} & \frac{b_{42}b_{43}}{\sigma_4^2} - \frac{b_{32}}{\sigma_3^2} & -\frac{b_{42}}{\sigma_4^2} \\ \frac{b_{41}b_{43}}{\sigma_4^2} - \frac{b_{31}}{\sigma_3^2} & \frac{b_{42}b_{43}}{\sigma_4^2} - \frac{b_{32}}{\sigma_3^2} & \frac{b_{43}^2}{\sigma_4^2} + \frac{1}{\sigma_3^2} & -\frac{b_{43}}{\sigma_4^2} \\ -\frac{b_{41}}{\sigma_4^2} & -\frac{b_{42}}{\sigma_4^2} & -\frac{b_{43}}{\sigma_4^2} & -\frac{1}{\sigma_4^2} \end{pmatrix} \quad (2.42)$$

2.5.3 Aprendendo os Parâmetros de uma Rede Gaussiana Bayesiana

A função de densidade de probabilidade encontrada na equação 2.38 não se trata de uma gaussiana, mas sim de uma distribuição t. Contudo, uma distribuição $t(\mathbf{x}; \alpha, \boldsymbol{\mu}, \mathbf{T})$

pode ser aproximada por uma distribuição $N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{T}^{-1})$, quando o limite de α tende ao infinito (DEGROOT, 2005).

Assim, com o uso dessa aproximação, pode-se encontrar a distribuição normal multivariada a partir da teoria apresentada na seção 2.5.1. Pela equação 2.38, obtém-se diretamente $\boldsymbol{\mu}$ e \mathbf{T}^{-1*} é dado por:

$$\mathbf{T}^{-1*} = \frac{v^*(\alpha^* - n + 1)}{v^* + 1} \beta^* \quad (2.43)$$

Fazendo o lado esquerdo da equação 2.43 igual a matriz simbólica obtida pelo algoritmo da seção 2.5.2 e calculando o valor numérico do lado direito da equação com uma amostra, como descrito na seção 2.5.1, obtêm-se um sistema de equações que, ao ser resolvido, permite obter os valores σ_i^{2*} e b_{ij}^* da rede bayesiana gaussiana de interesse. O valor de $\boldsymbol{\mu}^*$ da rede é obtido direto da equação 2.38.

2.6 Desvio Médio Absoluto

O desvio médio absoluto é uma métrica para descrever a performance de um determinado modelo (WILLMOTT; MATSUURA, 2005). Neste caso, essa métrica representa a média da diferença absoluta entre os valores reais e as predições do modelo.

O desvio médio absoluto é dado por:

$$DMA = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2.44)$$

onde n é o número de pontos comparados, y_i é o dado real (rótulo), \hat{y}_i é a predição do modelo, e DMA tem a mesma grandeza do que y_i e \hat{y}_i .

3 MATERIAIS E MÉTODOS

Neste capítulo será apresentada a metodologia empregada nos experimentos, bem como o conjunto de dados estudado e as tecnologias utilizadas para implementação de software necessário para este trabalho.

3.1 Conjunto de Dados: Mistura de Gases

O conjunto de dados utilizado neste trabalho trata-se de uma série temporal de 16 sensores expostos à uma mistura de gases com concentração variando ao longo do tempo, sendo as misturas etileno (C_2H_4) e metano (CH_4) no ar, e etileno e monóxido de carbono (CO) no ar, disponibilizado em arquivo CSV. Em particular, as análises deste trabalho foram desenvolvidas apenas para a mistura de etileno e CO .

Os sensores utilizados são sensores químicos da Figaro Inc., incluindo os modelos TGS-2600, TGS-2602, TGS-2610, TGS-2620, sendo que o sistema possui quatro sensores de cada tipo. Os sensores foram dispostos em um compartimento com volume de 60 ml, onde foi injetado um fluxo constante de gás de 300 ml/min, e as amostras foram feitas em uma frequência de 100 Hz ininterruptamente por um período de 12 horas (FONOLLOSA et al., 2015).

Cada um dos modelos de sensores utilizados tem aplicação e sensibilidade a diferentes gases. Os sensores TGS-2600 e TGS-2602 são utilizados para detecção de contaminantes no ar como etanol, amônia, CO e hidrogênio (FIGARO USA INC., 2015 e 2013). O modelo TGS-2610 é sensível a gases liquefeitos de petróleo como propano, iso-butano e metano (FIGARO USA INC., 2017). Já o TGS-2620 é geralmente utilizado para detecção de gases e vapores de solventes como etanol (FIGARO USA INC., 2014).

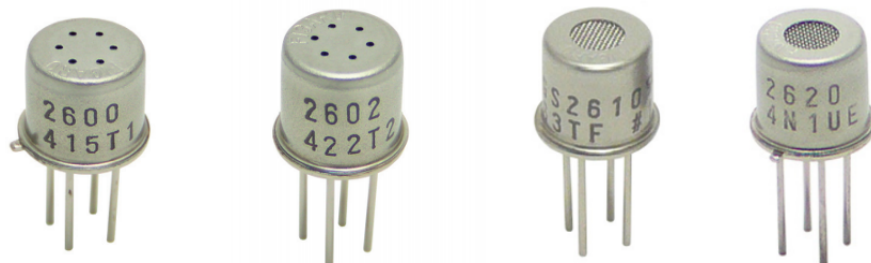


Figura 6: Da esquerda para direita: sensores TGS-2600, TGS-2602, TGS-2610, TGS-2620 (imagem adaptada da ficha de especificação dos componentes)

Na primeira hora de amostras os gases nunca se sobrepõem, isto é, pelo menos um dos gases sempre está com concentração igual a zero. A partir da segunda hora, os

gases começam a se sobrepor, porém os gases não variam simultaneamente em nenhum momento. As figuras 7 e 8 mostram o comportamento dos gases na primeira e segunda hora, respectivamente.

