

Minerando Dados para Caracterizar Padrões Sazonais de Colônias de Abelhas *Apis mellifera*

XIV SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO

Felipe Anderson Oliveira Maciel

Universidade Federal do Ceará

Programa de Pós-Graduação em Engenharia de Teleinformática

Grupo de Redes de Computadores, Engenharia de Softwares e Sistemas



UNIVERSIDADE
FEDERAL DO CEARÁ



GREat

Grupo de Redes de Computadores
Engenharia de Software
e Sistemas



Agenda

2

1. Introdução

- Motivação
- Contribuição

2. Trabalhos relacionados

3. Material e Métodos

- Bases de Dados
- Pré-Processamento dos Dados
- Estratégias de Aprendizado

4. Avaliação Experimental

5. Resultados

6. Conclusão

Introdução

Motivação

4

Polinização

- 75% das culturas vegetais dependentes
 - crescimento de 300% nos últimos 50-60 anos
- Principais agentes polinizadores: abelhas



Figura 1. Abelhas em atividade de polinização. Fonte: <http://www.jornalentreposto.com.br/99-arquivos/1481-insetos-polinizadores-melhoram-productividade-agricola>

Motivação

5

Diferentes estados observados durante o ciclo de vida de uma colônia

- Abandono
- Produção de crias

Inspeção visual

Apicultura de precisão

- Monitoramento remoto de apiários
- RSSF e IoT
- Semântica dos dados

Contribuição

6

Contribuição

- i. Implementação de um método que utiliza técnicas de mineração de dados para descobrir e caracterizar padrões sazonais de colônias de abelhas *Apis mellifera*,
- ii. Classificar o estado da colônia para novas amostras.

Trabalhos relacionados

Trabalhos relacionados

8

Meikle e Holst (2015)

- Principais grandezas sensoriadas: temperatura, umidade, massa, gases internos, áudio, vídeo e vibrações

Ruan et al. (2017)

- Massa pode ser utilizada para estimar a evaporação de néctar e água e o mel produzido

Kridi et al. (2014) e Kridi et al. (2016)

- Rede de sensores para detecção de temperaturas atípicas
- Algoritmo capaz de identificar os sinais do comportamento pré-migratório

Becher (2010)

- Impacto da variação da temperatura na dinâmica populacional da colônia

Trabalhos relacionados

9

Ostwald et al. (2016)

- A umidade é uma variável importante, associada ao resfriamento por evaporação

Kviesis et al. (2015)

- Aplicação de RSSF para monitoramento da temperatura e umidade

Murphy et al. (2016)

- Análise com árvores de decisão para descrever as condições internas e a atividade da colônia
- Interno: CO₂, O₂, gases poluentes, temperatura, umidade/ Externo: intensidade solar, chuva, temperatura

Zacepins et al. (2015)

