

ANÁLISE DE MÉTRICAS PARA RECUPERAÇÃO DE CASOS VISANDO À APLICAÇÃO NO PROBLEMA DE CONDUÇÃO DE TRENS DE CARGA

Aluno: Matheus Streisky
Orientador: Dr. André Pinz Borges

O modelo Ferroviário

- Importância do modal férreo:
 - *Grande volume transportado;*
 - *Redução do tráfego rodoviário;*
 - *Menor consumo de combustível comparado ao modal rodoviário.*

O modelo Ferroviário

- Importância do modal férreo:
 - *Grande volume transportado;*
 - *Redução do tráfego rodoviário;*
 - *Menor consumo de combustível comparado ao modal rodoviário.*

- Dificuldade existentes:
 - *Escolher a ação a ser executada (acelerar, frear, etc);*
 - *Escolher o ponto de aceleração (marcha) a ser usado em cada momento da estrada;*
 - *Conduzir com segurança e sem danos à via e ao trem;*
 - *Economizar combustível → menor impacto ambiental.*

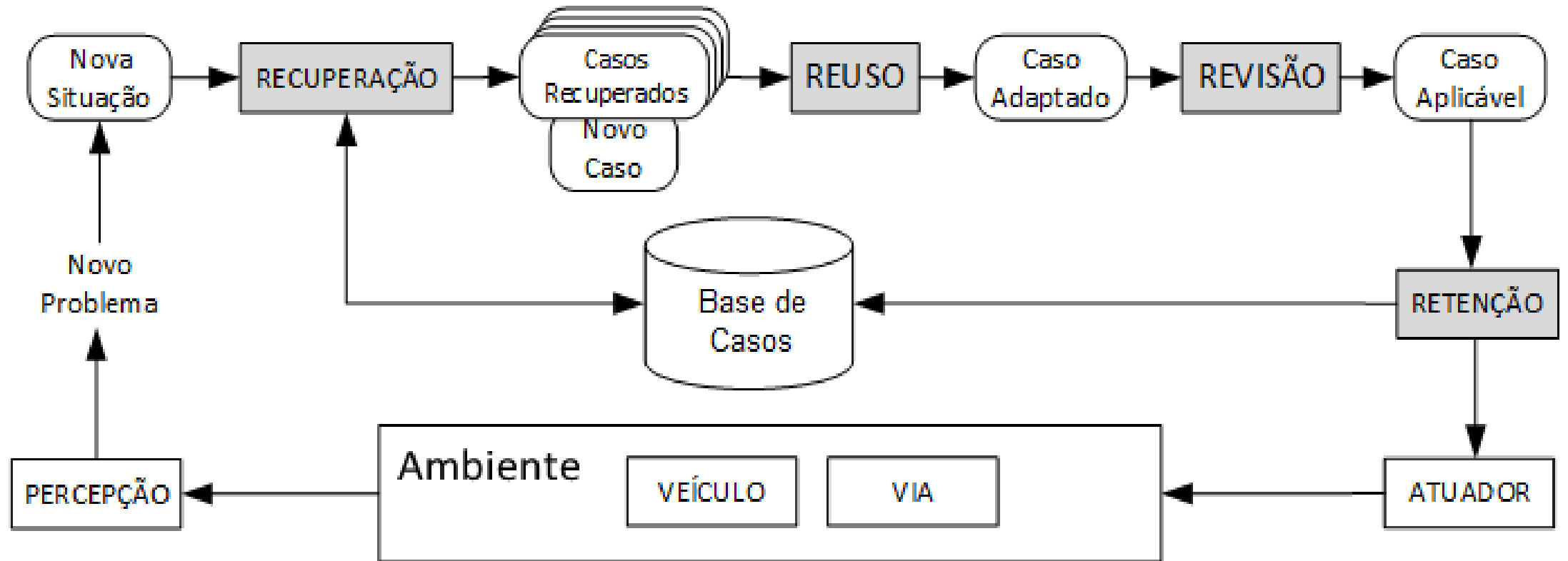
Raciocínio Baseado em Casos (RBC)

- Hipótese: problemas similares possuem soluções similares.
- Consiste em utilizar situações já registradas em memória para solucionar novos problemas.
- Exemplo:
 - *Um médico ao atender um paciente.*

Caso

- É todo o conhecimento armazenado em memória pelo método.
- São “memórias” de um problema que foi total ou parcialmente resolvido.
- Caso = características do problema + solução adotada.

O ciclo RBC



Fonte: O ciclo RBC (Borges, 2015).

Recuperação de Casos

- Dado determinado problema, o objetivo é recuperar casos (de uma base de casos) que mais se assemelham à situação atual.
- Base de casos: conjunto de casos lidos de históricos de viagens.
- A recuperação também deve ser precisa e eficiente.

Como saber se uma recuperação foi eficiente?

- Menos esforços empregados na etapa de Reuso.
 - *Quantidade menor de alterações do caso recuperado na etapa de Reuso.*

Como saber se uma recuperação foi eficiente?

