

ANÁLISE DE DADOS DE LÍNGUA ELETRÔNICA BASEADA EM ALGORITMOS FUZZY

L. P. O. SOUSA*, D. D. FERREIRA*, K. L. FUKUSHIMA*, J. E. OLIVEIRA*

**Departamento de Engenharia
Universidade Federal de Lavras
Lavras, Minas Gerias, Brasil*

Emails: lpsousa14@gmail.com, danton@deg.ufla.br, katiafuku@gmail.com,
juliano.oliveira@deg.ufla.br

Abstract— The design of devices that resemble biological recognition systems, inspired by the functioning of human organs, have made important advances in recent decades. As a result, interest in the concept of electronic languages has increased considerably in recent years. This study uses this type of device in order to identify different concentrations of Glutamate samples. The applied methodology involves algorithms of Fuzzy logic, where the tools used are unusual in this area. Among the methods applied, the best results reached 66% accuracy.

Keywords— Fuzzy Logic, E-tongue e Clustering.

Resumo— O projeto de dispositivos que se assemelham a sistemas de reconhecimento biológico, inspirados no funcionamento de órgãos humanos, tiveram avanços importantes nas últimas décadas. Em consequência disso, o interesse no conceito de línguas eletrônicas aumentou consideravelmente nos últimos anos. Este estudo utiliza esse tipo de dispositivo com o intuito de identificar diferentes concentrações de amostras de Glutamato. A metodologia aplicada envolve algoritmos de lógica Fuzzy, onde as ferramentas empregadas são pouco usuais nessa área. Dentre os métodos aplicados, os melhores resultados atingiram 66% de precisão.

Palavras-chave— Lógica Fuzzy, Língua Eletrônica e Clusterização.

1 Introdução

Os primeiros sistemas inspirados nos órgãos olfativos e gustativos foram propostos na década de 1980, a partir disso o desejo de modelar e melhorar esses sistemas despertou o interesse dos cientistas. O projeto desses dispositivos se assemelha aos sistemas de reconhecimento biológico, onde as técnicas sensoriais e instrumentais são utilizadas para determinar os sabores ou as substâncias presentes nas amostras analisadas. Em consequência disso, o interesse no conceito de línguas eletrônicas (LE) aumentou consideravelmente nos últimos anos (Nery and Kubota, 2016; Wei and Wang, 2014).

Basicamente, as LE são compostas por unidades de coleta de informações, também denominadas de arranjos de sensores, utilizadas em soluções aquosas. Para o processamento dos dados gerados utilizam-se técnicas de reconhecimento de padrões. As informações obtidas pelos sensores (que podem ser definidas como a “impressão digital”) correspondem ao conjunto dos diferentes constituintes da solução em análise e suas informações valiosas (como reconhecimento, identificação ou quantificação), as quais podem ser obtidas através de métodos quimiométricos e de inteligência artificial (Nery and Kubota, 2016; Wei and Wang, 2014).

Ao longo das últimas décadas houve avanços importantes no projeto desses sensores, normalmente direcionados para a implementação de novos conceitos, projetos ou configurações. A LE é

uma tendência recente da área sensorial, mas já existe uma grande quantidade de aplicações envolvendo LE em diversos segmentos, como na indústria alimentar, onde seu uso pode ser aplicado em questões de controle de qualidade, monitoramento de processos, otimização de biorreatores, controle de envelhecimento e controle automatizado de gosto. Na indústria química, pode-se aplicar na detecção de grupos funcionais e avaliação de pureza. Outro segmento que LE podem ser empregadas é na área medicinal, como monitoramento clínico, diagnóstico não-invasivo, avaliação de produtos farmacêuticos e crescimento de culturas celulares e ainda pode contribuir no monitoramento ambiental, como na detecção de resíduos de pesticidas (Cetó et al., 2015; Nery and Kubota, 2016; Wei, Wang and Zhang, 2013).

Ao realizar uma pesquisa rápida na literatura, é fácil observar que métodos como *Principal Component Analysis* (PCA) e *Partial Least Squares* (PLS) são utilizados com bastante frequência em análises de dados de LE. Porém, esses métodos possuem algumas limitações, como:

- O PCA não é a ferramenta mais adequada para ser usada na classificação de objetos, mesmo que seja possível observar uma separação nítida, entretanto há diversos estudos que o utilizam dessa forma. O recomendado é empregar uma técnica supervisionada em paralelo (Zielinski et al., 2014).
- No PLS as classes são divididas usando partições lineares no espaço de classificação que

podem ser problemáticas se a separação entre classes for não linear. Além disso, o ruído aleatório aumenta à medida que mais variáveis latentes são adicionadas (Zielinski et al., 2014).