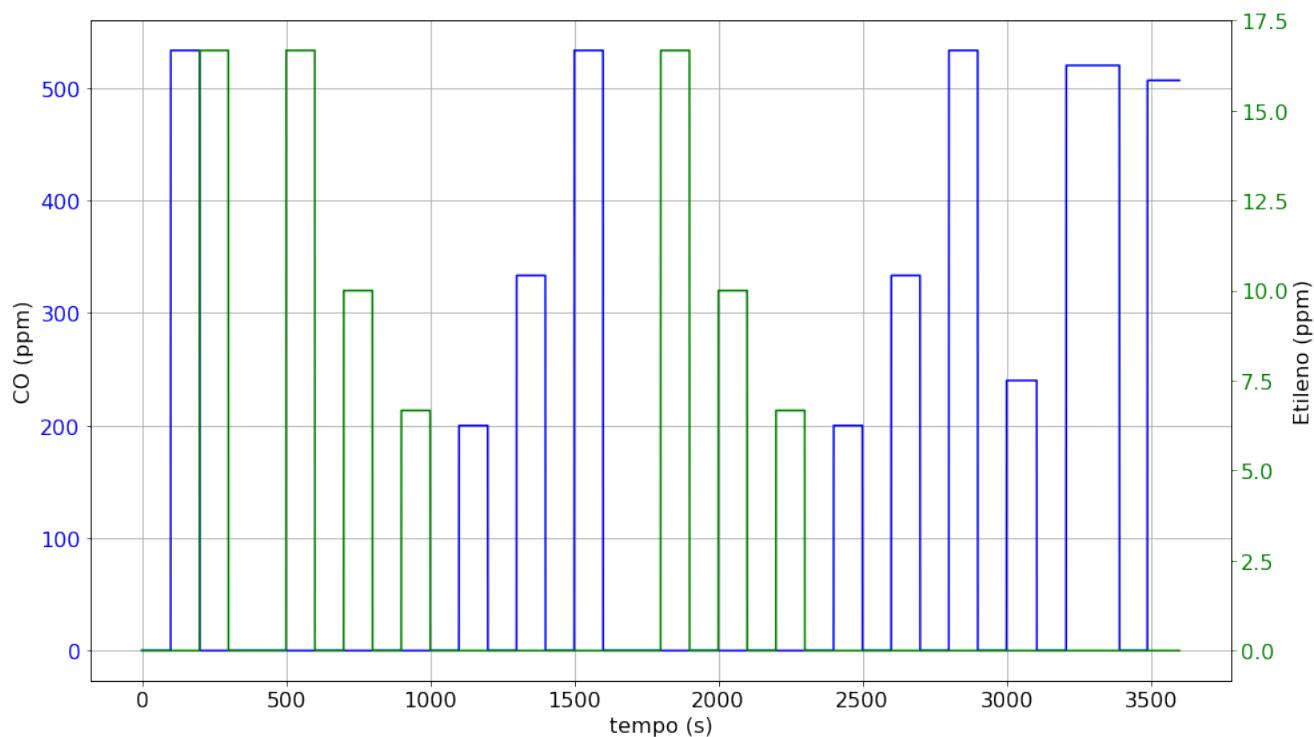


Figura 7: Concentração de CO e Etileno na primeira hora

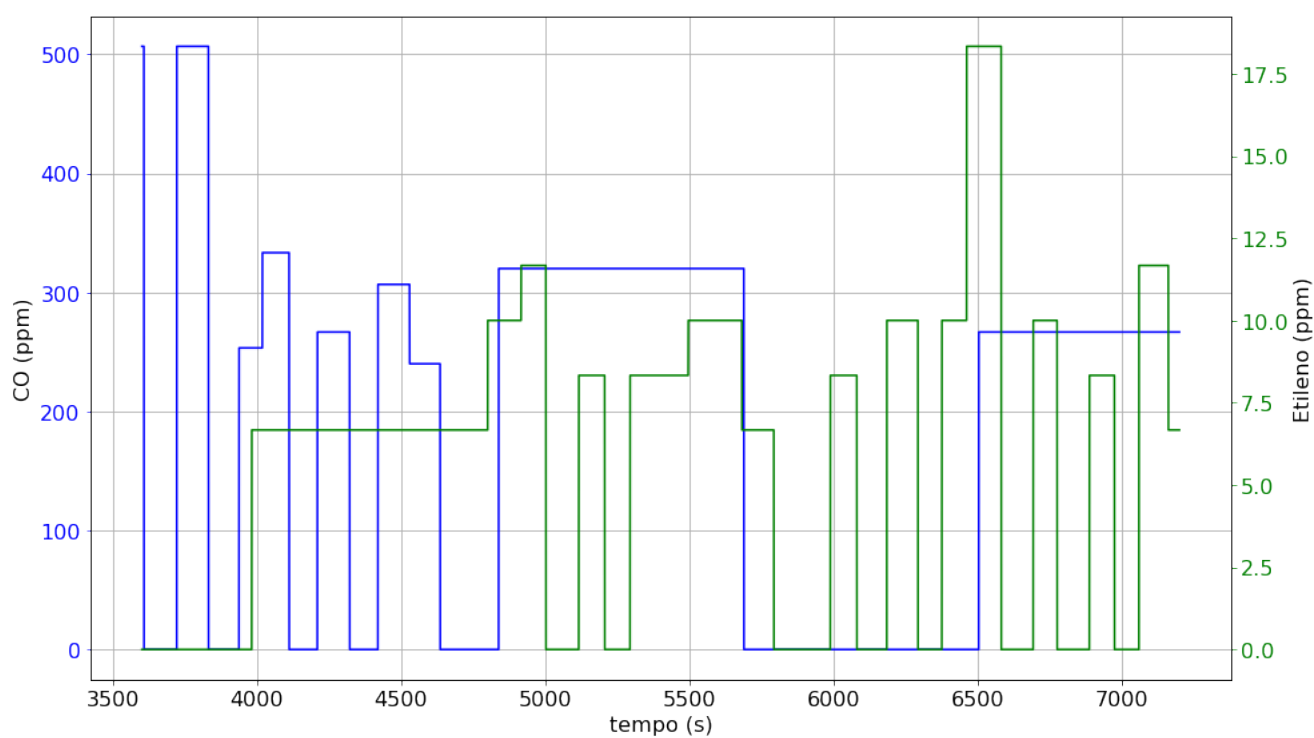


Figura 8: Concentração de CO e Etileno na segunda hora

3.2 Software

Foram utilizadas duas linguagens para implementação dos algoritmos necessários: Python e C++. A primeira foi utilizada em conjunto com as bibliotecas numpy, pandas, matplotlib e sympy para implementação dos algoritmos de aprendizado, parâmetros da rede e visualização de dados por ser uma linguagem de alto nível facilitando a implementação. Enquanto a segunda foi utilizada com a framework Qt5 para implementação do algoritmo de inferência em redes bayesianas gaussianas descrito na seção 2.4 por ser uma linguagem de alta eficiência.

3.2.1 Python

Com uma sintaxe concisa e com foco em legibilidade de código fonte, estruturas de alto nível nativas e uma grande comunidade, Python é hoje uma das linguagens de programação mais populares existentes.

Python foi lançada por Guido van Rossum em 1991, é uma linguagem de alto nível, interpretada por bytecode em máquina virtual Python e multi-paradigma, suportando programação modular, programação funcional e orientação a objeto (BORGES, 2014).

Numpy é um pacote de Python voltado para computação científica, incluindo vetores multidimensionais de objetos, operações otimizadas com vetores, rotinas de ordenação, operações matemáticas e estatísticas básicas, entre outras funcionalidades (OLIPHANT, 2006). Esse pacote foi utilizado devido sua eficiência em operações com matrizes.

Pandas é uma biblioteca de Python para manipulação de dados, incluindo estruturas e operações de manipulação de tabelas numéricas e séries temporais. Essa biblioteca fornece operações como amostragem, ordenação, filtro e concatenação de conjunto de dados e por isso se torna praticamente indispensável na análise de dados (MCKINNEY, 2012).

Sympy é uma biblioteca para computação simbólica em Python. Usando símbolos essa biblioteca é capaz de resolver equações aritméticas, cálculo e álgebra (MEURER et al., 2017).

Matplotlib é um pacote para criação de gráficos em Python, capaz de gerar desde gráficos de linha, histogramas, gráficos de dispersão e até mesmo gráficos de superfícies. Os gráficos deste trabalho foram criados com esta biblioteca (HUNTER, 2007).

3.2.2 C++

A linguagem de programação C++ é uma linguagem multi-paradigma, podendo ser imperativa, orientada a objetos ou genética, baseada na linguagem C. Idealizada para ser tão eficiente quanto C, porém com suporte a outros paradigmas, C++ pode ser utilizado em ambientes onde os recursos de hardware são extremamente limitados como sistemas embarcados (STROUSTRUP, 2000).

Qt é uma framework para desenvolvimento de aplicações com interface gráfica ou aplicações de console multiplataforma em C++ (ENG, 2016). A versão mais recente neste momento é Qt5, e é utilizado em diversos sistemas embarcados como *smart TVs* e centrais multimídia de automóveis. Alguns exemplos de empresas que usam Qt em seus produtos são: LG, Samsung, Autodesk e Ford.

Esta framework é construída em cima do conceito do *QObject* e *Event Loop*, permitindo programação orientada a eventos em C++. Além disso, possui implementações de estruturas de alto nível como *Hash*, vetores e listas otimizados, e interfaces como serial e rede, aumento o nível da linguagem e permitindo um desenvolvimento acelerado.

