



ANÁLISE EXPERIMENTAL DO MÉTODO DE COMPOSIÇÃO PROBABILÍSTICA DE PREFERÊNCIAS EM SEU MODO TRICOTÔMICO (CPP-TRI) NA CLASSIFICAÇÃO ORDENADA DE VINHOS DA REGIÃO DO MINHO

Glauco Barbosa da Silva

Centro de Análises de Sistemas Navais (CASNAV - UFF)
Praça Barão de Ladário s/n, Ilha das Cobras -RJ
glaucos@id.uff.br

Helder Gomes Costa

Universidade Federal Fluminense (UFF)
Rua Passos da Pátria, 156 - Bloco D – São Domingos, Niterói, RJ
helder.hgc@gmail.com

Resumo

O presente artigo tem como objetivo analisar a classificação ordenada do método CPP-TRI de um conjunto de dados de vinhos tinto. Para isso, a partir de um experimento inicial, os resultados alcançados são apresentados, analisados e comparados com os resultados da técnica SVM.

Palavras-Chaves: CPP-TRI; Apoio à Decisão Multicritério; MCDA; Sorting.

Tópicos: ADM – Apoio à Decisão Multicritério; IND- PO na Indústria; OA- Outras aplicações em PO.

Abstract

The present article aims to analyze the CPP-TRI method through the sorting task in the red wine dataset. For this, from an initial experiment, the results achieved are presented, analyzed and compared with the results from SVM technique.

Keywords: CPP-TRI; Multicriteria Decision Aid; MCDA; Sorting.

Paper topics: ADM- Multicriteria Decision Support; IND – OR in Industry; OA – Other applications in OR.



1. Introdução

O presente artigo tem como objetivo apresentar uma análise experimental do Método de Composição Probabilística de Preferências em seu modo Tricotômico (CPP-TRI). A partir dos resultados dos testes físico-químicos de uma amostra de vinhos da região norte de Portugal (Minho), um conjunto de dados é classificado pelo método CPP-TRI. Em seguida, os resultados obtidos pelo método CPP-TRI são analisados.

Impulsionado pelos benefícios que traz para a saúde e embasado por uma ciência (Enologia) responsável por observar e acompanhar todo o processo de produção, desde o plantio das uvas, passando pelo envelhecimento, até que as garrafas cheguem às prateleiras do mundo; o vinho é um produto cada vez mais apreciado por uma vasta gama de consumidores.

Frente a tal favorável cenário e de modo a apoiar o seu crescimento, a indústria vinícola tem investido em novas tecnologias para os processos de produção e venda de vinho. Neste contexto, são elementos-chave: a certificação e a avaliação da qualidade. Em geral, a certificação avalia: os resultados dos testes físico-químicos de laboratório, que determinam a densidade, teor alcoólico e pH, entre outros; e os testes sensoriais, decorrente da avaliação de especialistas.

Segundo [Smith and Margolskee \[2006\]](#), a descrição da percepção do gosto humano é influenciada por quatro fatores: salinidade, acidez, doçura e amargor. Conjuguar esses fatores torna a classificação de vinhos uma tarefa extremamente complexa e difícil, que tem atraído muitos pesquisadores para analisar a relação entre os testes físico-químicos e sensoriais.

Neste artigo, um conjunto de dados com 1599 instâncias de vinho tinto, onde cada instância é composta pelos resultados dos testes físico-químico e sensoriais da amostra, é classificado pelo método CPP-TRI; os dados foram coletados entre maio de 2004 e fevereiro de 2007 e estão disponíveis no repositório da Universidade da Califórnia, Irvine - *Machine Learning Repository* [Lichman 2013]. A modelagem de preferências proposta, baseada nos dados analíticos disponíveis da etapa de certificação, busca apoiar as avaliações dos especialistas, contribuindo para a melhoria do processo decisório.

Um estudo correlato foi realizado por Cortez *et al.* [2009], que propôs uma abordagem baseada em *Support Vector Machine* para classificar esse mesmo conjunto de dados.

A justificativa para o presente trabalho decorre da recente proposição do método CPP-TRI, que ainda possui poucos resultados avaliados e comparados com métodos mais difundidos na literatura; embora tenha sido aplicado a problemas de diferentes áreas como por exemplo: na classificação de municípios quanto a oferta de creches[Sant'anna *et al.* 2013]; na classificação de cursos por múltiplos avaliadores[Sant'anna *et al.* 2016]; na classificação das unidades da federação com base da população economicamente ocupada[Silva *et al.* 2016].