O objetivo principal dessa pesquisa é elaborar modelos de classificação que sejam capazes de discriminar amostras de diferentes concentrações. Os modelos elaborados utilizam algoritmos Fuzzy de clusterização, o que remete ao uso de metodologias pouco empregadas em análises de dados de LE. Os algoritmos Fuzzy de Clusterização utilizados nesse estudo foram o Fuzzy C-means, Gustafson-Kessel e Gath-Geva, os quais serão detalhados na Seção 3.

O restante do artigo está organizado da seguinte forma: a Seção 2 traz a revisão da literatura com o intuito de exemplificar os estudos que têm sido realizados nessa área de pesquisa. A Seção 3 é destinada a apresentação da configuração do experimento para obtenção dos dados e os detalhes dos métodos usadas. Os resultados e as análises alcançados são expostas na Seção 4. E por fim, na Seção 5, as conclusões do trabalho e os possíveis trabalhos futuros.

2 Estado da Arte

Devido à grande quantidade de aplicações possíveis relacionadas a esse tema, podem ser encontrados diversos estudos realizados na literatura. O estudo apresentado em (Cetó et al., 2015) relata a aplicação de outra LE como ferramenta para discriminação de diferentes origens de vinhos. Para tanto, ele utilizou uma série de ferramentas químicas para análise qualitativa e quantitativa e os escores globais alcançados com sucesso tiveram 92,9% da identificação correta para a aplicação qualitativa e um coeficiente de correlação de 0,830 para o quantitativo.

Ainda relacionado a vinhos, (Ouyang et al., 2013) apresenta uma pesquisa com objetivo de classificar amostras de vinho de acordo com seu tempo de envelhecimento (3, 5, 8 e 10 anos). Os resultados globais comprovam que o sistema de LE proposto com as ferramentas de classificação utilizadas poderia ser usado com sucesso na identificação de vinho com diferentes idades marcadas.

Há diversos estudos relacionados a detecção e quantificação de substâncias em diversos tipos de bebidas, como o trabalho proposto em (Cetó and Valle, 2014) que opera duas LE em diferentes arranjos para identificar tipos de cerveja. Além disso, o autor demonstra a criação de modelos que quantificam a quantidade do teor de álcool da cerveja.

A pesquisa feita em (Wei, Wang and Zhang, 2013) usa uma LE para monitorar a qualidade e o tempo de armazenamento de amostras de leite

pasteurizado não selado. As análises foram realizadas usando a voltametria de pulso com duas formas de ondas potenciais: a de retângulo e de escada em três segmentos de frequência: 1 Hz, 10 Hz, e 100 Hz. Os resultados indicam que o dispositivo manuseado é capaz de classificar as amostras com sucesso. Em outro estudo feito por esse pesquisador, ele utiliza a mesma metodologia descrita para discriminar seis categorias de iogurte. Todas as amostras foram separadas de maneira nítida e ainda foi possível determinar propriedades físicas das amostras de iogurte, como o estresse superficial e a viscosidade (Wei, Wang and Jin, 2013).

O trabalho (Lu et al., 2016) aborda uma análise atributiva para a caracterização e quantificação das particularidades do sabor de amostras de arroz (i.e., suavidade, viscosidade, doçura e aroma) empregando uma LE. Os coeficientes de correlação para todos os atributos foram superiores a 0,9 o que demonstra uma boa capacidade preditiva do modelo de análise proposto. Além do mais, os resultados indicaram que esta abordagem pode quantificar com precisão os atributos de sabor e pode ser uma ferramenta eficiente para o processamento de dados em um sistema de LE.

A pesquisa realizada em (Wei and Wang, 2014) propõem o uso de uma LE como instrumento para classificar e prever amostras de mel de diferentes origens florais e geográficas. As origens florais das amostras analisadas foram: acácia, trigo sarraceno e *motherwort* e as amostras referentes ao mel de acácia foram selecionadas de diferentes origens geográficas. As amostras foram classificadas por ambos os sistemas multissensores e os resultados certificam que os dois tipos de língua têm um bom potencial de classificação.

