

Taxa de Engajamento por Hora em Campanhas

Introdução

Este relatório apresenta uma análise detalhada do código, que realiza o processamento de dados provenientes do arquivo CSV "CampaignQueue.csv" para avaliar métricas de engajamento em campanhas publicitárias ou de comunicação, segmentadas por hora do dia. O estudo abrange etapas de pré-processamento de dados, cálculo de indicadores de desempenho, implementação de algoritmos de otimização (Greedy e A*), comparação de resultados para diferentes valores de L (número de horários selecionados) e geração de uma representação gráfica para o caso L=3.

O objetivo é identificar os intervalos temporais mais eficazes para o envio de campanhas, maximizando o engajamento (definido por leitura ou resposta) enquanto se considera os custos associados (envios sem engajamento).

Metodologia e Estrutura do Código

Variáveis Relevantes:

- Coluna de data/hora: Base para agrupamento por hora.
- Status: Indicador de engajamento (valor '4' ou 'read').
- Response: Campo textual não vazio como critério adicional de engajamento.

Pré-processamento e Cálculo de Indicadores

- Definição de Engajamento: A função `esta_engajado` avalia engajamento com base em status (inteiro '4' ou string 'read') ou presença de resposta não nula.

Agrupamento Temporal: Os dados são agregados por hora, calculando:

- Total de envios (soma de registros).
- Total de engajamentos (soma de casos positivos).
- Custo (diferença entre envios e engajamentos).
- Taxa de engajamento (razão entre engajados e enviados, com tratamento para divisões por zero).

Saída:

Gera uma tabela preenchida com zeros para horas sem registros, assegurando consistência nos dados.

Algoritmos de Otimização

- Algoritmo Greedy: Seleciona os L horários com as maiores taxas de engajamento, ordenando-os em ordem decrescente e subsequentemente por hora crescente, priorizando eficiência local.

- Algoritmo A*: Implementa uma busca heurística A* para identificar o conjunto de L horários que maximiza a função objetivo (engajados - β * custo, com $\beta=0.5$). A heurística considera os melhores engajamentos restantes, garantindo ordem crescente e evitando redundâncias via min-heap.
- Comparação: Para L variando de 1 a 5, computa totais de engajados, custos e taxas médias para ambos os métodos.

Visualização Gráfica

- Cria um gráfico de barras da taxa de engajamento por hora para L=3.
- Representa horários selecionados pelo Greedy com linhas verticais tracejadas.
- Indica horários escolhidos pelo A* com marcadores em diamante.

Resultados

A tabela a seguir sumariza as métricas calculadas para cada hora do dia (00:00 a 23:00). A taxa de engajamento varia entre aproximadamente 0.372 (06:00) e 0.496 (13:00), com picos significativos em 13:00, 15:00 e 17:00, possivelmente refletindo padrões de comportamento dos destinatários.

Hora	Total Enviado	Total Engajado	Custo	Taxa de Engajamento	Label Hora
0	230	98	132	0.426	00:00
1	191	75	116	0.393	01:00
2	211	94	117	0.445	02:00
3	235	109	126	0.464	03:00
4	195	87	108	0.446	04:00
5	212	86	126	0.406	05:00
6	196	73	123	0.372	06:00
7	216	89	127	0.412	07:00
8	199	80	119	0.402	08:00
9	205	75	130	0.366	09:00
10	220	92	128	0.418	10:00
11	224	96	128	0.429	11:00
12	196	79	117	0.403	12:00

13	228	113	115	0.496	13:00
14	243	106	137	0.436	14:00
15	205	97	108	0.473	15:00
16	201	93	108	0.463	16:00
17	198	93	105	0.470	17:00
18	186	77	109	0.414	18:00
19	199	81	118	0.407	19:00
20	194	84	110	0.433	20:00
21	195	82	113	0.421	21:00
22	223	99	124	0.444	22:00
23	198	90	108	0.455	23:00

