

Obtenção de rota representativa para navios atracando no Porto do Rio de Janeiro

Com a crescente monitoração de diversas atividades, a quantidade de informação disponível para análise tem aumentado significativamente, propiciando um novo paradigma de descoberta de padrões apoiado em algoritmos de *data mining*. Na área de navegação marítima, graças a introdução e uso obrigatório do AIS pelos navios de maior porte, não é diferente. Enormes quantidades de dados de posição podem ser facilmente acumulados sobre a movimentação dos navios numa região. Com a recente integração do AIS com satélite e a interconexão de fontes de dados via Internet, não só a movimentação costeira mas também boa parte da movimentação global de navios também pode ser analisada em detalhe para identificar padrões recorrentes, acompanhar e estimar movimentações futuras e detectar comportamentos atípicos. Esse estudo de caso investiga uma possibilidade de aplicação de um desses algoritmos: o TraClus através de uma extensão implementada no Sistema de Tráfego Aquaviário (STAQ) desenvolvido pela CASH Computadores e Sistemas.

Suponhamos uma base de dados contendo os relatos de posição AIS acumulados ao longo de um período numa região de interesse. Seleccionemos as sequencias de posições de navios de grande porte entrando na Baía de Guanabara e atracando no Porto do Rio de Janeiro. Considerando o conjunto de trajetórias descritas por esses rastros, pergunta-se: existe algo em comum entre elas? Seria possível simplificar o complexo conjunto de trajetórias substituindo-o por um punhado de outras linhas bem mais simples capazes de capturar a tendência geral dos deslocamentos desses navios na baía? Qual o caminho mais comum ou “rota representativa” que um navio desses segue da entrada da baía até o porto e qual a velocidade média ao longo dos principais trechos? Considerando essa “rota representativa”, qual o tempo estimado que um novo navio compatível com os navios estudados e que acaba de chegar à entrada da baía deve levar até o porto?

Em *Data Mining* [1], algoritmos de *clustering* em geral procuram agrupar conjunto de dados em múltiplos grupos ou *clusters* tal que objetos num *cluster* tem uma alta similaridade, mas são distintos de objetos em outros *clusters*. Entre esses algoritmos podemos destacar os algoritmos de *trajectory clustering* que operam sobre trajetórias ou partes de trajetórias.

O TraClus [3] é um algoritmo bastante conhecido na área e já com diversas variações propostas, inclusive uma recente para rotas de navios [4]. Aqui, nos deteremos sobre sua formulação original de 2007 e no final, introduziremos uma pequena extensão.

Em linhas gerais, o TraClus recebe como entrada um conjunto de trajetórias, simplifica e particiona essas trajetórias em segmentos, descobre *clusters* de segmentos e finalmente retorna um conjunto simplificado de trajetórias representativas calculadas sobre cada um dos *clusters* de segmentos. Ele é baseado no algoritmo consagrado DBSCAN [2] e, como este último, também encontra *clusters* extensos, pois ele inclui tanto os segmentos diretamente quantos os indiretamente atingíveis a partir do segmento identificado como núcleo em cada *cluster*.

Uma vantagem muito interessante desse algoritmo é que ele é capaz de identificar subtrajetórias, ou seja, porções das trajetórias suficientemente próximas, ainda que as trajetórias não sigam próximas o tempo todo. Na verdade, ele particiona previamente as trajetórias e opera sobre seus segmentos de pares de pontos. Depois de reunir os segmentos em *clusters*, ele traça uma trajetória representativa composta de novos pontos calculados usando uma linha de varredura perpendicular à direção média dos segmentos reunidos de cada *cluster* e calculando uma coordenada média entre as interseções conseguidas, conforme figura 1. Diferente de um mapa de calor – outra ferramenta muito interessante para identificar

padrões visuais - o Traclus retorna uma trajetória que ainda pode ser trabalhada numericamente, por exemplo, transformada em rota.