A LE é composta por arranjos de sensores, de modo que cada sensor pode gerar uma grande quantidade de dados para serem analisados. Esses dados são manipulados com ferramentas matemáticas e estatísticas. Os métodos de classificação mais utilizados para caracterização de amostras podem ser divididos em dois grupos:

- O primeiro das técnicas de discriminação de classe que definem delimitadores entre as classes estabelecidas, de maneira que as amostras sejam sempre atribuídas a uma dessas classes (Borràs et al., 2015).
- O segundo grupo corresponde ao uso de técnicas que calculam um modelo separado para cada classe estabelecida, de modo que uma amostra desconhecida pode ser atribuída a essa classe ou rejeitada (Borràs et al., 2015).

3 Materiais e Métodos

3.1 Língua Eletrônica

A teoria da LE se origina de mecanismos do sistema gustativo do ser humano e pode ser conside-

rada como instrumento analítico baseado em uma série de sensores químicos de baixa seletividade com especificidade parcial (sensibilidade cruzada), que com o auxílio de processamentos quimiométricos é capaz de caracterizar amostras líquidas complexas. Se for configurado e/ou treinado (calibrado), a LE é capaz de reconhecer a composição qualitativa e quantitativa de soluções multi-espécies de diferentes naturezas (Wei, Wang and Zhang, 2013; Peris and Escuder-Gilabert, 2013).

As LE desenvolvidas nos últimos anos podem ser classificadas em três tipos: as que são baseadas em eletrodos de espectroscopia potenciométrica, voltamétrica e de impedância (Wei and Wang, 2014; Liu et al., 2013). De modo geral, uma LE é composta por três elementos: (1) uma matriz de sensores químicos com diversos seletivos, (2) um amplificador e gerador de sinal e (3) um *software* com o algoritmo apropriado para processar os sinais e analisar os resultados, como apresentado na Figura 1.

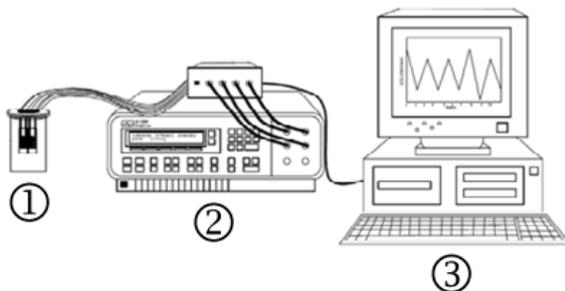


Figura 1: Elementos de uma LE.
Fonte: Adaptado de (Fukushima, 2017).

A LE utilizada nesse estudo, foi desenvolvida em (Fukushima, 2017), onde a mesma tinha como objetivo desenvolver, caracterizar e aplicar sensores nanoestruturados para análise de realçadores de sabor em soluções aquosas. Em sua composição foram utilizados eletrodos interdigitados de ouro fabricados sobre lâminas de vidro modificados através da técnica *Layer-by-Layer*.

Como o intuito do presente artigo é analisar os resultados obtidos pela LE com diferentes ferramentas, os detalhes sobre o preparo e a caracterização das soluções e dos filmes utilizados podem ser encontrados no estudo realizado em (Fukushima, 2017).

3.2 Amostras

As amostras utilizadas no experimento são compostas por Glutamato Monossódico (GM), que é um realçador de sabor muito utilizado em indústrias alimentícias. Realçadores de sabor são aditivos alimentares usados para causar um efeito de gosto e sabor mais marcante nos alimentos ao paladar do consumidor (Milani, 2013).

Esse realçador proporciona o quinto sabor básico, o “Umami”, que em japonês significa sabroso. Além disso, ele também é utilizado pelas indústrias alimentícias para o desenvolvimento de produtos com a redução da concentração de sódio (Fukushima, 2017).

Os modelos de agrupamento propostos têm como objetivo classificar diferentes concentrações da substância em análise. As concentrações que esse estudo pretende classificar estão expostas na Tabela 1.

Tabela 1: Concentrações do realçador.

Concentração	GM (mol L^{-1})
C1	$3,54 \times 10^{-4}$
C2	$3,54 \times 10^{-10}$
C3	$1,00 \times 10^{-6}$

Os dados obtidos pela LE foram capturados usando seis unidades sensoriais conectadas nos canais de leituras, as quais foram feitas utilizando os softwares ZPlot 2.6 e ZView 2.6. Os dados foram coletados a partir das medidas de resistência em função da frequência de 10 Hz a 1 MHz. Mais detalhes relacionados a coleta dos dados podem ser encontradas em (Fukushima, 2017).