3.3 Rede Bayesiana

A metodologia do experimento consiste em três etapas principais: aprendizado dos parâmetros da rede bayesiana gaussiana a partir dos dados, aplicação dos parâmetros em uma rede e finalmente, a inferência dos valores de interesse na rede.

Para aprendizado da rede, o conjunto de dados foi separado em duas partes: treino e teste. O conjunto de teste foi escolhido como sendo as quatro primeiras horas do dado e o conjunto de treino, as oito horas seguintes. E com o conjunto de dados de treino, excluindo-se a coluna de tempo, foi executado o procedimento descrito na seção 2.5 com as 18 colunas restantes, assumindo ignorância prévia.

O procedimento para aprendizado de parâmetros da rede bayesiana gaussiana, nesse caso, consiste dos seguintes passos:

1. Encontrar matriz de precisão numérica a partir do conjunto de dados utilizando a equação 2.43.
2. Encontrar a matriz de precisão simbólica utilizando o algoritmo descrito na figura 5 para n igual a 18.
3. Igualar a matriz simbólica à matriz numérica.
4. Resolver o sistema de equações construído a partir da igualdade.

A execução do procedimento foi feito com *script* em Python, durante o qual, cada variável aleatória da tabela foi nomeada como X_i , onde o índice i varia de 1 a 18, conforme consta na tabela 2. Ao fim do procedimento de aprendizado foram obtidos os parâmetros média atualizada μ_i^* , variância atualizada σ_i^{2*} e coeficiente de correlação entre as variáveis atualizada b_{ij}^* , onde j é menor do que i .

Tabela 2: Relação entre variável e índice para algoritmo de aprendizado de parâmetros

Descrição	Nome	Descrição	Nome
CO	X_1	TGS-2620-2	X_{10}
Etileno	X_2	TGS-2602-3	X_{11}
TGS-2602-1	X_3	TGS-2602-4	X_{12}
TGS-2602-2	X_4	TGS-2600-3	X_{13}
TGS-2600-1	X_5	TGS-2600-4	X_{14}
TGS-2600-2	X_6	TGS-2610-3	X_{15}
TGS-2610-1	X_7	TGS-2610-4	X_{16}
TGS-2610-2	X_8	TGS-2620-3	X_{17}
TGS-2620-1	X_9	TGS-2620-4	X_{18}

Como discutido na seção anterior, existem duas formas de modelar um variável aleatória contínua em uma rede bayesiana gaussiana, a primeira é uma variável onde a média é dada por uma média própria somada a influência dos vértices pais, e a segunda, uma variável onde a média é dada estritamente pelos vértices pais, mas, que tem um vértice pai auxiliar que representa a média da variável aleatória.

Considerando essas duas abordagens, são propostas duas estruturas. A primeira, chamada de Estrutura A, ilustrada na figura 9, representando a primeira abordagem e segue a nomenclatura da tabela 2.

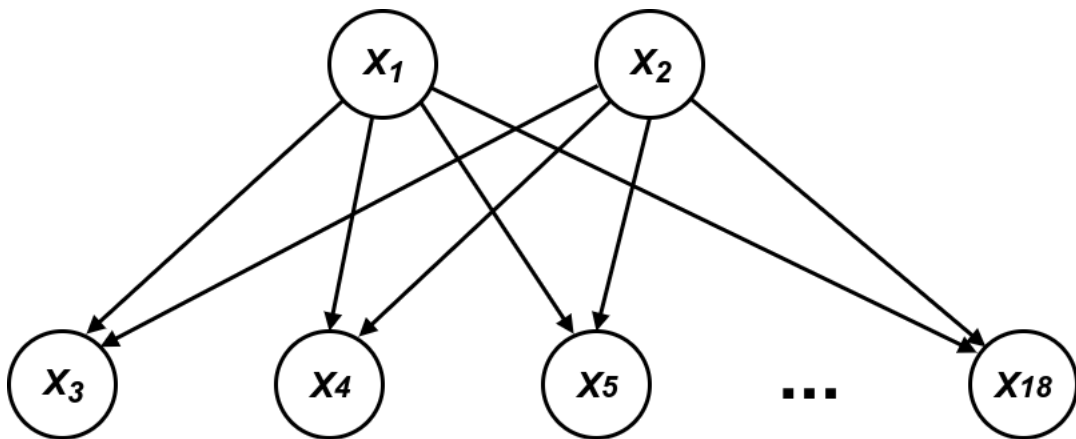


Figura 9: Rede bayesiana gaussiana para inferência de concentração em mistura de gases: Estrutura A

A segunda estrutura proposta, chamada de Estrutura B, ilustrada na figura 10 leva em consideração a segunda abordagem. Neste caso, são necessários a inclusão de mais 16 vértices que tem função de vértices auxiliares contendo a média dos 16 vértices dos sensores. A nova nomenclatura é especificada na tabela 3.

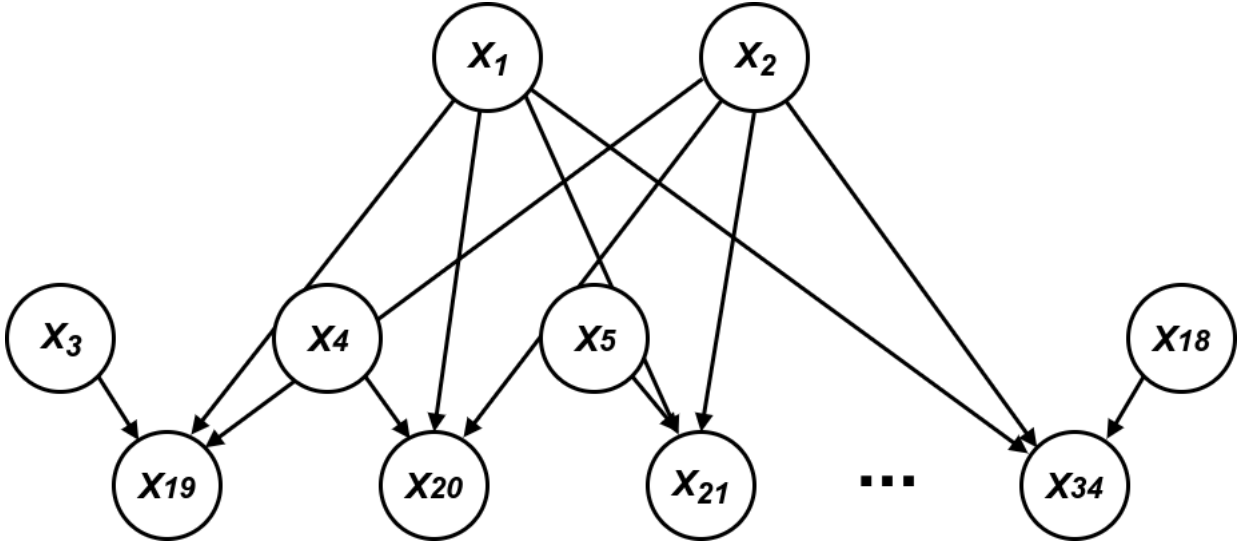


Figura 10: Rede bayesiana gaussiana para inferência de concentração em mistura de gases: Estrutura B

Tabela 3: Relação entre variável e índice para Estrutura B

Descrição	Nome	Descrição	Nome
CO	X_1	$\mu_{TGS-2620-4}$	X_{18}
Etileno	X_2	TGS-2602-1	X_{19}
$\mu_{TGS-2602-1}$	X_3	TGS-2602-2	X_{20}
$\mu_{TGS-2602-2}$	X_4	TGS-2600-1	X_{21}
$\mu_{TGS-2600-1}$	X_5	TGS-2600-2	X_{22}
$\mu_{TGS-2600-2}$	X_6	TGS-2610-1	X_{23}
$\mu_{TGS-2610-1}$	X_7	TGS-2610-2	X_{24}
$\mu_{TGS-2610-2}$	X_8	TGS-2620-1	X_{25}
$\mu_{TGS-2620-1}$	X_9	TGS-2620-2	X_{26}
$\mu_{TGS-2620-2}$	X_{10}	TGS-2602-3	X_{27}
$\mu_{TGS-2602-3}$	X_{11}	TGS-2602-4	X_{28}
$\mu_{TGS-2602-4}$	X_{12}	TGS-2600-3	X_{29}
$\mu_{TGS-2600-3}$	X_{13}	TGS-2600-4	X_{30}
$\mu_{TGS-2600-4}$	X_{14}	TGS-2610-3	X_{31}
$\mu_{TGS-2610-3}$	X_{15}	TGS-2610-4	X_{32}
$\mu_{TGS-2610-4}$	X_{16}	TGS-2620-3	X_{33}
$\mu_{TGS-2620-3}$	X_{17}	TGS-2620-4	X_{34}

Utilizando programa para inferência desenvolvido em C++, é feito a inferência da concentração dos gases CO e Etileno no conjunto de dados de treino e de teste para cada uma das estruturas utilizando todos os 16 sensores como evidência.