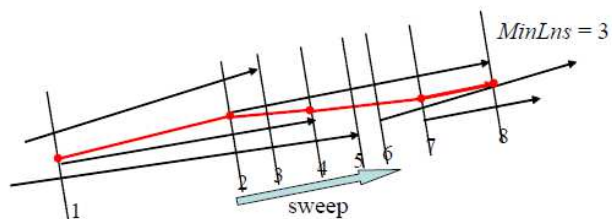


Fig. 1. Um *cluster* de segmentos é identificado e a trajetória representativa no *cluster* é calculada varrendo os segmentos no *cluster* com uma perpendicular a direção média e calculando os pontos médios das interseções desses segmentos com a perpendicular

O TraClus admite dois parâmetros que regulam seu funcionamento: ***eps*** ou distância máxima de reunião dos segmentos de um *cluster*, e ***minlns*** ou quantidade mínima de segmentos para formar um *cluster*. Quanto menor o *eps*, maior a proximidade exigida dos segmentos para considerá-los solidários no mesmo *cluster*. Quanto maior o *minlns*, maior a quantidade de segmentos exigida para caracterizar um *cluster*.

Outra característica interessante do algoritmo é que ele inicia simplificando as trajetórias de entrada, diminuindo significativamente a quantidade de segmentos que serão analisados, o que o deixa bem rápido, pois ele opera sobre um conjunto de segmentos representativo, mas bem reduzido se comparado ao conjunto de entrada.

A seguir, vamos descrever o uso que fizemos do algoritmo TraClus aplicado sobre os rastros de navios no STAQ, nas seguintes etapas:

1. Busca preliminar de rastros candidatos segundo um critério de interesse
2. Descoberta dos *clusters* de segmentos de rastros
3. Extração das trajetórias representativas para cada *cluster* de segmentos de rastros
4. Plotagem das trajetórias representativas sobre a carta
5. Transformação de uma trajetória representativa em rota representativa
6. Estimativa de ETA para um novo navio ao longo da rota representativa

Adotamos o princípio que os melhores resultados serão obtidos na combinação da expertise do analista humano com a capacidade de processamento e descoberta de padrões da automação. Por isso achamos importante que algoritmo já opere sobre uma seleção preliminar de rastros candidatos, de alguma forma, já relacionados a uma categoria de deslocamento sendo estudada (navios atracando no Porto do Rio de Janeiro) e não uma amostra qualquer de rastros.

Iniciamos o roteiro previsto abrindo a janela de Busca de Rastros do STAQ e selecionando rastros que satisfaçam as seguintes condições:

- Navios de grande porte (exclui, por exemplo, os rebocadores)
- Duração máxima do rastro até duas horas
- Posições AIS coletadas durante 8 dias na Baía de Guanabara (16/7 a 27/7/19)
- Rastro inicia com navio entrando na baía (retbg)
- Rastro termina com navio entrando e parando na área do Porto do Rio (areaporto)

Nessa busca, usamos duas áreas auxiliares: uma área retangular maior para marcar a região da baía (retbg) e permitir detectar a entrada na baía, assim como uma outra área circular para permitir detectar a chegada à área do porto (areaporto) conforme ilustrado na figura 2.

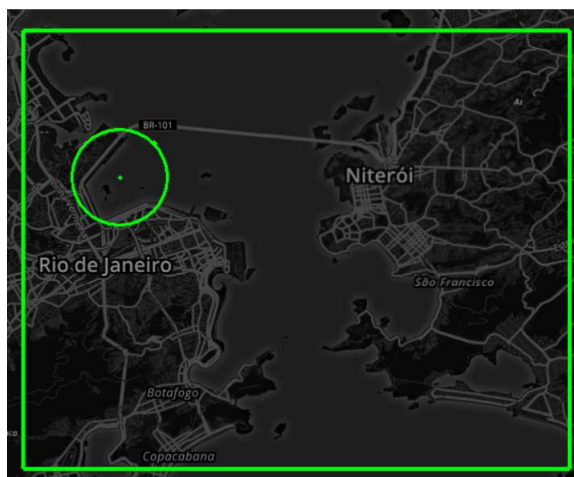


Fig. 2. Áreas auxiliares usados na busca de rastros: uma retangular para delimitar entrada na Baía de Guanabara e outra circular já na região do porto

Rodamos a busca preliminar de rastros e encontramos 21 rastros com um total de 1627 segmentos conforme a tabela 1.