Os dados obtidos pela LE serão analisadas através das ferramentas de agrupamentos, *Fuzzy C-means* (FCM), Gustafson-Kessel (GK) e Gath-Geva (GG), os quais são métodos de particionamento difusos com diferentes normas de distância.

A estrutura dos dados gerados pela LE utilizada, consiste em cinco variáveis medidas, agrupadas em um vetor linha $x_k = [x_{k1}, x_{k2}, \dots, x_{k5}]^T$, $x_k \in R^n$. As variáveis obtidas pela LE foram: Frequência em Hertz (F), Impedância real (Zr), Impedância Imaginária (Zi), Capacitância (C) e Resistência (R). O conjunto possui 3672 observações, as quais são denotadas por $X = x_k, k = 1, 2, \dots, 3672$.

Note que as observações obtidas de cada concentração têm as mesmas dimensões, ou seja, 1224 observações de cada. O pré agrupamento realizado foi aplicar os algoritmos nesse banco de dados $A_{3672 \times 5}$. Os dados foram normalizados conforme a Equação 1.

$$X = \frac{X_{old} - \bar{X}}{\sigma_X} \quad (1)$$

onde X_{old} é a observação anterior, \bar{X} é a média das observações e σ_X é o desvio padrão das observações.

3.3 Métodos

A característica principal dos métodos de agrupamento é a capacidade de detectar a estrutura

subjacente nos dados, não só para classificação e reconhecimento de padrões, mas para redução e otimização de modelos. As técnicas de agrupamento podem ser aplicadas em dados quantitativos (numéricos), qualitativos (categóricos) ou em ambos (Balasko et al., 2010). Neste caso, será considerado agrupamento de dados quantitativos, haja vista que os dados obtidos são tipicamente observações de um processo químico.

O objetivo da análise de agrupamentos é a classificação de observações de acordo com as similaridades dentre elas e a organização desses dados em grupos. As técnicas usadas nesse estudo estão entre os métodos não supervisionados, ou seja, eles não usam identificadores de classe anteriores (Balasko et al., 2010). A seguir uma breve explanação dos métodos aplicados.

3.3.1 Fuzzy C-means

Esse algoritmo de agrupamento é baseado na minimização da função objetiva chamada *C-means functional*, Equação 2. Essa função pode ser vista estatisticamente como a medida da variância total de x_k e de v_i (Balasko et al., 2010).

$$J(X; U, V) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|x_k - v_i\|_A^2, \quad (2)$$

onde $X = [x_k, k = 1, 2, \dots, N]$ é o vetor de observações, a matriz $U = [\mu_{ik}]$ representa o grau de pertinência de cada observação e $V = [v_1, v_2, \dots, v_c]$, $v_i \in R^n$ é o vetor de protótipos dos *clusters*.

Essa minimização representa um problema de otimização não-linear, o que pode ser resolvido usando uma variedade de métodos disponíveis, mas o método mais popular é a simples iteração de Picard através das condições de primeira ordem dos pontos estacionários. Os pontos estacionários da função objetivo podem ser encontrados por meio dos multiplicadores Lagrange:

$$\bar{J}(X; U, V, \lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ikA}^2 + \sum_{k=1}^N \lambda_k \left(\sum_{i=1}^c \mu_{ik} - 1 \right), \quad (3)$$

onde D_{ikA}^2 , Equação 4, representa a distância do produto interno ao quadrado.

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i). \quad (4)$$

Caso $D_{ikA}^2 > 0, \forall i, k, m > 1$, então $(U, V) \in M_{fc} \times R^{n \times c}$ pode minimizar a função objetivo somente se:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (D_{ikA}/D_{jkA})^{2/(m-1)}}, \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (5)$$

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^m x_k}{\sum_{k=1}^N \mu_{ik}^m}, \quad 1 \leq i \leq c. \quad (6)$$