Este trabalho está organizado em cinco seções, como a seguir: na Seção 2, o método CPP-TRI é descrito; a Seção 3 apresenta as características do conjunto de dados; na Seção 4, O método CPP-Tri é aplicado e os resultados são analisados; ao fim, as considerações são apresentadas na Seção 5.

2. O Método CPP-TRI

O método de Composição Probabilística de Preferências em seu modo Tricotômico (CPP-TRI) proposto por [Sant'anna *et al.* \[2015\]](#), é definido como uma abordagem para o problema de classificação ordenada, que consiste em alocar alternativas(instâncias) em classes pré-determinadas por perfis de referência(vetor de avaliações segundo cada critério); baseia-se na transformação probabilística proposta por [Sant'anna \[2002\]](#), similarmente a transformação dos números *crisp* em intervalos de pertinência da Teoria dos Conjuntos Fuzzy [Zadeh 1965].

O método CPP-TRI tem como principal característica levar em conta as imprecisões existentes nas avaliações segundo cada critério; busca associar medidas numéricas a distribuições de probabilidade. Em geral, assume-se, além de distribuições normais, idêntica distribuição e independência entre as perturbações que provocam a imprecisão nas medidas; os dados das



avaliações disponíveis em cada critério são usados para estimar as variâncias dessas distribuições normais [Sant'anna *et al.* 2012].

Sant'anna *et al.* [2015] afirma que uma vantagem da modelagem com distribuição de probabilidades é liberar o decisor da tarefa de atribuir pesos ou determinar parâmetros de fronteira.

Uma vez que as avaliações são transformadas em distribuições de probabilidade, o modelo de preferências é construído pelas medidas de distância probabilística (δ), calculada para todas as classes j (ordenadas), como a diferença entre as probabilidades estar acima do perfil da classe j e de estar abaixo do perfil da classe j . Para uma alternativa (instância) qualquer a_i ; a_i é comparada com o perfil de referência de cada classe j ; resultando numa probabilidade de estar acima da classe j e em outra probabilidade de estar abaixo da classe j . a_i é alocada na classe mais próxima, isto é, a classe j que o possui o menor valor absoluto de (δ).

3. Características dos Dados

A amostra analisada é composta por 1599 instâncias de vinhos tintos com características da região produtora do norte de Portugal (Minho). Para possibilitar um melhor entendimento das avaliações e da classificação dos vinhos, algumas considerações acerca dos seus componentes principais são destacadas, sem a pretensão de esgotar o assunto.

Dentre os componentes presentes nos testes laboratoriais, [Nierman \[2004\]](#) descreve os principais, como a seguir:

- a acidez é uma propriedade fundamental, que contribui principalmente para o sabor e resistência à infecção microbiana; afeta o sabor e o pH, que influenciam a cor, a estabilidade à oxidação e, conseqüentemente, a vida útil global de um vinho. Alguns ácidos surgem como resultado do processo de fermentação de leveduras e / ou bactérias, ou seja, esses parâmetros são relevantes na qualidade do vinho. A acidez total é dividida em dois grupos: *acidez volátil* e *acidez fixa*. A *acidez volátil* define a deterioração do vinho, composta em grande parte de ácido acético (vinagre).
- O ácido cítrico é um ácido orgânico que tem muitos usos na produção de vinho; muitas vezes usado como um conservante natural. Em outras adicionado aos vinhos para aumentar a acidez e dar um sabor "fresco". A adição de ácido cítrico apresenta como desvantagem a instabilidade microbiana, uma vez que as bactérias usam ácido cítrico no seu metabolismo, pode aumentar o crescimento de micróbios indesejados.
- Açúcar residual é a quantidade de açúcar remanescente após a fermentação. Tem como razões principais: a incapacidade da levedura de converter açúcar em álcool; a interrupção da fermentação pela presença de um álcool mais "pesado"; e a queda significativa da temperatura, que acaba matando a levedura. O açúcar residual é equilibrado pela acidez, teor alcoólico e taninos.
- Cloreto (cloreto de sódio) é a quantidade de sal no vinho.
- Dióxido de enxofre é um conservante presente em todos os vinhos.
- Densidade pode ser usada como medida de verificação rápida de homogeneidade e consistência.
- pH descreve o quão ácido ou básico um vinho está, variando numa escala de 0 (muito ácido) a 14 (muito básico). A maioria dos vinhos está entre 3 e 4.