MMSI	NAVIO	TIPO	COMPR	DURAÇÃO	DISTÂNCIA	PONTOS
259022000	AKER WAYFARER	90 - OUTRO	157 m	1.5 h	8.6 M	114
636016986	ANTHEA Y	7x - CARGO	300 m	1.1 h	8.6 M	64
432783000	ATLAS LEADER	7x - CARGO	199 m	1.5 h	8.9 M	56
305466000	BBC CONGO	7x - CARGO	142 m	1.0 h	8.2 M	81
305152000	BBC DOLPHIN	7x - CARGO	120 m	1.2 h	9.0 M	74
710014030	BRAM BUCCANEER	7x - CARGO	93 m	1.3 h	8.7 M	85
710024720	BRAM BUCK	7x - CARGO	93 m	1.3 h	8.6 M	66
710003790	CASTILLO DE HERRERA	8x - TANKER	140 m	1.2 h	8.3 M	74
710069000	CBO GUANABARA	99 - OUTRO	77 m	1.8 h	8.9 M	136
710025850	CBO OCEANIA	7x - CARGO	88 m	1.5 h	8.4 M	93
710004920	CD SERRA DA ESTRELA	8x - TANKER	65 m	1.5 h	4.6 M	78
636017202	CHRISTOS	7x - CARGO	199 m	1.0 h	7.6 M	58
636091645	GRAL. SAN MARTIN	7x - CARGO	176 m	1.3 h	9.0 M	115
236542000	GRANDE AFRICA	7x - CARGO	214 m	1.6 h	8.9 M	120
247078800	GRANDE FRANCIA	7x - CARGO	214 m	1.5 h	8.9 M	76
257406000	ISLAND PRIDE	90 - OUTRO	103 m	1.3 h	8.7 M	62
636090866	LETAVIA	7x - CARGO	220 m	59 min	8.3 M	95
710003840	LOG IN PANTANAL	7x - CARGO	182 m	1.0 h	8.2 M	56
710003840	LOG IN PANTANAL	7x - CARGO	182 m	51 min	7.5 M	49
477305900	MSC ANTIGUA	7x - CARGO	299 m	1.3 h	8.8 M	53
256858000	MSC ATHENS	7x - CARGO	300 m	1.2 h	8.6 M	43

Tabela 1. Rastros previamente selecionados no STAQ

Na figura 3, plotamos alguns dos rastros encontrados. Note que entre os rastros plotados, um deles (rastros 7) representa uma outra trajetória bem diversa das demais e foi selecionado porque o navio também entra no retângulo auxiliar mas “por cima” e não representa uma entrada na baía. O STAQ permite descartá-lo previamente.