Observe que a Equação 6 é a média ponderada de x_k com os graus de associação (pertinência) ao *cluster* μ_{ik}^m , por isso que o algoritmo é chamado de "*c-means*". Esse algoritmo utiliza em seus cálculos a norma padrão da distância Euclidiana, o que induz grupos de hiperesferas. Devido a isso, ele só pode detectar agrupamentos com a mesma forma e orientação, pois a escolha comum da matriz de indução da norma é definido como o inverso da matriz de covariância, dada por (Balasko et al., 2010):

$$F = \frac{1}{N} \sum_{i=1}^k (x_k - \bar{x})(x_k - \bar{x})^T. \quad (7)$$

3.3.2 Gustafson-Kessel

Gustafson e Kessel estenderam o algoritmo padrão FCM empregando uma norma de distância adaptativa, para detectar agrupamentos de diferentes formas geométricas em um conjunto de dados. Cada *cluster* possui sua própria matriz A_i , o que implica a seguinte norma de produto interno (Gustafson and Kessel, 1978):

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \leq i \leq c, \quad 1 \leq k \leq N. \quad (8)$$

As matrizes A_i são usadas como variáveis de otimização no modo C-means, permitindo que cada *cluster* se adapte a norma de distância à estrutura topológica local dos dados. A função objetivo J do algoritmo GK é obtido por:

$$J(X; U, V, A) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ikA_i}^2. \quad (9)$$

Para obter uma solução viável, A_i deve ser constrangido de alguma forma e a maneira usual de conseguir isso é restringindo seu determinante. Com isso, a matriz A_i irá variar com seu determinante fixo para otimizar a forma do *cluster* enquanto seu volume permanece constante:

$$A_i = [\rho_i \det(F_i)]^{1/n} F_i^{-1}, \quad (10)$$

onde ρ é fixado para cada v_i e F_i é definido por:

$$F_i = \frac{\sum_{k=1}^N (\mu_{ik})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N (\mu_{ik})^m}. \quad (11)$$

Note que a substituição das Equações 10 e 11 em 8 fornece a norma de distância quadrada generalizada de Mahalanobis entre x_k e \bar{v}_i (média de cluster), onde F_i é ponderada por μ_{ik}^m (Gustafson and Kessel, 1978).

3.3.3 Gath-Geva

Esse é um algoritmo de agrupamento proposto por Bezdek e Dunn, que emprega uma norma de distância baseada nas estimativas de máxima verossimilhança difusa, através da Equação 12 (Bezdek and Dunn, 1975):

$$D_{ik} = \frac{\sqrt{\det(F_{wi})}}{\alpha_i} \exp\left(\frac{1}{2}(x_k - v_i^{(l)})^T F_{wi}^{-1}(x_k - v_i^{(l)})\right), \quad (12)$$

com a matriz de covariância difusa do i -ésimo *cluster* dada pela Equação 13:

$$F_{wi} = \frac{\sum_{k=1}^N (\mu_{ik})^w (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N (\mu_{ik})^w}, 1 \leq i \leq c. \quad (13)$$

Observe que a diferença entre a matriz F_i no algoritmo GK e F_{wi} definido acima é que o último não envolve o expoente de ponderação m , com intuito de generalizar a expressão. O valor de α_i é a probabilidade anterior de selecionar v_i , dada pela equação:

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N (\mu_{ik}), \quad (14)$$

onde μ_{ik} são interpretados como as probabilidades posteriores de selecionar v_i dado o ponto de dados x_k .

F_{wi} é usada em conjunto com uma distância “exponencial” e v_i não são limitados pelo volume. Entretanto, esse algoritmo é menos robusto, pois devido à norma de distância exponencial, pode convergir para um próximo ótimo local. Por isso esse algoritmo precisa de uma boa inicialização (Bezdek and Dunn, 1975).

Esses métodos foram implementados em rotinas do *software* Matlab e as funções usadas como base fazem parte da *Fuzzy Clustering e Data Analysis Toolbox* disponíveis em (Balasko et al., 2010).

4 Resultados e Discussões

Os parâmetros definidos foram: o número de *clusters* (baseado na quantidade de classes, foi determinado igual a 3), o expoente de difusão igual a 2 (valor padrão, segundo (Balasko et al., 2010)) e a tolerância de agrupamento igual a 10^{-5} .

Os resultados estão expostos na Tabela 2, os mesmos dizem respeito ao percentual de observações classificadas corretamente.

Com intuito de eliminar as redundâncias dos dados, foram calculados os índices de Correlação de Pearson de todas variáveis entre si. A Tabela 3, mostra os resultados desses índices.

Tabela 2: Índices de precisão dos algoritmos com o banco de dados A (%).