Para os onze parâmetros principais da amostra provenientes dos testes laboratoriais, e considerados como critérios para avaliação, a tabela 1 apresenta a uma consolidação das medidas de interesse.

Tabela 1. Medidas de Interesse - Testes laboratoriais.

Critério	Descrição	
----------	-----------	--



		Min	Max	Média	Desvio-padrão
g ₁	Acidez Fixa	4.6	15.9	8.32	1.741
g ₂	Acidez Volátil	0.12	1.58	0.528	0.179
g ₃	Ácido Cítrico	0.0	1.0	0.271	0.195
g ₄	Açúcar Residual	0.9	15.5	2.539	1.41
g ₅	Cloreto de Sódio	0.012	0.611	0.087	0.047
g ₆	Dióxido de Enxofre Livre	1.0	72.0	15.875	10.46
g ₇	Dióxido de Enxofre Total	6.0	289.0	46.468	32.895
g ₈	Densidade	0.99	1.004	0.997	0.002
g ₉	pH	2.74	4.01	3.311	0.154
g ₁₀	Sulfato	0.33	2.0	0.658	0.17
g ₁₁	Álcool	8.4	14.9	10.422	1.006

Nos testes sensoriais, o atributo *qualidade* é decorrente da média das avaliações de pelo menos três especialistas, que atribuem notas que variam de 0(muito ruim) a 10(excelente). Para cada instância, a avaliação dos especialistas é registrada no atributo *Qualidade*; para o conjunto analisado foram identificadas seis classes. A distribuição das instâncias conforme a avaliação dos especialistas (qualidade) é apresentada na figura 1.

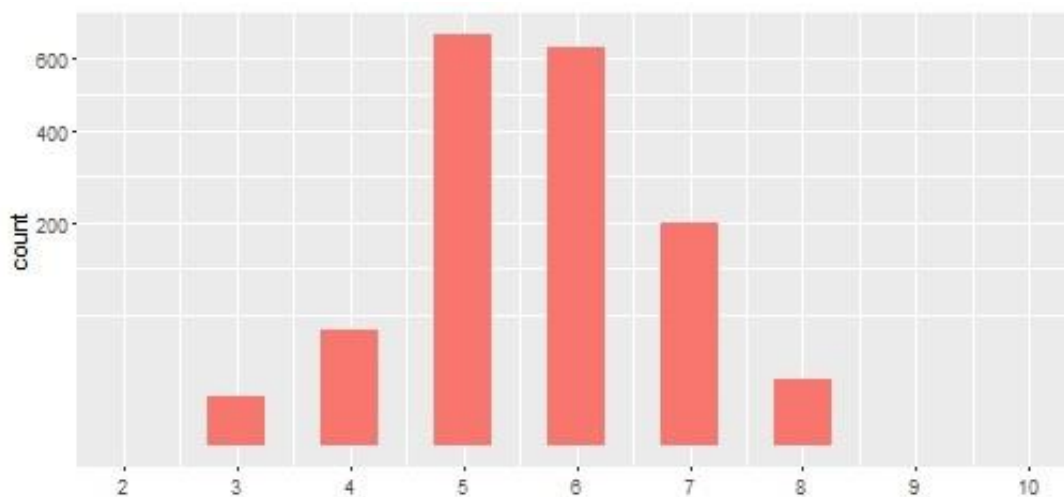


Figura 1. Distribuição das instâncias por atributo Qualidade.

	3	4	5	6	7	8	total
Instâncias	10	53	681	638	199	18	1599



(Vinhos Tinto)												
-------------------	--	--	--	--	--	--	--	--	--	--	--	--

4. Aplicação

Nesta seção o método CPP-TRI é aplicado ao conjunto de dados. Conforme proposto por Sant'anna [2014], na formulação de um problema de classificação, adotamos:

- $G = (g_1, \dots, g_k)$, como um conjunto de k critérios.
- $A = (a_1, \dots, a_k)$, como um vetor de dimensão R^k que guarda as avaliações de acordo com cada k critério para a instância a ser classificada.
- $C = (C_1, \dots, C_j)$, é um conjunto de j classes, ordenadas da pior para a melhor, de modo que uma instância é tão melhor quanto maior for o índice da classe que ela for classificada.