Para comparar a eficiência do modelo, é calculado o desvio médio absoluto dos valores inferidos contra os valores reais tanto no conjunto de treino quanto no conjunto de

teste. Ainda, é calculado o desvio médio absoluto da média dos valores reais do conjunto de treino contra os valores reais, para comparação.

Um modelo que prevê sempre a média do conjunto de dados utilizado para treino é a regressão mais simples possível, e por isso é utilizada como base de comparação com o modelo obtido.

4 RESULTADOS

Neste capítulo serão apresentados os resultados obtidos nos experimentos descritos no capítulo anterior, passando pelas duas estruturas propostas.

4.1 Aprendizado

Ao final do método de aprendizado de parâmetros, obteve-se o vetor de média atualizada μ^* e o vetor de variância atualizada σ^{2*} , como apresentado na tabela 4. Também obteve-se os coeficientes de correlação atualizados b_{ij}^* , conforme as tabelas 5, 6 e 7.

Tabela 4: Médias μ_i^* e variâncias σ_i^{2*} atualizadas

i	μ_i^*	σ_i^{2*}	i	μ_i^*	σ_i^{2*}	i	μ_i^*	σ_i^{2*}
1	1.110E+02	1.355E+03	7	1.789E+03	2.888E+02	13	4.948E+03	3.657E+02
2	5.177E+00	1.840E+00	8	2.094E+03	2.356E+01	14	3.942E+03	3.893E+01
3	1.918E+03	1.882E+04	9	5.001E+03	9.234E+02	15	8.643E+02	1.824E+01
4	1.518E+02	7.428E+04	10	5.310E+03	6.178E+01	16	9.817E+02	7.188E+00
5	4.329E+03	7.459E+04	11	1.300E+03	2.994E+02	17	5.379E+03	1.271E+02
6	4.738E+03	6.790E+01	12	1.295E+03	2.756E+01	18	4.328E+03	6.355E+01

Tabela 5: Coeficientes de correlação atualizados b_{ij}^* (colunas 1 à 6)

$b_{i,j}^*$	1	2	3	4	5	6
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
2	-2.33E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
3	3.08E+00	1.62E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4	-3.88E-01	1.56E+01	-2.70E-02	0.00E+00	0.00E+00	0.00E+00
5	1.17E-01	1.70E+02	8.36E-01	8.23E-02	0.00E+00	0.00E+00
6	7.48E-02	-3.12E-01	8.10E-02	2.10E-04	1.06E+00	0.00E+00
7	4.30E-01	3.45E+00	1.52E-01	2.02E-04	1.58E-01	2.31E-02
8	6.98E-02	-1.63E+00	1.20E-02	-1.54E-04	4.62E-02	-1.19E-01
9	-8.61E-02	2.00E+01	-1.87E-01	-2.23E-03	2.12E+00	-1.27E+00
10	4.21E-02	1.48E+00	1.21E-02	4.23E-04	-2.39E-02	1.20E-02
11	1.72E-01	2.81E-01	4.58E-01	-2.51E-03	-1.62E+00	1.41E+00
12	-1.12E-02	-9.98E-01	2.10E-03	2.56E-04	3.13E-02	5.21E-03
13	1.49E-01	-3.05E+00	-8.17E-02	-2.52E-03	-9.97E-02	1.30E+00
14	-2.73E-03	-6.55E-01	4.91E-02	3.06E-04	3.13E-01	-4.10E-01
15	-8.42E-03	1.09E-01	2.29E-02	2.51E-04	2.98E-01	-2.37E-01
16	-6.20E-03	-2.10E-01	-5.07E-03	-1.56E-04	1.31E-01	-1.32E-01
17	6.61E-02	2.74E+00	6.94E-03	-4.30E-05	2.74E-01	-1.13E+00
18	5.82E-02	-1.16E-01	-6.40E-03	-5.47E-04	-5.37E-01	7.37E-03

Tabela 6: Coeficientes de correlação atualizados b_{ij}^* (colunas 7 à 12)

$b_{i,j}^*$	7	8	9	10	11	12
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
5	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
6	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
8	1.24E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
9	2.68E+00	-7.86E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00
10	-6.20E-02	-7.91E-02	1.09E+00	0.00E+00	0.00E+00	0.00E+00
11	-5.70E-01	-2.45E-01	-3.51E-01	5.61E-01	0.00E+00	0.00E+00
12	-1.52E-01	9.35E-02	-1.15E-01	9.24E-02	1.01E+00	0.00E+00
13	-3.20E+00	1.54E+00	-6.54E-01	7.48E-01	-3.23E-01	7.42E-01
14	-4.59E-01	7.30E-01	8.41E-02	6.65E-02	-4.52E-02	7.37E-02
15	7.27E-01	-2.25E-01	4.90E-02	-1.04E-01	1.58E-01	-1.77E-01
16	2.33E-01	-2.54E-02	3.28E-02	-4.51E-02	1.07E-01	-8.27E-02
17	4.28E-01	-1.07E+00	-2.23E-01	1.06E+00	4.25E-01	-6.31E-01
18	-1.65E-01	1.39E+00	3.23E-01	2.63E-02	-1.51E-01	2.35E-01

Tabela 7: Coeficientes de correlação atualizados b_{ij}^* (colunas 13 à 18)

$b_{i,j}^*$	13	14	15	16	17	18
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
5	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
6	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
9	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
10	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
12	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
13	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
14	6.96E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
15	-2.45E-02	-1.05E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
16	3.02E-02	-1.63E-02	6.54E-01	0.00E+00	0.00E+00	0.00E+00
17	8.88E-01	3.00E-01	-6.95E-02	3.85E-01	0.00E+00	0.00E+00
18	8.21E-02	6.09E-01	-1.66E+00	2.46E-01	2.46E-01	0.00E+00

4.2 Estrutura A

No aprendizado de parâmetro obteve-se uma tabela de coeficientes de correlação contendo mais relações do que previsto pela Estrutura A, ilustrada pela figura 9. Todos os coeficientes não previstos pelo modelo foram ignorados.

As figuras 11 e 12 mostram, respectivamente, o resultado da inferência dos gases CO e Etileno durante a primeira hora do conjunto de treino (quinta hora absoluta) contra os dados reais. O resultado do cálculo do DMA do conjunto de treino contra a inferência da estrutura A, DMA do conjunto de treino contra a média, bem como o percentual de melhora do erro da Estrutura A em relação a média é descrito na tabela 8.

Já as figuras 13 e 14 mostram, respectivamente, o resultado da inferência dos gases CO e Etileno durante a segunda hora do conjunto de teste contra os dados reais. Os resultados das métricas para o conjunto de teste estão na tabela 9.