- Nota-se uma carência de soluções automatizadas para extração do conhecimento

Material e Métodos

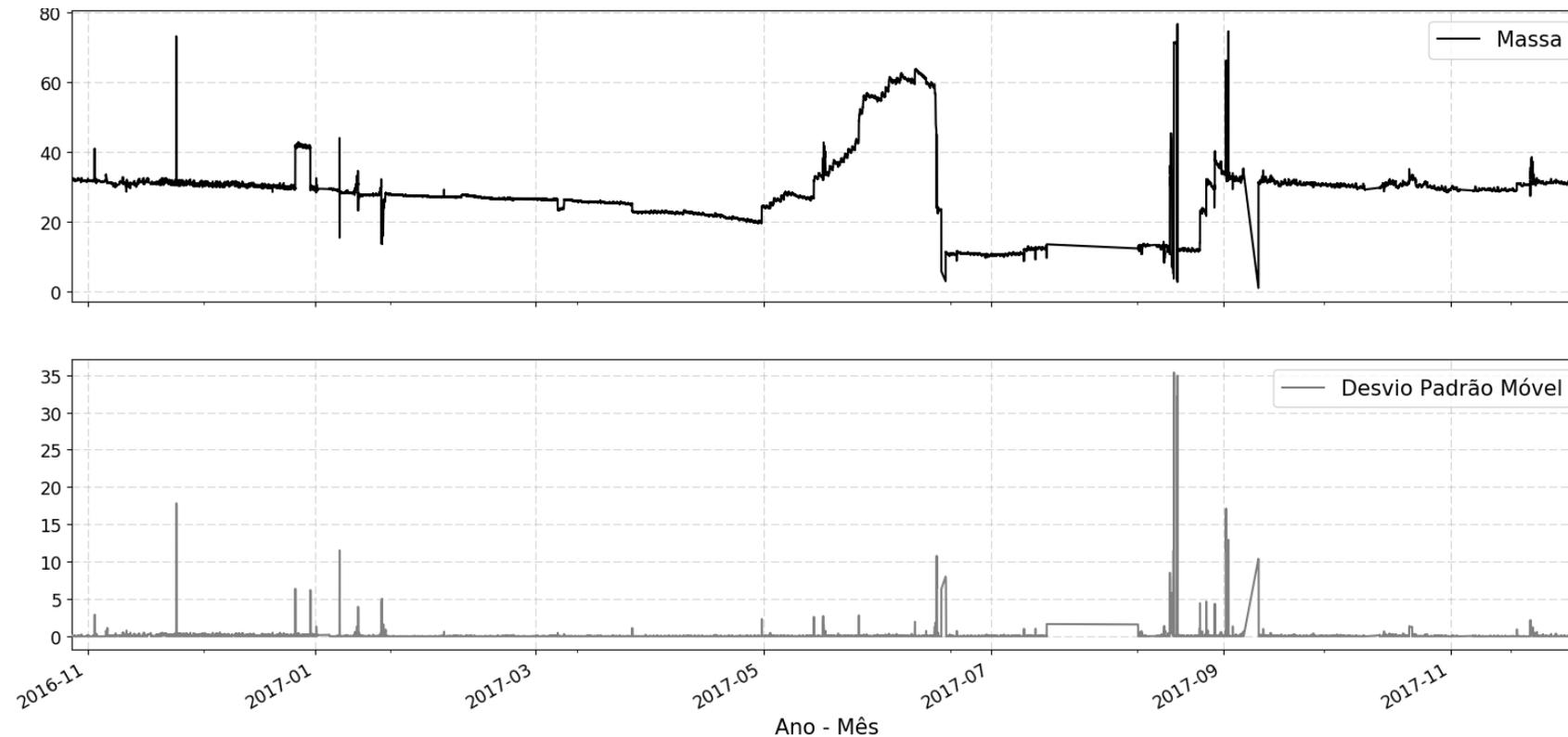
Bases de Dados

11

- HiveTool.net
- Arnas (Dinamarca) – 104.912 amostras
- 27/10/2016 a 05/12/2017
- Temperatura, umidade e massa

Pré-Processamento dos Dados

- Remoção de anomalias – desvio padrão móvel



Estratégias de Aprendizado

13

- Normalização dos dados

$$Z = \frac{x - \mu}{\sigma}$$

- Clusterização
 - SSD (*Sum of Squared Distances*): avaliação do posicionamento dos protótipos