Baseado nos testes laboratoriais, para cada classe C_j é estabelecido um perfil representativo da classe – (C_{jk}) , onde cada dimensão do perfil é definida pela média e desvio-padrão para cada g_k critério; o perfil representativo de cada classe é apresentado na tabela 2.

Tabela 2. Perfis representativos por Classe.

Classe		g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
C_8	Média	8.57	0.42	0.39	2.58	0.07	13.28	33.44	0.995	3.27	0.77	12.09
	DSVP	2.12	0.14	0.20	1.30	0.01	11.16	25.43	0.002	0.20	0.12	1.22
C_7	Média	8.87	0.40	0.38	2.72	0.08	14.05	35.02	0.996	3.29	0.74	11.47
	DSVP	1.99	0.15	0.19	1.37	0.03	10.18	33.19	0.002	0.15	0.14	0.96
C_6	Média	8.35	0.50	0.27	2.48	0.08	15.71	40.87	0.997	3.32	0.68	10.63
	DSVP	1.80	0.16	0.20	1.44	0.04	9.94	25.04	0.002	0.15	0.16	1.05
C_5	Média	8.17	0.58	0.24	2.53	0.09	16.93	56.51	0.997	3.30	0.62	9.90
	DSVP	1.56	0.16	0.18	1.36	0.05	10.96	36.99	0.002	0.15	0.17	0.74
C_4	Média	7.78	0.69	0.17	2.69	0.09	12.26	36.25	0.997	3.38	0.60	10.27
	DSVP	1.63	0.22	0.20	1.79	0.08	9.03	27.58	0.002	0.18	0.24	0.93
C_3	Média	8.36	0.88	0.17	2.64	0.12	11.00	24.90	0.997	3.40	0.57	9.96
	DSVP	1.77	0.33	0.25	1.40	0.07	9.76	16.83	0.002	0.14	0.12	0.82

Sant'anna [2014] estabelece como requisito para a aplicação do método que os perfis estejam ordenados, isto é, para todos os critérios, nenhuma avaliação do perfil de uma classe superior pode ser menor que a avaliação do perfil de uma classe inferior no mesmo critério; e pelo menos um critério do perfil da classe superior deve ser maior que a avaliação do perfil da classe inferior.

Numa breve análise da tabela 2, é possível identificar que os perfis não atendem o requisito de ordenação para aplicação do CPP-TRI. Para possibilitar a aplicação do método, e de forma a atender ao requisito, os perfis das classes são ajustados observando o algoritmo para eliminar inversões (quadro 1), de modo a evitar atributos variando em diferentes direções.

Quadro 1. Algoritmo para Eliminar Inversões em Perfis Representativos*



Objetivo: As coordenadas dos perfis de uma classe $i1$ qualquer devem estar todas abaixo das coordenadas dos perfis das classes $i2$ com $i2 > i1$.

Notação. Para um critério qualquer, denote por P_i o valor inicial do critério no perfil da classe i e por Q_i o valor final (isto é, após a eliminação da inversão).

Algoritmo

Identifique para ser a classe central (C) a classe com mais observações.

Se houver mais de uma classe nessa condição e houver inversão em algum critério entre elas, use a média aritmética dos valores invertidos em substituição a eles.

Para as classes i com $i < C$: $Q_i = \text{mínimo}(Q_{i+1}, P_i)$.

Para as classes i com $i > C$: $Q_i = \text{máximo}(Q_i - 1, P_i)$.

*Cedido pelo autor do método.

Aplicando o algoritmo acima, a partir da classe central (C_5) - maior número de observações – novos perfis representativos são estabelecidos, como mostrado na tabela 3.