Tanto no conjunto de treino quanto de teste, o modelo obteve um desempenho ruim na inferência da concentração do monóxido de carbono, enquanto para o etileno o modelo foi capaz de reproduzir o comportamento da concentração desse gás, com um erro 30,41% menor do que erro para média do valor das concentrações no conjunto de treino.

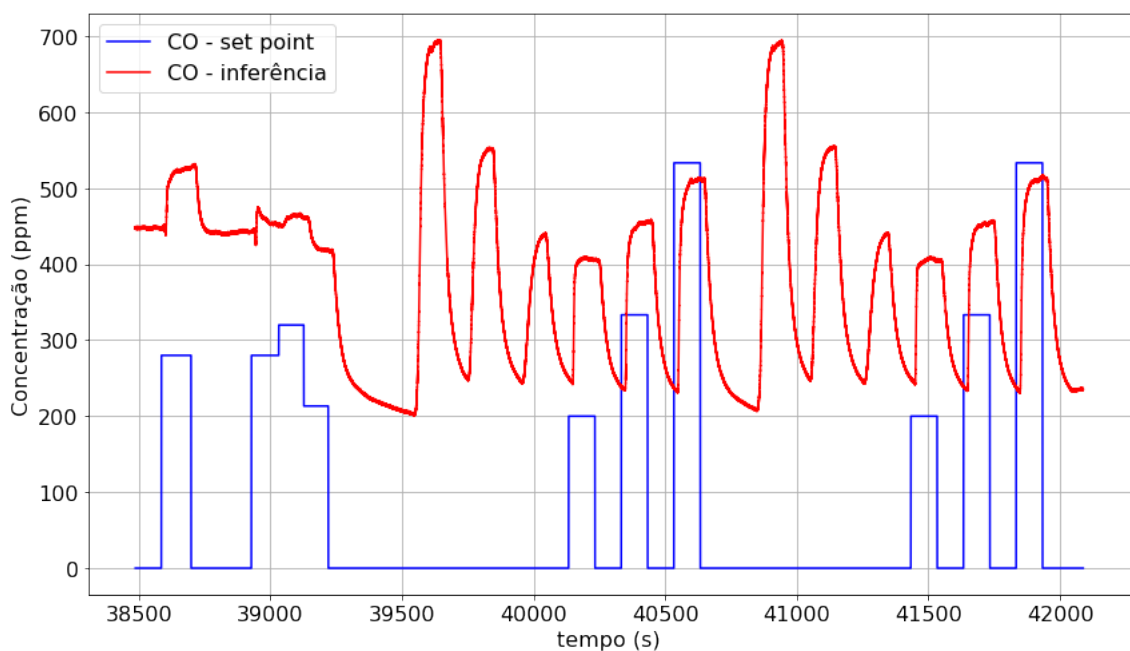


Figura 11: Inferência e *set point* para CO com Estrutura A em conjunto de treino

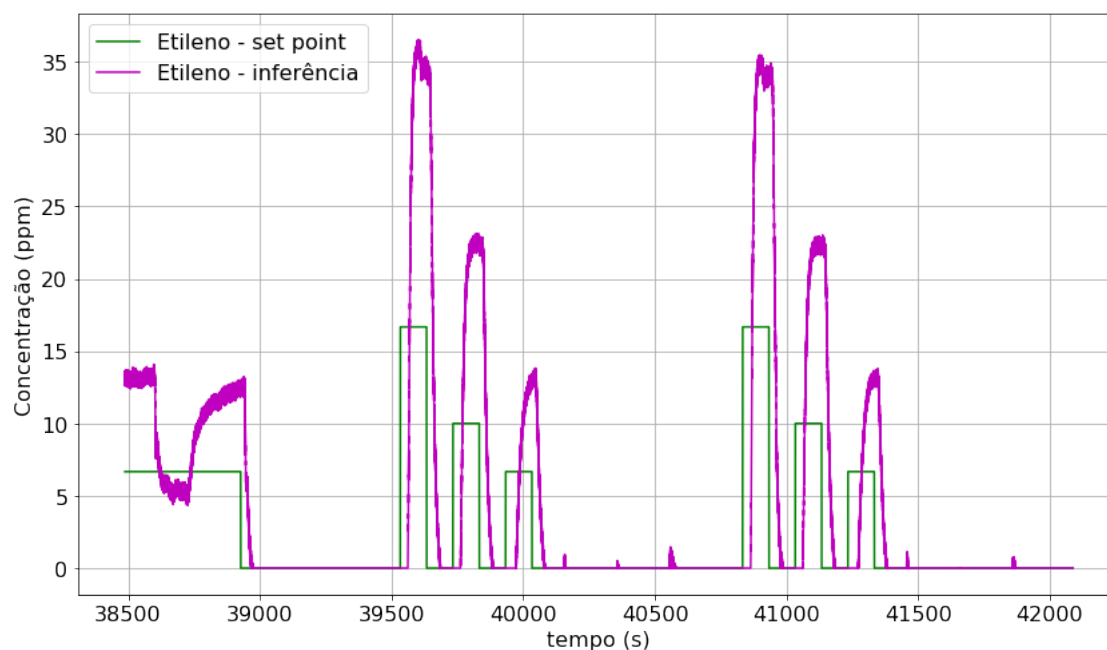


Figura 12: Inferência e *set point* para Etileno com Estrutura A em conjunto de treino

Tabela 8: Comparação do desvio médio absoluto do conjunto de treino com inferência e média dos dados (Estrutura A)

Gás	DMA inferência (ppm)	DMA média (ppm)	Percentual de melhora
Etileno	4.1937	5.3059	20.96%
CO	339.6126	140.8641	-141.09%