$$SSD = \sum_{\forall x \in V} |x - w|^2$$

- Validação da clusterização – índice Calinski-Harabasz

$$CH(K) = \frac{tr(B_K)/(K - 1)}{tr(W_K)/(N - K)}$$

Estratégias de Aprendizado

14

- Classificação
 - Naive Bayes
 - k-Nearest Neighbors
 - Random Forest

Avaliação Experimental

Avaliação Experimental

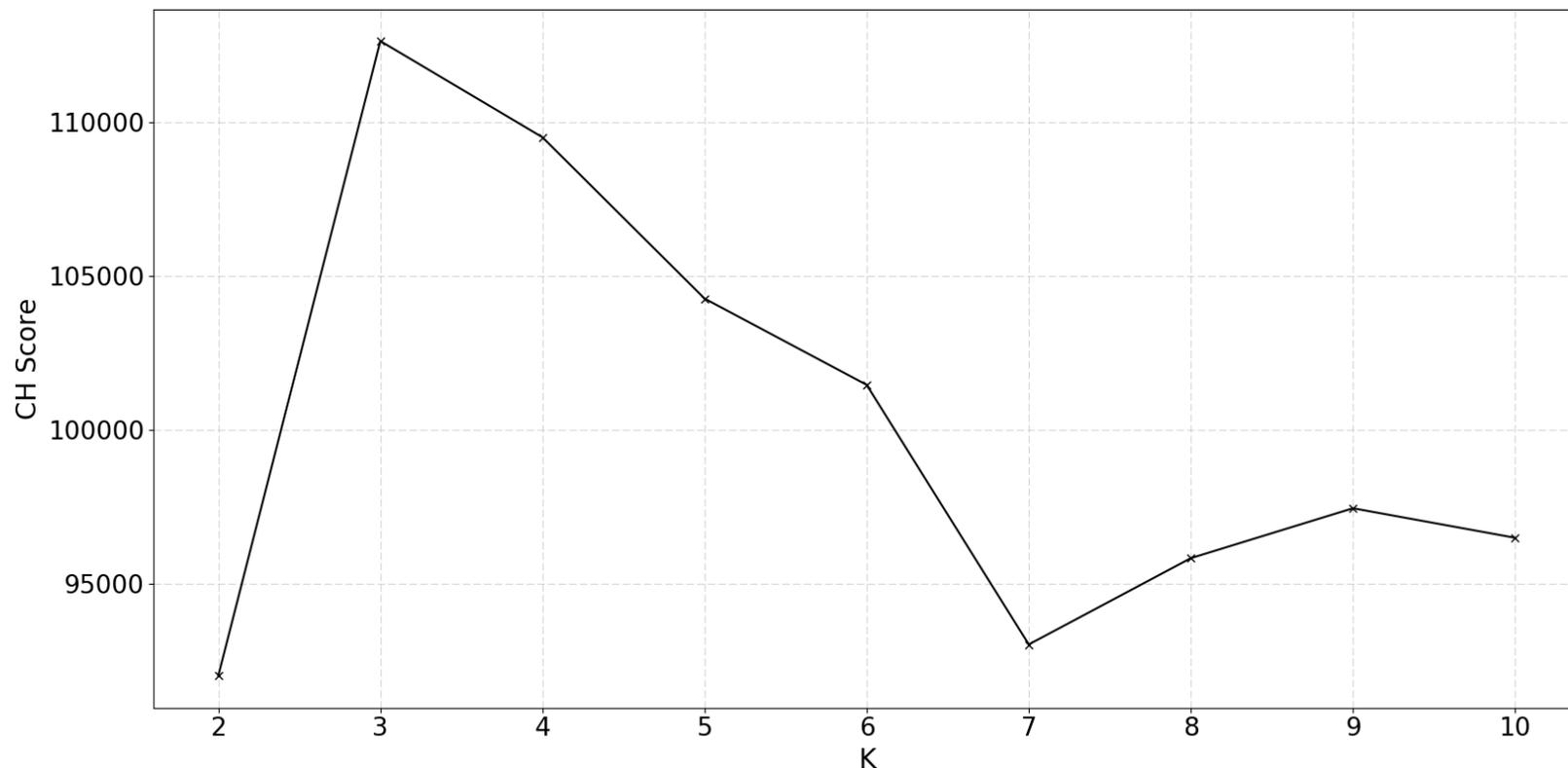
16

- Remoção de anomalias e normalização dos dados
- Aplicação do índice Calinski-Harabasz para definir o K mais adequado
- Aplicação do *K-means* e avaliação dos melhores protótipos pelo SSD
- Aplicação dos algoritmos de classificação para identificar o padrão de novas amostras