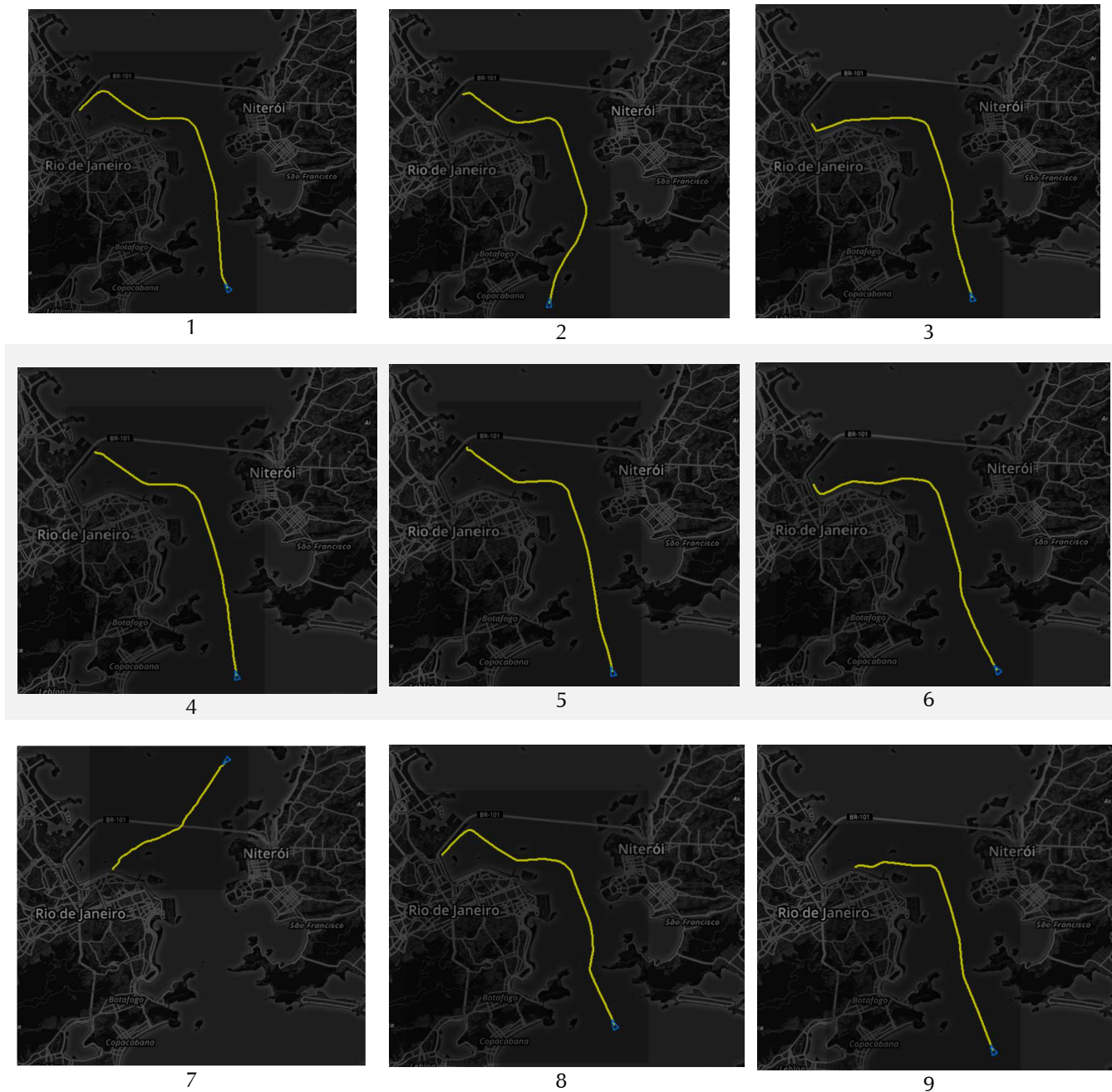


Fig. 3. Alguns rastros encontrados sendo exibidos na carta

Rodamos o TraClus com: $eps = 0,4 M$ (740 metros) e $minLns = 10$, ou seja, a distância de *clusterização* como 740 metros e exigimos ao menos 10 segmentos para formar um *cluster*. A título de comparação a largura da boca da barra, onde a baía é mais estreita e por onde os navios entram e saem, é de aproximadamente 1600 metros, então nossa distância é aproximadamente metade disso.

O algoritmo inicia reduzindo as trajetórias dos rastros imputados, ficando com apenas 117 segmentos. Em seguida, identifica dois *clusters* de segmentos: um contendo 25.8% dos segmentos e o outro com 2.6%. A tabela 2 mostra alguns detalhes dos resultados obtidos.

CLUSTER	% SEGMENTOS	TRAJETÓRIA REPRESENTATIVA	VMED	TEMPO ESTIMADO
1	25.8%	29 pontos (6.8 M)	8.7 nós	46 min
2	2.6%	3 pontos (1.0 M)	5.0 nós	12 min

Tabela 2. *Clusters* encontrados

A figura 4 ilustra as trajetórias representativas calculadas para cada *cluster* pelo algoritmo.

A trajetória representativa mais longa calculada sobre o primeiro *cluster* contém 29 pontos e captura o deslocamento dos navios da entrada da baía até o porto. A outra trajetória representativa calculada sobre o segundo *cluster* captura um “desvio” em relação à primeira que alguns navios fazem, já nas proximidades do porto, em direção a um berço alternativo.