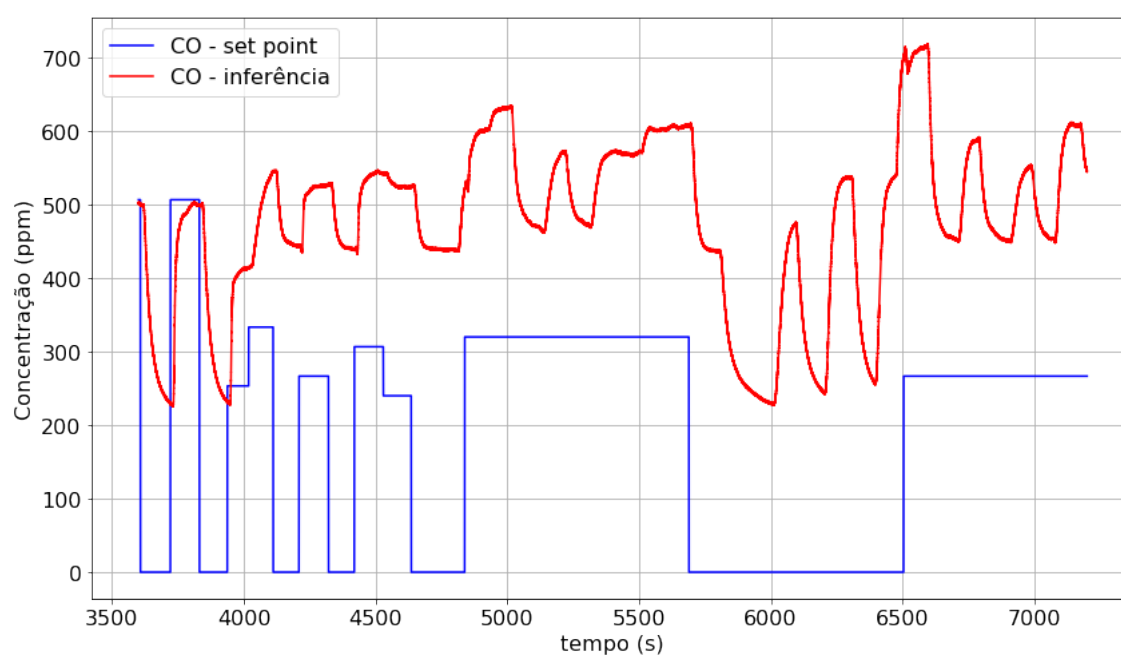


Figura 13: Inferência e *set point* para CO com Estrutura A em conjunto de teste

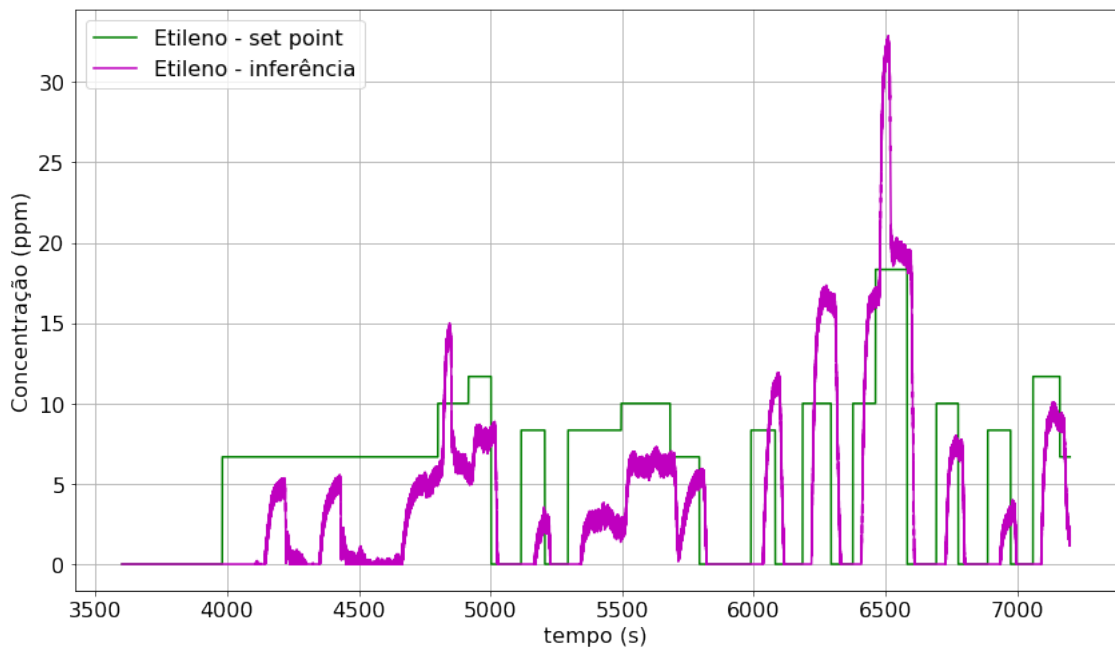


Figura 14: Inferência e *set point* para Etileno com Estrutura A em conjunto de teste

Tabela 9: Comparação do desvio médio absoluto do conjunto de teste com inferência e média dos dados (Estrutura A)

Gás	DMA inferência (ppm)	DMA média (ppm)	Percentual de melhora
Etileno	3.4009	4.8869	30.41%
CO	320.9922	153.7820	-108.73%

4.3 Estrutura B

Assim como na primeira estrutura, todos os coeficientes não previstos pela Estrutura B, ilustrada pela figura 10, foram ignorados. Neste caso os coeficientes de correlação entre os gases e os sensores foram mantidos, os vértices dos sensores mantiveram sua variância porém sua média foi modificada para zero, os vértices auxiliares recebem variância zero a média de seus respectivos vértices de sensores. Ainda, a coeficiente entre o vértice auxiliar e seu respectivo vértice de sensor é igual a um.

As figuras 15 e 16 mostram, respectivamente, o resultado da inferência dos gases CO e Etileno durante a primeira hora do conjunto de treino (quinta hora absoluta) contra os dados reais. Os resultados das métricas para o conjunto de teste estão na tabela 10. Ainda, as figuras 17 e 18 mostram, respectivamente, o resultado da inferência dos gases CO e Etileno durante a segunda hora do conjunto de teste contra os dados reais, e as métricas para o conjunto de teste estão na tabela 11.

Neste modelo obteve-se um resultado expressivo na inferência do monóxido de carbono, obtendo um erro 39,72% e 23,63% menor do que a média do conjunto de treino

na inferência do conjunto de treino e de teste respectivamente. Porém, apresentou um desempenho bastante ruim para inferência da concentração de etileno.

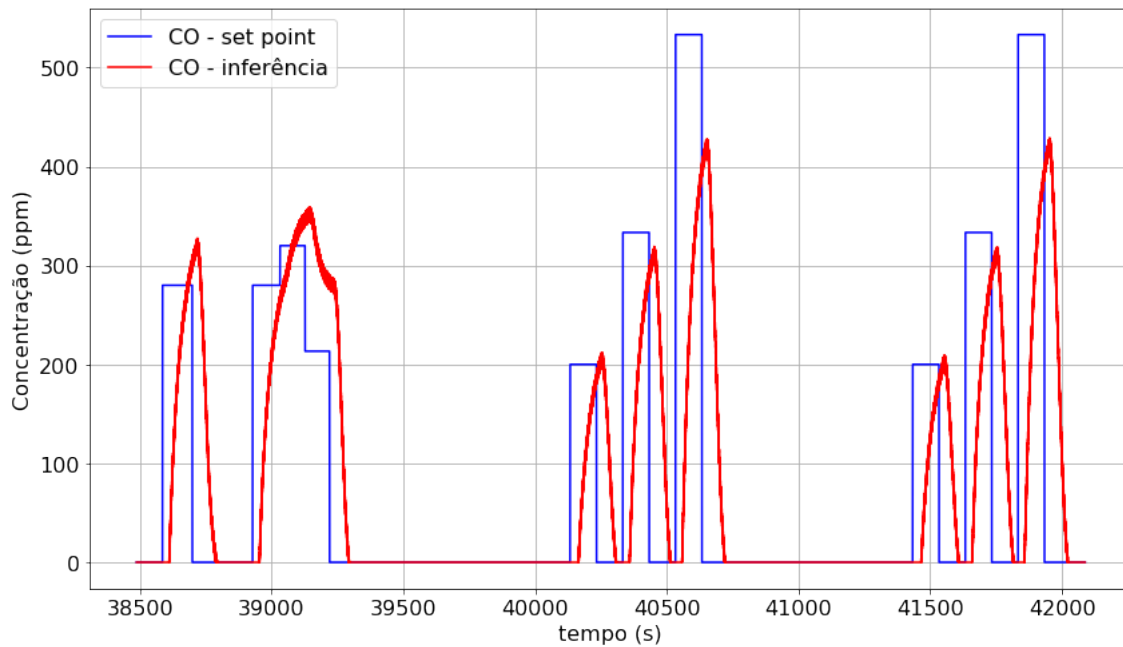


Figura 15: Inferência e *set point* para CO com Estrutura B em conjunto de treino