Resultados

Resultados – Clusterização

Quantidade ideal de *clusters*



Resultados – Clusterização

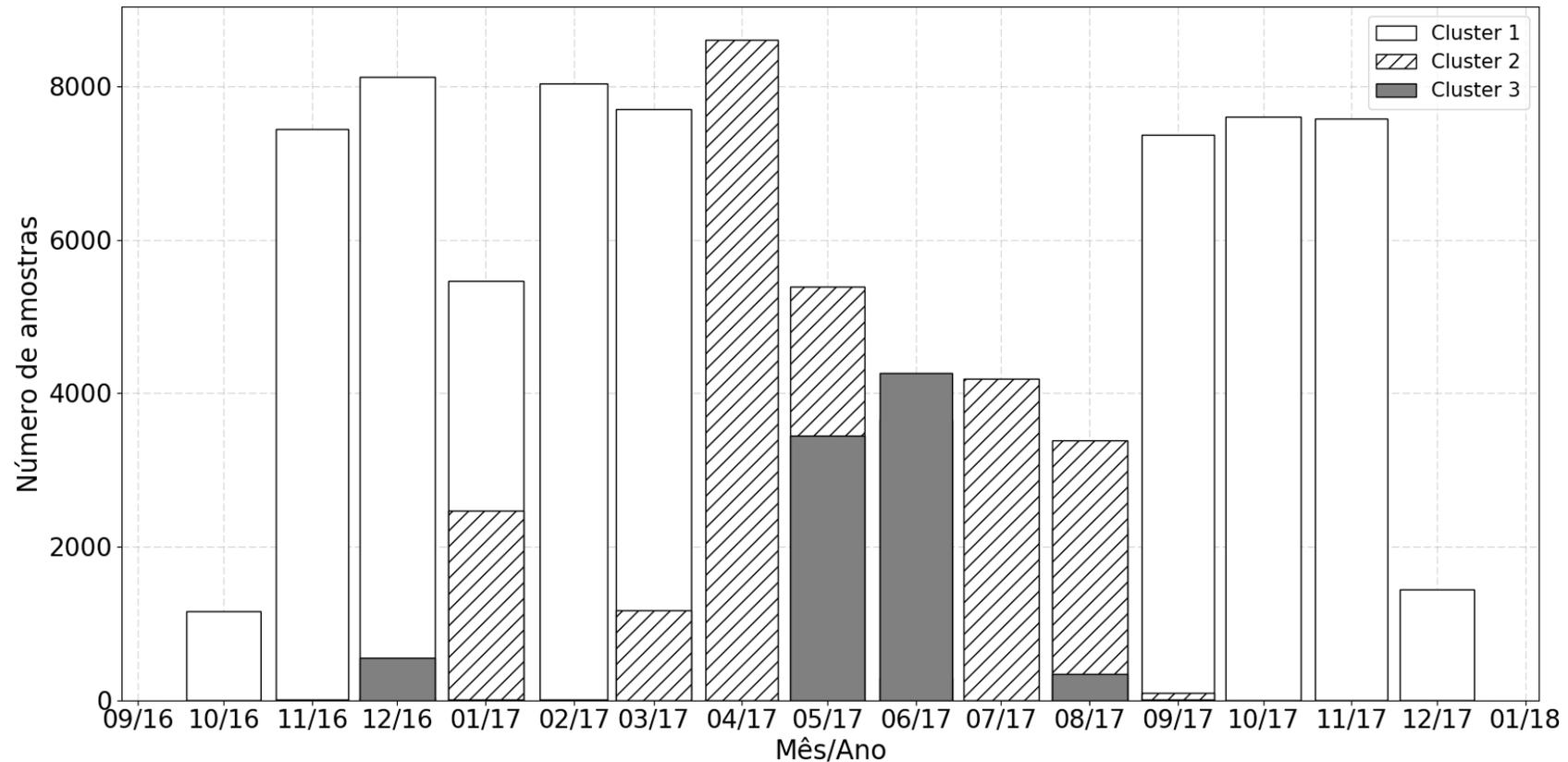
Centroides obtidos

Tabela 1: Estatísticas descritivas dos atributos de cada *cluster*.

		CENTROIDE	MÍNIMO	MÉDIA	MÁXIMO
Cluster 1 64476 amostras	T (°C)	11.97	-2.11	11.98	32.11
	U (%)	85.34	44.00	85.43	103.00
	M (kg)	29.20	11.35	29.21	51.72
Cluster 2 28983 amostras	T (°C)	29.70	10.28	29.68	44.28
	U (%)	54.70	19.00	54.68	95.00
	M (kg)	20.01	0.98	19.89	38.03
Cluster 3 8609 amostras	T (°C)	30.16	10.61	30.30	35.00
	U (%)	66.70	38.00	66.35	100.00
	M (kg)	53.54	35.37	53.42	74.45

Resultados – Clusterização

Frequência dos *clusters* durante o ano



Resultados – Classificação

Matrizes de confusão

Tabela 2: *k*-NN sem Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	624	5	0
	2	4	283	2
	3	0	1	81

Tabela 4: *Random Forest* sem Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	624	4	1
	2	5	284	0
	3	0	0	82

Tabela 6: *Naive Bayes* sem Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	12715	52	128
	2	237	5560	0
	3	0	121	1601

Tabela 3: *k*-NN com Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	626	2	1
	2	4	283	2
	3	0	0	82

Tabela 5: *Random Forest* com Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	625	3	1
	2	5	284	0
	3	0	0	82

Tabela 7: *Naive Bayes* com Cross-Validation

Classe	Predições			
	1	2	3	
Reais	1	12858	28	9
	2	200	5507	90
	3	17	2	1703

Resultados – Classificação

22

Taxas de acerto dos algoritmos

Tabela 8: Taxas de sucesso dos algoritmos de classificação.

Algoritmo	sem <i>cross-validation</i>	com <i>cross-validation</i>
N. Bayes	97.36	98.31
k-NN	98.00	99.10
R. Forest	98.90	99.20

Conclusão

Conclusão

24

Objetivo geral da proposta

- Caracterizar os padrões de colônias de abelhas *Apis mellifera* segundo as mudanças das estações do ano

Principal contribuição

- Implementação de um método que utiliza o índice de validação Calinski-Harabasz e o algoritmo de clusterização *K-means*

Contribuição secundária

- Classificação de novas amostras

Resultados

- Os padrões encontrados foram reconhecidos como coerentes; modelo de classificação apresentou alta precisão

Em perspectiva

- Aplicação da metodologia em dados de colônias de *Apis mellifera* africanizadas

Obrigado!

felipe.anderson@gmail.com
felipe.anderson@alu.ufc.br