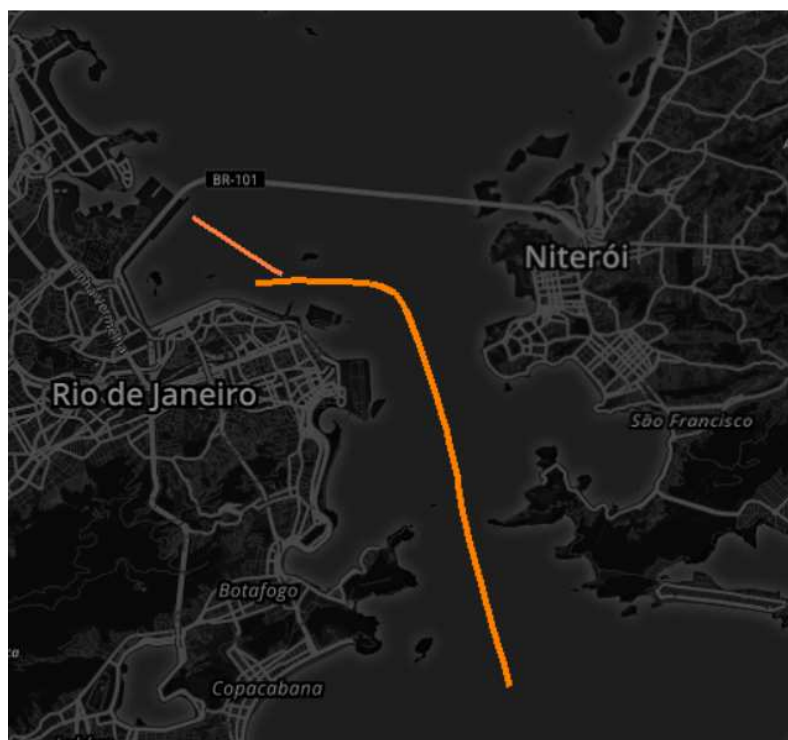


Fig. 4. As duas trajetórias representativas calculadas pelo algoritmo: uma mais longa desde a entrada da baía até o porto. A outra, um desvio já no porto

Uma adaptação que julgamos oportuna fazer na implementação do TraClus foi calcular uma velocidade média para cada um dos pontos da trajetória representativa do *cluster*. Isso foi feito calculando para cada ponto da trajetória representativa a velocidade média dos pontos dos segmentos do *cluster* usados no cálculo daquele ponto da trajetória.

A tabela 3 mostra as velocidades para cada um dos 29 pontos da trajetória representativa mais longa, incluindo o número de segmentos (de rastros distintos) NSEGS envolvidos para calcular cada ponto. Além da velocidade média incluímos o seu desvio padrão e ele dá uma medida da semelhança dos segmentos de rastro naquele trecho. Os desvios padrão das velocidades estão bem baixos, o que sugere que possamos tomar com boa confiança aquela velocidade como representativa no trecho.

ID	LAT	LON	SOG	NSEGS
1	22°58,334'S	043°07,777'W	8.1 +/- 0.0	10
2	22°58,195'S	043°07,812'W	7.7 +/- 1.0	11
3	22°58,035'S	043°07,864'W	8.4 +/- 0.2	13
4	22°57,612'S	043°07,997'W	8.7 +/- 0.8	13
5	22°57,105'S	043°08,144'W	8.9 +/- 1.1	13
6	22°56,559'S	043°08,309'W	9.0 +/- 0.1	14
7	22°56,313'S	043°08,382'W	9.2 +/- 0.6	15
8	22°56,133'S	043°08,421'W	9.2 +/- 0.0	14
9	22°55,912'S	043°08,460'W	9.3 +/- 0.2	15
10	22°55,744'S	043°08,490'W	8.6 +/- 0.2	16
11	22°55,586'S	043°08,521'W	9.0 +/- 0.6	17
12	22°55,354'S	043°08,576'W	8.9 +/- 0.1	18
13	22°55,171'S	043°08,632'W	8.9 +/- 0.6	18
14	22°55,046'S	043°08,666'W	9.1 +/- 0.9	19
15	22°54,924'S	043°08,706'W	9.0 +/- 0.3	19
16	22°54,666'S	043°08,791'W	9.2 +/- 0.6	20
17	22°54,498'S	043°08,848'W	9.3 +/- 0.2	20
18	22°54,327'S	043°08,908'W	9.3 +/- 0.7	20
19	22°54,107'S	043°08,993'W	9.3 +/- 0.0	20
20	22°53,706'S	043°09,159'W	9.3 +/- 0.2	20
21	22°53,602'S	043°09,217'W	9.1 +/- 0.2	20
22	22°53,499'S	043°09,316'W	9.1 +/- 0.1	20
23	22°53,411'S	043°09,472'W	8.7 +/- 0.2	20
24	22°53,381'S	043°09,633'W	8.1 +/- 0.5	20
25	22°53,365'S	043°09,864'W	8.0 +/- 0.2	20
26	22°53,349'S	043°10,476'W	7.7 +/- 0.4	20
27	22°53,347'S	043°10,682'W	7.6 +/- 0.4	18
28	22°53,365'S	043°10,948'W	7.3 +/- 0.4	11
29	22°53,376'S	043°11,171'W	7.3 +/- 0.4	10