	FCM	GK	GG
Todas Freqüências	35,16	41,09	40,28

Tabela 3: Índices de correlação do banco de dados A.

	F	Zr	Zi	C
F	1			
Zr	-0,23	1		
Zi	0,2	-0,89	1	
C	-0,26	0,07	-0,19	1
R	-0,2	0,73	-0,93	0,22

Note que as variáveis que obtiveram índices elevados mais vezes foram Zi e R. Porém, Zi foi a variável que teve os índices de correlação mais altos e por isso a mesma foi removida do banco de dados e a dimensão do novo banco de dados é $B_{3672 \times 4}$.

Os algoritmos Fuzzy foram aplicados novamente e os resultados de precisão alcançados estão na Tabela 4. Observe que os resultados não tiveram avanços significativos se comparados com os valores da Tabela 2.

Tabela 4: Índices de precisão dos algoritmos com banco de dados B (%).

	FCM	GK	GG
Todas Freqüências	38,64	38,56	41,55

Após os testes realizados inicialmente e observando as características da estruturação dos dados, exposto na Figura 2, é possível observar que os dados estão distribuídos de forma alongada, de tal modo que os centroídes não podem representar bem as classes.

Em consequência disso, o próximo passo foi definir novos bancos de dados, com a finalidade de obter resultados de classificação mais expressivos. Assim, o banco de dados que será analisado vai levar em consideração apenas as observações que possuem os mesmos valores de Freqüência. Com isso, o novo banco de dados possui 72 observações. Os valores de Freqüência analisadas serão: 10 Hz (**F1**), 10^2 Hz (**F2**), 10^3 Hz (**F3**), 10^4 Hz (**F4**), 10^5 Hz (**F5**) e 10^6 Hz (**F6**).

A Tabela 5 mostra a precisão da classificação de cada método aplicada em cada valor de

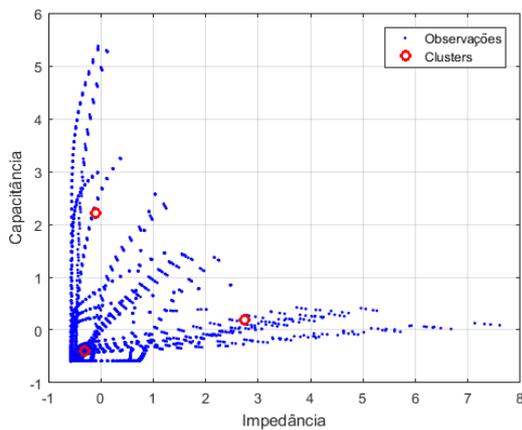


Figura 2: Estrutura dos dados.

frequência determinado.

Tabela 5: Resultados da precisão dos algoritmos (%).

	F1	F2	F3	F4	F5	F6
FCM	44,44	38,89	50,00	50,00	66,67	44,44
GK	44,44	50,00	61,11	65,28	61,11	61,11
GG	38,89	50,00	52,78	61,11	63,88	61,11

Observe que o melhor resultado obtido, foi através do algoritmo FCM na faixa de Frequência F5 com 66,67 %. Note também que, o algoritmo que teve resultados com menor variância foi o GK com média igual 57,17 % e desvio padrão 8,07. Com relação as Frequências analisadas, F5 teve os melhores resultados de classificação com os algoritmos usados, alcançando uma média de 63,88 % com desvio padrão igual a 2,78.

A Tabela 6 traz a matriz de confusão do algoritmo FCM com F5.

Tabela 6: Matriz de confusão de FCM com F5 (%).

	C1	C2	C3
C1	100	16,67	87,33
C2	0	83,33	0
C3	0	0	16,67

A Tabela 6 mostra que a C1 foi a única classe que obteve uma classificação perfeita, já a C2 teve um bom índice de classificação e por fim, a C3 foi a que teve o pior resultado, onde o algoritmo classificou uma grande quantidade de observações como sendo de C1.

5 Conclusões

O intuito desse trabalho foi testar a aplicação de novas ferramentas para classificar amostras com diferentes concentrações.

Com os métodos determinados e após a aplicação dos mesmos no banco de dados, os resultados alcançados através de cálculos de precisão demonstraram que dependendo a faixa de frequência das observações os algoritmos chegaram a atingir a precisão de 66%.