Tabela 3. Novos perfis representativos das Classes (ajustado)

Classe		g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
C_8	Média	8.87	0.40	0.39	2.48	0.07	-13.28	-33.44	-0.995	-3.267	0.77	12.09
	DSVP	2.12	0.14	0.20	1.30	0.01	11.16	25.43	0.002	0.20	0.12	1.22
C_7	Média	8.87	0.40	0.38	2.48	0.08	-14.05	-35.02	-0.996	-3.291	0.74	11.47
	DSVP	1.99	0.15	0.19	1.37	0.03	10.18	33.19	0.002	0.15	0.14	0.96
C_6	Média	8.35	0.50	0.27	2.48	0.08	-15.71	-40.87	-0.997	-3.318	0.68	10.63
	DSVP	1.80	0.16	0.20	1.44	0.04	9.94	25.04	0.002	0.15	0.16	1.05
C_5 (Central)	Média	8.17	0.58	0.24	2.53	0.09	-16.93	-56.51	-0.997	-3.318	0.62	9.90
	DSVP	1.56	0.16	0.18	1.36	0.05	10.96	36.99	0.002	0.15	0.17	0.74
C_4	Média	7.78	0.69	0.17	2.69	0.09	-16.93	-56.51	-0.997	-3.382	0.60	9.90
	DSVP	1.63	0.22	0.20	1.79	0.08	9.03	27.58	0.002	0.18	0.24	0.93
C_3	Média	7.78	0.88	0.17	2.69	0.12	-16.93	-56.51	-0.997	-3.398	0.57	9.90
	DSVP	1.77	0.33	0.25	1.40	0.07	9.76	16.83	0.002	0.14	0.12	0.82

Na classificação das instâncias do conjunto de dados em análise, para uma classe em um k -ésimo critério, tem-se:

A_{ijk+} ← a probabilidade da instância a_i apresentar um valor acima do perfil representativo da classe C_{jk} .



A_{ijk-} ← a probabilidade da instância a_i apresentar um valor abaixo do perfil representativo da classe C_{jk} .

A_{ij+} ← a probabilidade da instância a_i apresentar um valor acima do perfil representativo da classe C_j .

$$A_{ij+} = \prod_k A_{ijk+}$$

A_{ij-} ← a probabilidade da instância a_i apresentar um valor abaixo do perfil representativo da classe C_j .

$$A_{ij-} = \prod_k A_{ijk-}$$

O procedimento para associar uma instancia à uma determinada classe, classificação, baseia-se na diferença entre as probabilidades de estar acima ou abaixo de cada classe, distância probabilística (δr). Sendo a instância classificada na classe na qual a distância probabilística(δr) é mínima.

$$\delta r = |A_{ij+} - A_{ij-}|$$

No cálculo das probabilidade definidas acima, Sant'anna(2014) destaca que se a variância é pequena, um “único” critério pode ser suficiente para determinar a pertinência de uma instância à classe. Ou seja, a variância deve ser suficientemente grande para que a probabilidade associada ao critério não seja demasiadamente pequena e torne os demais critérios irrelevantes, isto é, evitar o poder de veto de um único critério.

A partir do conjunto de dados de vinhos tinto [Lichman 2013] e dos perfis representativos das classes, arquivos *profilesbyclass.txt*(média) e *stdev.txt*(desvios-padrão), um experimento para o cálculo das probabilidades, distâncias probabilísticas e classificação foi conduzido com o código R, processado na ferramenta *Rstudio*[*Rstudio 2017*], conforme apresentado no quadro 2 abaixo:

Quadro 2. Código R para Classificação com o método CPP-TRI

```
## i = instance index ( 1 -- 1599)
## j = classe index ( 1 -- 6)
## k = criterion index ( 1 -- 11)
WinesMod <- read.csv2("/WinesMod.csv")
stdevbyclass<-read.table("/stdevbyclass.txt", quote="\\"",
                        comment.char="")
profilebyclass<-read.table("/profilebyclass.txt",quote="\\"",
                           comment.char="")

A<-WinesMod
B<- profilebyclass
D<- stdevbyclass
acima<-A
abaixo<-A
Pacima<- A
Pabaixo<- A
Pacima<-Pacima[1:nrow(A), 1:nrow(B)]
Pabaixo<-Pabaixo[1:nrow(A), 1:nrow(B)]
C<-Pacima
Classe<-Pabaixo
for (i in 1:1599)
```



```

{
for (j in 1:6)
{
for (k in 1:11)
{
abaixo[j,k]<-pnorm(B[j,k],A[i,k],D[j,k])
acima[j,k]<-1-pnorm(B[j,k],A[i,k],D[j,k])
}
Pabaixo[i,j]<-prod(abaixo[j,])
Pacima[[i,j]]<-prod(acima[j,])
}
}
C<-abs(Pacima-Pabaixo)
for (i in 1:nrow (A))
{
for (j in 1:nrow (B))
{
Classe[i,j]<-ifelse (abs(C[i,j])==min(abs(C[i,])), 9-j, ":")
}
}
View(Classe)
write.csv(Classe, file="ClasseRed.csv")

```

Os resultados da classificação das instâncias pelo método CPP-TRI estão sumarizados no quadro 3 e na tabela 4, como a seguir.