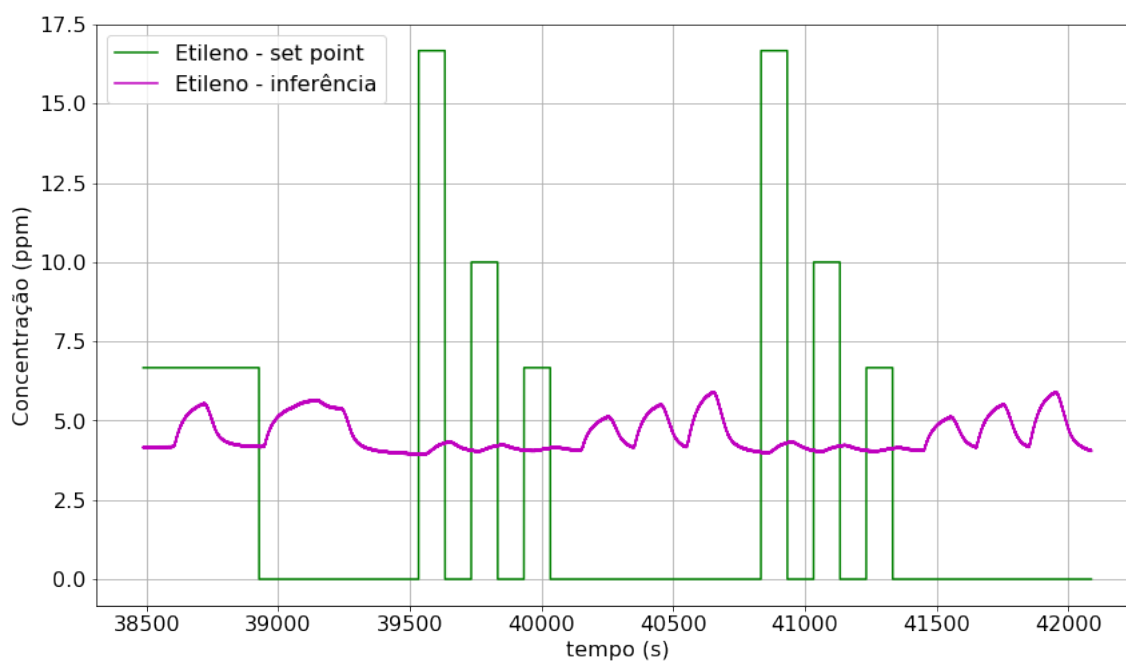


Figura 16: Inferência e *set point* para Etileno com Estrutura B em conjunto de treino

Tabela 10: Comparação do desvio médio absoluto do conjunto de treino com inferência e média dos dados (Estrutura B)

Gás	DMA inferência (ppm)	DMA média (ppm)	Percentual de melhora
Etileno	5.2143	5.3059	1.73%
CO	84.9102	140.8641	39.72%

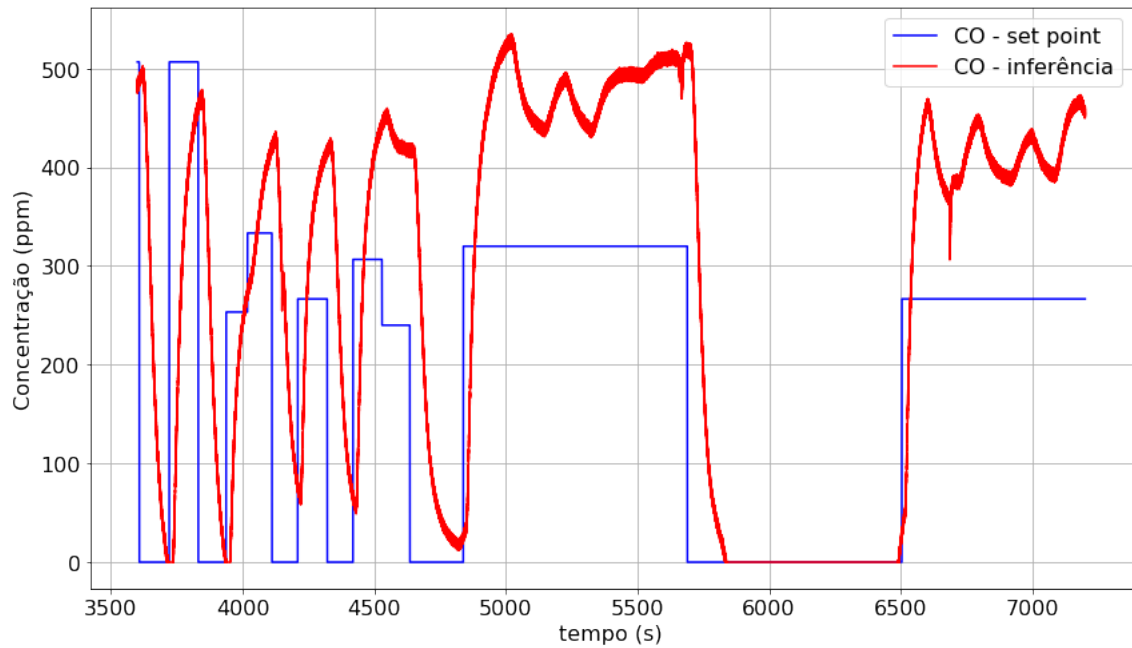


Figura 17: Inferência e *set point* para CO com Estrutura B em conjunto de teste

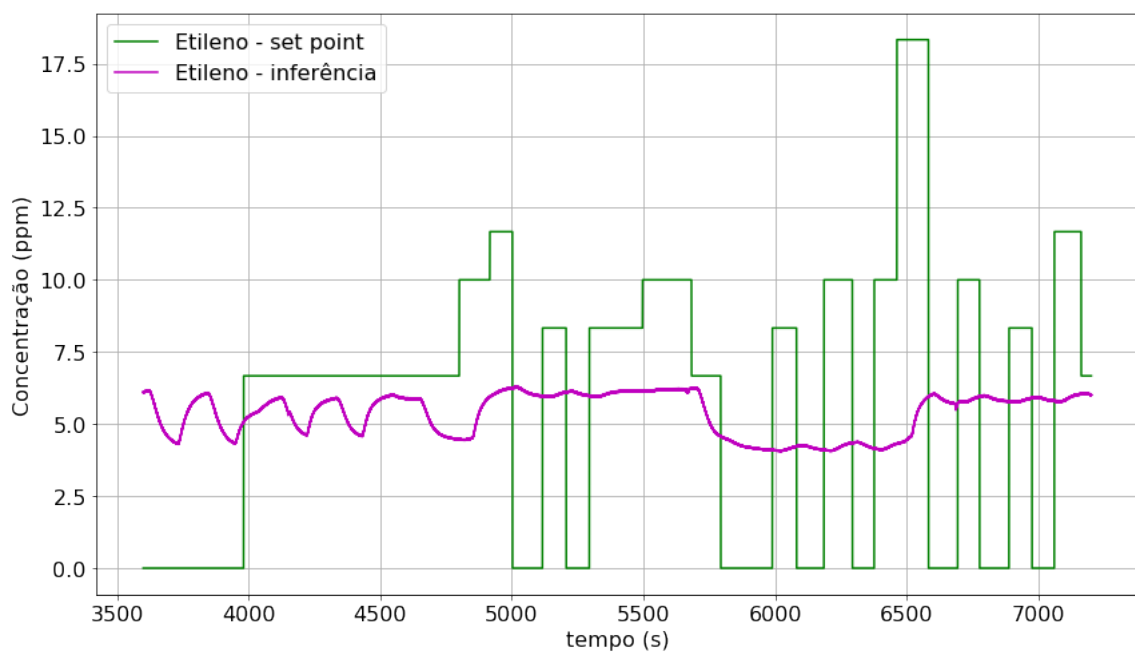


Figura 18: Inferência e *set point* para Etileno com Estrutura B em conjunto de teste

Tabela 11: Comparação do desvio médio absoluto do conjunto de teste com inferência e média dos dados (Estrutura B)

Gás	DMA inferência (ppm)	DMA média (ppm)	Percentual de melhora
Etileno	4.7477	4.8869	2.85%
CO	117.4392	153.7820	23.63%

5 CONCLUSÕES

Nos resultados obtidos no treinamento, pode-se observar os parâmetros da distribuição gaussiana encontrada para cada um dos nós e seus coeficientes de correlação, que ilustram a característica de Redes Bayesianas possuírem parâmetros interpretáveis. Observa-se, a partir da tabela de coeficientes de correlação, o quão sensível cada sensor é a cada um dos gases da mistura, por exemplo, nota-se que sensores correspondentes a variáveis aleatórias X_{15} e X_{16} , ambos do modelo TGS-2610, tem coeficiente da ordem de 10^{-3} em relação a concentração de CO, a menor correlação entre sensores e gases da tabela 5, implicando uma baixa influência da concentração de CO no valor destes sensores.