Após as análises realizadas nesse estudo, os resultados alcançados indicam que a Frequência de cada observação influencia na estrutura dos dados, o que nesse caso afetou o desempenho dos algoritmos. E como pôde ser observado, a faixa de Frequência que obteve o melhor desempenho foi F5.

Como trabalhos futuros, os autores têm como objetivo fazer novos testes envolvendo diferentes faixas de frequência com o intuito de verificar em qual faixa de frequência os métodos atingem o melhor resultado. E a aplicação de outros métodos *Fuzzy* que se encaixem melhor com o problema, métodos supervisionados, como por exemplo Redes Neurais e *Fuzzy* ARTMAP.

Agradecimentos

Agradecimento à CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa de mestrado concedida e a FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais).

Referências

- Balasko, B., Abonyi, J. and Feil, B. (2010). *Manual for Fuzzy Clustering and Data Analysis Toolbox (For Use with Matlab)*, Department of Process Engineering University of Veszprem. pp. 1 - 77.
- Bezdek, J. C. and Dunn, J. C. (1975). Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions, *IEEE Transactions on Computers* **C-24**(8): 835 – 838.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L. and Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment—a review, *Analytica Chimica Acta*, Vol. 891, Elsevier, pp. 1 – 14.
- Cetó, X., González-Calabuig, A., Capdevila, J., Puig-Pujol, A. and del Valle, M. (2015). Instrumental measurement of wine sensory descriptors using a voltammetric electronic

- tongue, *Sensors and Actuators B: Chemical* **207**(Part B): 1053 – 1059.
- Cetó, X. and Valle, M. (2014). Application of electronic tongues in the qualitative and quantitative analysis of beers, *IEEE SENSORS*, IEEE.
- Fukushima, K. L. (2017). *Desenvolvimento de sensores nanoestruturados para análise de realçadores de sabor em água*, Master's thesis, Programa de Pós-Graduação em Engenharia de Biomateriais, Universidade Federal de Lavras.
- Gustafson, D. E. and Kessel, W. C. (1978). Fuzzy clustering with a fuzzy covariance matrix, *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pp. 761 – 766.
- Liu, M., Wang, M., Wang, J. and Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and chinese vinegar, *Sensors and Actuators B: Chemical* **177**(Supplement C): 970 – 980.
- Lu, L., Hu, X., Tian, S., Deng, S. and Zhu, Z. (2016). Visualized attribute analysis approach for characterization and quantification of rice taste flavor using electronic tongue, *Analytica Chimica Acta*, Vol. 919, Elsevier, pp. 11 – 19.
- Milani, T. M. G. (2013). *Estudo da aromatização pré-extrusão e das condições de extrusão sobre as características físicas e retenção de aroma em proteína isolada de soja*, Master's thesis, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista Julio de Mesquita Filho.
- Nery, E. W. and Kubota, L. T. (2016). Integrated, paper-based potentiometric electronic tongue for the analysis of beer and wine, *Analytica Chimica Acta*, Vol. 918, Elsevier, pp. 60 – 68.
- Ouyang, Q., Zhao, J. and Chen, Q. (2013). Classification of rice wine according to different marked ages using a portable multi-electrode electronic tongue coupled with multivariate analysis, *Food Research International* **51**(2): 633 – 640.
- Peris, M. and Escuder-Gilabert, L. (2013). Online monitoring of food fermentation processes using electronic noses and electronic tongues: a review, *Analytica chimica acta*, Vol. 804, Elsevier, pp. 29 – 36.
- Wei, Z. and Wang, J. (2014). Tracing floral and geographical origins of honeys by potentiometric and voltammetric electronic tongue, *Computers and Electronics in Agriculture*, Vol. 108, Elsevier, pp. 112 – 122.
- Wei, Z., Wang, J. and Jin, W. (2013). Evaluation of varieties of set yogurts and their physical properties using a voltammetric electronic tongue based on various potential waveforms, *Sensors and Actuators B: Chemical* **177**(Supplement C): 684 – 694.
- Wei, Z., Wang, J. and Zhang, X. (2013). Monitoring of quality and storage time of unsealed pasteurized milk by voltammetric electronic tongue, *Electrochimica Acta*, Vol. 88, Elsevier, pp. 231 – 239.
- Zielinski, A. A., Haminiuk, C. W., Nunes, C. A., Schnitzler, E., Ruth, S. M. and Granato, D. (2014). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline, *Comprehensive reviews in food science and food safety* **13**(3): 300 – 316.