Quadro 3. Distribuição das instâncias classificadas com pelo CPP-TRI.

CPP-TRI							
	3	4	5	6	7	8	total
Instâncias (Vinhos)	310	194	310	300	237	248	1599

Table 4. Matriz de Confusão.- CPP-TRI.

Classe Original	CPP-TRI Classificação						
	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	Total
C ₃	4	2	2	2	0	0	10
C ₄	10	9	19	9	2	4	53
C ₅	226	102	136	125	62	30	681
C ₆	62	78	130	133	122	113	638
C ₇	8	3	23	29	44	92	199
C ₈	0	0	0	2	7	9	18
	310	194	310	300	237	248	

A partir da classificação com o CPP-TRI, tabela 4, pode-se observar que a acurácia (taxa de acerto) foi de 20.95% das instâncias foram classificadas na classe original. Aumentando o nível de



tolerância para uma classe adjacente, a taxa de acerto sobe para 53.22%, isto é, considera-se como acerto a classificação na classe original ou em uma classe adjacente acima ou abaixo.

Em um experimento análogo, [Cortez et al. \[2009\]](#) propuseram um classificador baseado *Support Vector Machine(SVM)*, com diferentes níveis de tolerância. Os resultados alcançados com a técnica SVM para uma tolerância ($T=0$), os resultados do classificador SVM e CPP-TRI são bastante similares, em torno de 20%. Para uma tolerância ($T=0.5$), os resultados gerais do SVM são superiores, 62.35% contra 53.22% do CPP-TRI. Entretanto, para as classes com menor número de instâncias, a taxa de acertos foi maior para o CPP-TRI. O modelo SVM não classificou nenhuma instância na classe 3 e também não obteve acerto para a classe 8.

A acurácia e precisão do método CPP-TRI, por classe, estão consolidadas na tabela 5.

Tabela 5. Acurácia e Precisão por classe.

	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Acurácia(%)	40	16.98	19.87	20.85	22.11	50
Precisão(%)	1.29	4.64	43.87	44.33	18.57	3.63

Embora os resultados não sejam tão distantes dos resultados apresentados pelo classificador SVM, é possível que melhores resultados possam ser alcançados. [Cortez et al. \[2009\]](#) sugerem que nem todos os atributos sejam relevantes. Dessa forma, a primeira ação demanda foi a avaliar a força da associação entre dois atributos, correlação de Pearson(r). Foram identificadas forte associações entre os seguintes atributos: Acidez Fixa – Ácido Cítrico; Acidez Fixa – Densidade; Acidez Fixa – pH; Acidez Volátil - Ácido Cítrico; Ácido Cítrico – pH; Dióxido de Enxofre Livre – Dióxido de Enxofre Total; e Densidade – Álcool. A correlação Acidez Fixa - Densidade é ilustrada na figura 2 abaixo.

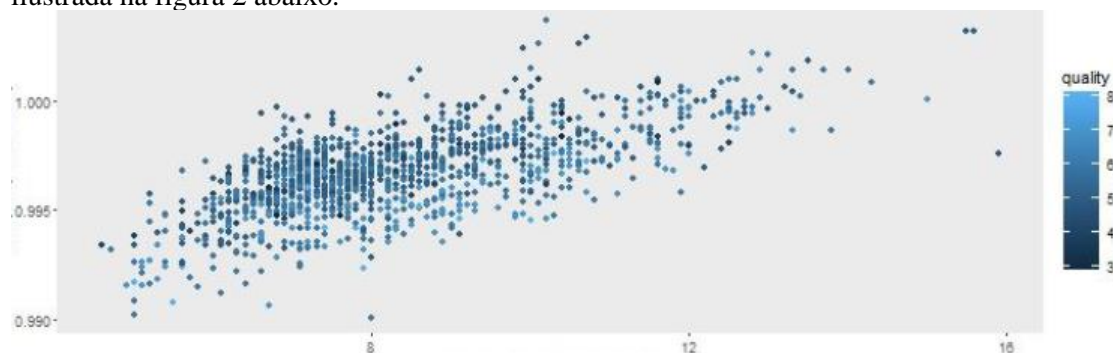


Figura 2. Correlação Acidez Fixa – Densidade ($r=0.668$).