Ainda, observa-se que os sensores correspondentes as variáveis aleatórias X_7 e X_8 , os dois também do modelo TGS-2610, apresentam coeficiente de correlação da ordem de 10^{-1} e 10^{-2} em relação a concentração de CO. O mesmo fenômeno pode ser observado entre os outros pares de sensores. Assim, a diferença entre os coeficientes encontrados para sensores do mesmo modelo sugere a existência de uma dependência espacial no valor medido pelo sensor.

Dentre as estruturas A e B, construídas a partir da relação causal das variáveis e com parâmetros obtidos a partir do algoritmo de aprendizado descrito na seção 2, nenhuma das duas obteve sucesso na inferência de ambos os gases simultaneamente. Cada estrutura obteve sucesso relativo na inferência da concentração de um dos gases, enquanto obteve desempenho péssimo na inferência da concentração do outro gás.

A estrutura A obteve péssimo desempenho na inferência da concentração de CO, com um desvio médio absoluto no conjunto de treino e de teste de 339 ppm e 320 ppm, respectivamente, enquanto simplesmente utilizar a média do conjunto de dados como resultado da inferência obtém um desempenho bastante superior, com desvio médio absoluto de 140 ppm e 153 ppm, para conjunto de treino e de teste, respectivamente. Porém, na inferência da concentração de etileno, a estrutura obteve uma melhora de 20.96% e 30.41% em relação a média, para os conjuntos de treino e teste, respectivamente.

A estrutura B obteve um desempenho razoável na inferência da concentração de CO, com uma melhora de 39.72% e 23.63% no desvio médio absoluto em relação a média, para os conjuntos de treino e de teste, respectivamente. Porém, a estrutura não foi capaz de obter um desempenho semelhante para inferência da concentração de etileno, tendo um desvio médio absoluto de 5.2 ppm e 4.7 ppm para os conjuntos de treino e de teste, respectivamente, sendo os valores muito próximos do resultado obtido para inferência com média.

Embora as estruturas sejam baseadas nos mesmos parâmetros, os resultados foram

bastante diferentes para cada uma delas. Essa divergência pode ser explicada pelo algoritmo de inferência em redes bayesianas gaussianas utilizado neste trabalho, ter como exigência que o grafo que forma a estrutura seja individualmente conectado. Apesar dessa exigência, as redes causais possíveis para esse problema não são capazes de satisfazer a condição, causando comportamento inesperado nos resultados da inferência. A escolha de outro algoritmo de inferência pode melhorar os resultados, porém não existem muitas bibliotecas gratuitas com inferência para variáveis contínuas completamente implementadas devido sua complexidade.

Também é notável que a resposta de inferência da concentração dos gases apresenta certo atraso em relação a concentração real. Esse comportamento surge devido ao atraso da resposta dos próprios sensores de gases frente a uma variação abrupta da concentração dos gases a que estão expostos. O modelo estático proposto não consegue contornar esse problema, uma vez que sempre considera aquele instante isolado do histórico da concentração dos gases. O modelo ideal para este caso seria um modelo dinâmico, que é capaz de considerar os últimos valores adquiridos em uma série temporal e prever as concentrações com maior acurácia mesmo enfrentando variações abruptas das concentrações de interesse e sensores com respostas lentas.

Visto que o algoritmo de inferência não é adequado para as estruturas causais propostas, os trabalhos futuros podem conter uma pesquisa sobre outros algoritmos de inferências para redes bayesianas contínuas na literatura e sua possível implementação para comparação com os resultados obtidos neste trabalho. Ainda, pode-se discretizar o problema e fazer uso de redes bayesianas discretas, para comparação dos resultados das diferentes abordagens.

REFERÊNCIAS

- Bermak, A.; Belhouari, S. B. Bayesian learning using gaussian process for gas identification. **IEEE Transactions on Instrumentation and Measurement**, v. 55, n. 3, p. 787–792, June 2006.
- BORGES, L. E. **Python para desenvolvedores: aborda Python 3.3**. [S.l.]: Novatec Editora, 2014.
- BURRELL, G. A.; SEIBERT, F. M. Experiments with small animals and carbon monoxide. **Industrial & Engineering Chemistry**, ACS Publications, v. 6, n. 3, p. 241–244, 1914.
- COOPER, G. F. An overview of the representation and discovery of causal relationships using bayesian networks. **Computation, causation, and discovery**, AAAI Press/MIT Press Cambridge, MA, p. 4–62, 1999.
- DEGROOT, M. H. **Optimal statistical decisions**. [S.l.]: John Wiley & Sons, 2005. v. 82.
- ENG, L. Z. **Qt5 C++ GUI Programming Cookbook**. [S.l.]: Packt Publishing Ltd, 2016.
- FIGARO USA INC. **TGS 2600 - for the detection of Air Contaminants**. [S.l.], 2013.
- FIGARO USA INC. **TGS 2620 - for the detection of Solvent Vapors**. [S.l.], 2014.
- FIGARO USA INC. **TGS 2602 - for the detection of Air Contaminants**. [S.l.], 2015.
- FIGARO USA INC. **TGS 2610-C00 - for the detection of LP Gas**. [S.l.], 2017.
- FONOLLOSA, J. et al. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. **Sensors and Actuators B: Chemical**, Elsevier, v. 215, p. 618–629, 2015. Dados disponíveis em: <<https://archive.ics.uci.edu/ml/datasets/Gas+sensor+array+under+dynamic+gas+mixtures>>.
- HUNTER, J. D. Matplotlib: A 2d graphics environment. **Computing in science & engineering**, IEEE Computer Society, v. 9, n. 3, p. 90, 2007.
- LIU, Y.; ZHOU, K.; LEI, Y. Using bayesian inference framework towards identifying gas species and concentration from high temperature resistive sensor array data. **Journal of Sensors**, Hindawi, v. 2015, 2015.
- MCKINNEY, W. **Python for data analysis: Data wrangling with Pandas, NumPy, and IPython**. [S.l.]: "O'Reilly Media, Inc.", 2012.
- MEURER, A. et al. Sympy: symbolic computing in python. **PeerJ Computer Science**, PeerJ Inc., v. 3, p. e103, 2017.

NEAPOLITAN, R. E. **Learning Bayesian Networks**. New Jersey: Pearson Prentice Hall, 2004.

OLIPHANT, T. E. **A guide to NumPy**. [S.l.]: Trelgol Publishing USA, 2006. v. 1.

SHACHTER, R. D.; KENLEY, C. R. Gaussian influence diagrams. **Management science**, INFORMS, v. 35, n. 5, p. 527–550, 1989.

STROUSTRUP, B. **The C++ programming language**. [S.l.]: Pearson Education India, 2000.

SUNDGREN, H. et al. Artificial neural networks and gas sensor arrays: quantification of individual components in a gas mixture. **Measurement Science and Technology**, IOP Publishing, v. 2, n. 5, p. 464, 1991.

WALPOLE, R. E. **Probability and Statistics for Engineers and Scientists**. 8. ed. [S.l.]: Pearson Prentice Hall, 2006. ISBN 0131877119,9780131877115.

WEST, D. B. **Introduction to Graph Theory**. 2. ed. [S.l.]: Pearson Prentice Hall, 2000.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. **Climate research**, v. 30, n. 1, p. 79–82, 2005.

YUNUSA, Z. et al. Gas sensors: A review. **Sensors and Transducers**, v. 168, p. 61–75, 04 2014.