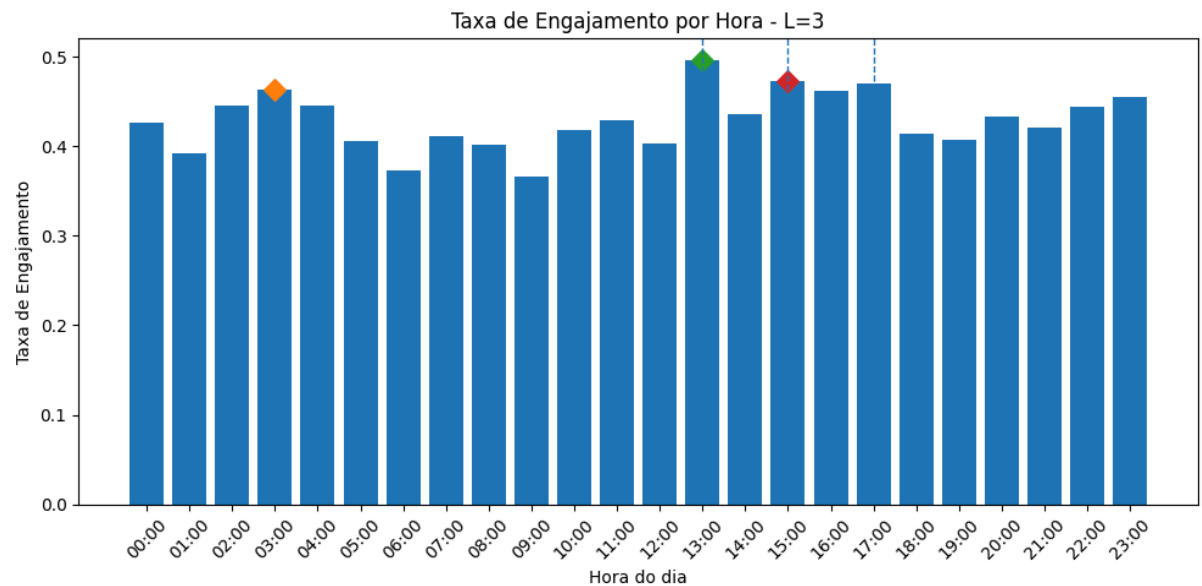
Comparação Greedy vs. A* (L=1 a 5)

A tabela compara os horários selecionados, totais de engajados, custos e taxas médias para ambos os algoritmos. O A* apresenta superioridade em engajados para L=2 e L=3, refletindo sua capacidade de otimizar globalmente, enquanto converge com o Greedy para L>=4.

L	Horas Greedy	Engajados Greedy	Custo Greedy	Taxa Média Greedy	Horas A*	Engajados A*	Custo A*	Taxa Média A*
1	[13:00]	113	115	0.496	[13:00]	113	115	0.496
2	[13:00, 15:00]	210	223	0.484	[03:00, 13:00]	222	241	0.480
3	[13:00, 15:00, 17:00]	303	328	0.479	[03:00, 13:00, 15:00]	319	349	0.478
4	[03:00, 13:00, 15:00, 17:00]	412	454	0.476	[03:00, 13:00, 15:00, 17:00]	412	454	0.476

5	[03:00, 13:00, 15:00, 16:00, 17:00]	505	562	0.473	[03:00, 13:00, 15:00, 16:00, 17:00]	505	562	0.473
---	-------------------------------------	-----	-----	-------	-------------------------------------	-----	-----	-------

Análise do Gráfico (L=3)



O gráfico de barras ilustra a taxa de engajamento por hora, com variação de ~0.3 a ~0.5.

- **Seleção Greedy:** Linhas tracejadas em 13:00, 15:00 e 17:00, refletindo os picos locais de engajamento.
- **Seleção A*:** Marcadores em diamante em 03:00, 13:00 e 15:00, indicando uma estratégia otimizada que incorpora o pico matinal de 03:00. No gráfico fornecido, picos destacados (triângulo laranja em ~03:00, verde em ~13:00, vermelho em ~15:00) corroboram a escolha do A*, sugerindo que a inclusão de horários matinais pode equilibrar engajamento e custo.

Discussão e Análise

- **Desempenho dos Algoritmos:** O A* demonstra superioridade para L=2 e L=3 (ex.: 319 engajados vs. 303), devido à otimização global que pondera custos. Para L>=4, a convergência sugere limitação no dataset ou saturação das melhores opções.
- **Padrões Temporais:** Picos entre 13:00-17:00 e 03:00 podem estar relacionados a fusos horários ou hábitos de uso (ex.: atividade noturna em regiões específicas).
- **Limitações:** A ausência de dados sobre dias da semana ou segmentação geográfica pode mascarar variações adicionais.

Conclusões e Recomendações

- **Resultados Principais:** O A^* é preferível para L pequeno, enquanto ambos os algoritmos convergem para L maior. Horários como 13:00, 15:00 e 03:00 são ideais para maximizar engajamento.