5. Considerações Finais

O presente trabalho analisou a classificação ordenada de um conjunto de vinhos tinto com o método CPP-TRI. A tarefa de classificação de vinhos é bastante complexa e difícil uma vez que combina parâmetros físico-químicos e sensoriais. Os resultados alcançados pelo método foram comparados aos resultados obtidos em um trabalho anterior que utilizou a técnica de SVM para a classificação do mesmo conjunto de dados.

Sem tolerância, os resultados do método CPP-TRI foram similares ao SVM. Considerando que o SVM é uma técnica amplamente difundida, pode-se dizer que os resultados do CPP-TRI foram bastante satisfatórios.

Entretanto, a classificação pode ter sido influenciada (e comprometida) pela natureza dos dados: as classes não são balanceadas e existe correlação entre os atributos.

Sugere-se que em trabalhos futuros, seja incluída a etapa de seleção de atributos antes da aplicação do método.



6. Referências

Cortez, P.; Cerdeira, A.; Almeida, F.; Matos, T.; Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. **Decision Support Systems**, v. 47, n. 4, p. 547-553, ISSN 01679236 (ISSN). Disponível em: < <https://www.scopus.com/inward/record.uri?eid=2-s2.0-72049084727&partnerID=40&md5=40272319d9be0fe13b3f021fefadccc7> >.

Lichman, M. (2013) **UCI Machine Learning Repository**.

Nierman, D. Whats in Wine? **Waterhouse Lab**, University of California, Davis - UC Davis, 2004. Disponível em: < www.waterhouse.ucdavis.edu/whats-in-wine/fixed-acidity >. Acesso em: 20/03/2017.

Rstudio. (2017). Rstudio: Integrated Development Environment for R. Boston, MA. Disponível em: < <http://www.rstudio.org> >.

Sant'anna, A. P. (2002). Aleatorização e Composição de Medidas de Preferência. **Pesquisa Operacional**, v. 22, p. 17,

Sant'anna, A. P. (2014). **Probabilistic Composition of Preferences, Theory and Applications**. Springer International Publishing, ISBN 9783319112770. Disponível em: < <https://books.google.com.br/books?id=eFDCBAAAQBAJ> >.

Sant'anna, A. P.; Costa, H. G.; Nepomuceno, L. D. D. O.; Pereira, V. (2016). A Probabilistic Approach Applied to the Classification of Courses by Multiple Evaluators. **Pesquisa Operacional**, v. 36, n. 3, p. 17,

Sant'anna, A. P.; Costa, H. G.; Pereira, V. (2015). CPP-TRI: A sorting method based on the probabilistic composition of preferences. **International Journal of Information and Decision Sciences**, v. 7, n. 3, p. 193-212, Disponível em: < <http://www.scopus.com/inward/record.url?eid=2-s2.0-84940495090&partnerID=40&md5=32382e90f963eff5ecca607b6c1b1d5> >.

Sant'anna, A. P.; Costa, H. G.; Pereira, V. (2012). CPP-TRI: Um Método de Classificação Ordenada Baseada em Composição Probabilística. **Relatórios de Pesquisa em Engenharia de Produção**, v. 12, n. 8, p. 14, ISSN 1678-2399.

Sant'anna, A. P.; Faria, F.; Costa, H. G. (2013). Aplicação da Composição Probabilística e do Método das K-Médias à Classificação de Municípios Quanto à Oferta de Creches. **Cadernos do IME - Série Estatística**, v. 34, p. , ISSN 1413-9022 / 2317-4536.

Silva, G. B.; Barros, M.D.; Costa, H. G. (2016) **Abordagem Híbrida Multicritério - Mineração de Dados Aplicada a Classificação de Unidades da Federação com Base na População Economicamente Ocupada**. XVIII Simpósio de Pesquisa Operacional & Logística da Marinha (SPOLM). Rio de Janeiro: Blucher Marine Engineering Proceedings. 2.

Smith, D.; Margolskee, R. (2006). Making sense of taste. **Scientific American**, v. 16(3), n. Special issue, p. 84-92,

Zadeh, L. A. (1965). Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338-353, 6// ISSN 0019-9958. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S001999586590241X> >.