Tabela 3. Velocidades ao longo da trajetória representativa

No STAQ, podemos transformar a linha correspondente à trajetória representativa em rota, considerando a velocidade diferenciada ponto a ponto ou usando a velocidade média total da trajetória, ou ainda, usando a velocidade instantânea e atual do navio sendo acompanhado. Nesse último caso, estaríamos aproveitando apenas a informação do formato histórico da trajetória.

Assim que um novo navio, compatível com as condições que usamos para achar a rota representativa, chega na entrada da baía, podemos imaginá-lo navegando na rota representativa e estimar seu ETA de chegada ao porto, conforme janela Rota do Rastreador e ilustrado na figura 5.

Em relação à pertinência de considerar ou não aquele novo navio como seguindo aquela rota representativa encontramos na literatura alguns trabalhos interessantes [5][6] que propõe o cálculo de uma probabilidade dessa pertinência. São cálculos complexos que procuram comparar as posições, velocidades e rumos mais recentes do novo navio com o histórico na região da rota. Usando uma estimativa com funções kernel calculam uma probabilidade daquele navio também seguir a rota que os anteriores seguiram. Pretendemos no futuro implementar e investigar essa ideia no STAQ. Por ora, contamos com o expertise do analista para reavaliar periodicamente essa pertinência à medida que o navio prossegue.

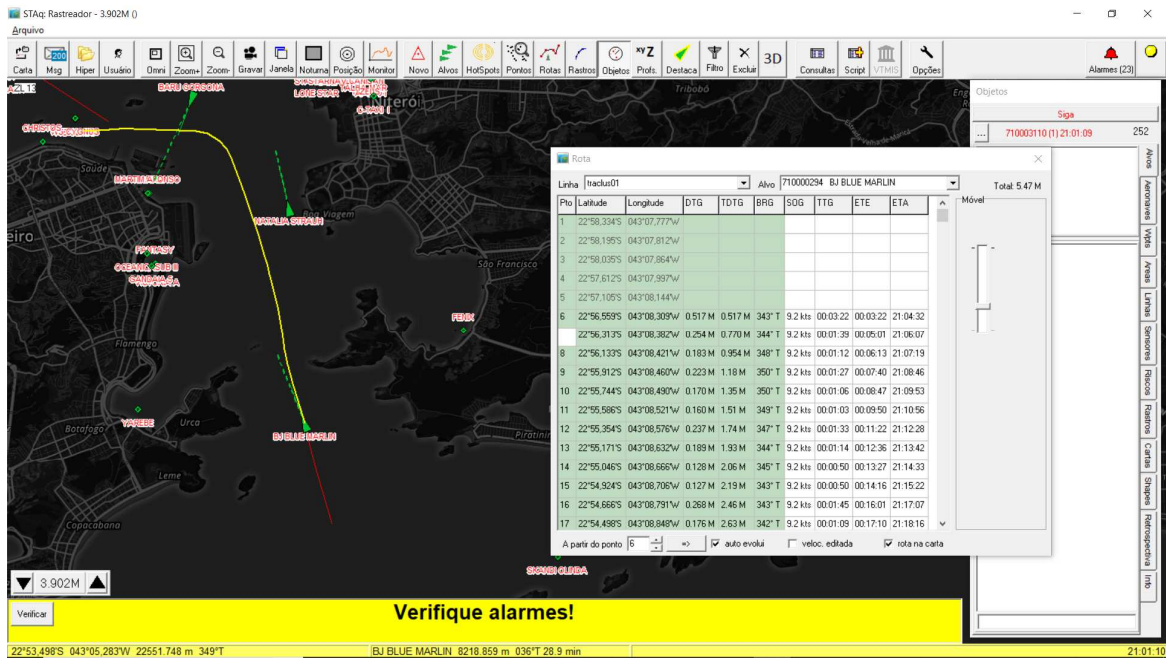
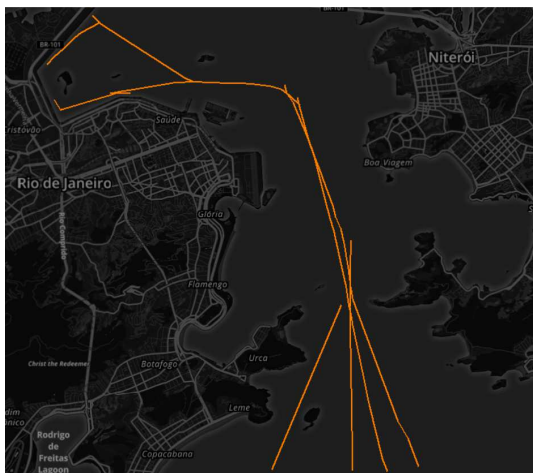
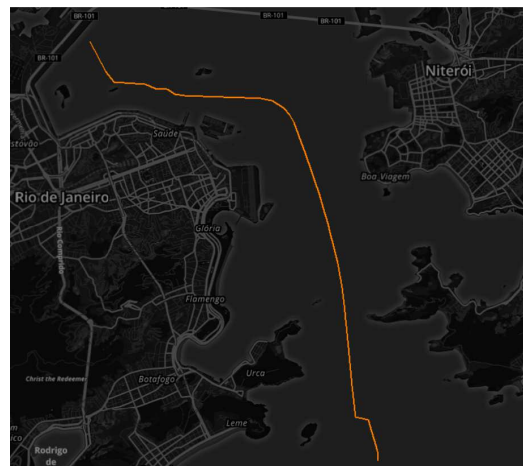


Fig. 5. Acompanhando e estimando chegada de um novo navio através da trajetória representativa transformada em rota no Rastreador

Para melhor entendimento do potencial desse algoritmo, verifiquemos o efeito de variar os parâmetros de entrada, conforme ilustra a figura 6.



eps = 0,25M (460 m), minIns= 2 segmentos
7 trajetórias



eps = 1 M (1850 m), minIns = 15 segmentos
1 trajetória

Fig. 6. Análise do efeito da variação dos parâmetros

Em resumo, podemos ajustar esses dois parâmetros conforme a tabela a seguir:

	eps (distância máxima entre segmentos)	minlns (mínimo de segmentos)
menor	mais trajetórias porém mais curtas	mais trajetórias, algumas pouco frequentadas
maior	menos trajetórias porém mais longas	menos trajetórias, porém mais frequentadas

Tabela 4. Resumo do efeito da variação dos parâmetros

Nesse texto, procuramos avaliar o uso de um algoritmo de *trajectory clustering* através de um estudo de caso em pequena escala. Acreditamos que com isso conseguimos ilustrar o potencial dessas técnicas como ferramentas para descoberta de padrões que podem ser usados para prever movimentações futuras (e também comportamentos atípicos) dos navios a partir da extensa base coletada de posições AIS e, quem sabe, inspirar outros estudos nessa linha.



Autor: Mauricio Peixoto Pacheco da Rocha
Engenheiro de Computação formado pelo Instituto Militar de Engenharia (IME) e Mestre em Engenharia de Sistemas pela COPPE/UFRJ. Engenheiro de software responsável pelo sistema STAQ

mauricio.peixoto@hotmail.com

Referências:

- [1] Data mining: concepts and techniques, 3rd edition, Morgan-Kaufmann 2011
- [2] Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering *clusters* in large spatial databases with noise. In Proceedings of Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA , 2-4 August 1996; pp. 226-231.
- [3] Trajectory *Clustering*: A Partition-and-Group Framework, Jae-Gil Lee, Jiawei Han and Kyu-Young Whang, SIGMOD´07, June 11-14, 2007, Beijing, China
- [4] Extracting Shipping Routing Patterns by Trajectory *Clustering* Model Based on Automatic Identification System Data, Pan Sheng and Jingbo Yin, Sustainability 2018
- [5] R. Laxhammar, G. Falkman, E. Sviestins. Anomaly detection in sea traffic - a comparison of the Gaussian Mixture Model and the Kernel Density Estimator. 12th International Conference on Information Fusion, Seattle, WA, USA, 6-9 July. 2009.
- [6] Giuliana Pallotta, Michele Vespe and Karna Bryan. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. NATO Science and Technology Organization (STO), Centre for Maritime Research and Experimentation (CMRE), Viale San Bartolomeo 400, 19126, La Spezia, Italy