- Menos esforços empregados na etapa de Reuso.
 - *Quantidade menor de alterações do caso recuperado na etapa de Reuso.*

- No problema de condução de trens (avaliado na etapa de Revisão):
 - *Não ocorrer patinação do trem.*
 - *Menos mudanças de marcha.*
 - *Menor consumo de combustível.*

Exemplo de Recuperação de Casos

- Caso problema:

Vel	Vel Final	Vel Max	% Rampa	Solução
2	5	35	0,2	?

Exemplo de Recuperação de Casos

■ Caso problema:

Vel	Vel Final	Vel Max	% Rampa	Solução
2	5	35	0,2	?

■ Casos Recuperados:

Caso	Vel	Vel Final	Vel Max	% Rampa	Solução	Índice Dissimilaridade
1	3	7	40	0,3	(5, 0), (2, 10)	5,4781
2	4	6	45	0	(7, 9), (1, 10)	10,2489

Algoritmos de Recuperação de Casos

- Algoritmos:

- *busca serial (Simoudis et al, 1993);*
- *busca hierárquica (Maher e Zhang, 1991);*
- *busca por simulação paralela (Domeshek, 1993).*

K-NN (k-Nearest Neighbor)

- Geralmente é utilizado para agrupamento de dados.
- k é o número de casos recuperados ao problema atual.
- k -NN é um dos algoritmos de recuperação mais difundidos pela literatura, pois possui uma fácil implementação (Zia et al, 2014).
- Para seleção dos vizinhos mais próximos, é utilizado como métrica de dissimilaridade a Distância Euclidiana.

Métricas de similaridade

- Podem influenciar se a recuperação de casos será boa ou ruim.
- Não há um consenso sobre uma medida de similaridade que seja aplicável a todos os tipos de variáveis que podem existir numa base de casos.
- Geralmente são empregadas medidas variadas como:
 - *Distância Euclidiana;*
 - *Distância Euclidiana Ponderada;*
 - *Distância de Manhattan;*
 - *Distância de Minkowski.*

Métricas

Distância Euclidiana:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Distância Euclidiana
Ponderada:

$$d = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2}$$

x e y representam os atributos dos casos a serem comparados
 w é o peso atribuído à esses atributos

Métricas

Distância de *Manhattan*:

$$d = \sum_{i=1}^n |x_i - y_i|$$

Distância de *Minkoski*:

$$d = \left(\sum_{i=1}^n |x_i - y_i|^q \right)^{\frac{1}{q}}$$

Distância de *Chebyshev*:

$$d = \max_k |x_k - y_k|$$

x e y representam os atributos dos casos a serem comparados
 q é o grau de sensibilidade da métrica

Experimentos

- Foi assumido os valores de $k=3$ e $k=5$ para um problema fictício a ser solucionado, com os seguintes valores:

[5.6, 3.7, 5.2, 0.4]

- Obs: Recomenda-se que k possua um valor ímpar, assim a ocorrência de empates se tornam mais difícil de ocorrer.

Experimentos

- Foi utilizada uma base de casos contendo 150 elementos com informações simuladas de condução de trens de carga.
- Os casos contidos nesta base possuem 4 tipos de informações triviais sobre condução (dados fictícios, mas dentro da faixa aceitável):
 - *Velocidade;*
 - *Pressão dos freios;*
 - *Peso total dos vagões (em toneladas);*
 - *Inclinação (subida, descida ou reta).*

Experimentos

- Para utilizar a Distância Euclidiana e da Distância de *Manhattan*, não há outros fatores a serem declarados.
- Na Distância Euclidiana Ponderada, será assumido os seguintes valores de w (peso) em cada uma das características dos casos:
 - 4 para velocidade,
 - 5 para a pressão dos freios,
 - 3 para peso dos vagões e
 - 1 para grau de inclinação.
- Na Distância de *Minkowski*, q (grau de sensibilidade) assumirá valor igual ao infinito.

Resultados

Tabela 2 - Resultados obtidos com k=3.

Métrica	k = 3	
	Casos	d
Distância Euclidiana	[6,1; 2,8; 4,7; 1,2]	1,396
	[6,1; 3,0; 4,6; 1,4]	1,449
	[5,7; 3,0; 4,2; 1,2]	1,449
Distância Euclidiana Ponderada	[6,0; 3,4; 4,5; 1,6]	2,000
	[5,9; 3,0; 5,1; 1,8]	2,190
	[5,4; 3,0; 4,5; 1,5]	2,300
Distância de Manhattan	[5,6; 3,0; 4,5; 1,5]	2,500
	[5,9; 3,0; 5,1; 1,8]	2,500
	[6,0; 3,4; 4,5; 1,6]	2,600
Distância de Minkowski	[5,7; 2,8; 4,5; 1,3]	0,900
	[6,1; 2,8; 4,7; 1,2]	0,900
	[6,4; 2,9; 4,3; 1,3]	0,900

Resultados

Métrica	k = 5	
	Casos	d
Distância Euclidiana	[6,1; 2,8; 4,7; 1,2]	1,396
	[6,1; 3,0; 4,6; 1,4]	1,449
	[5,7; 3,0; 4,2; 1,2]	1,449
	[5,4; 3,0; 4,5; 1,5]	1,493
	[5,7; 2,9; 4,2; 1,3]	1,568
Distância Euclidiana Ponderada	[6,0; 3,4; 4,5; 1,6]	2,000
	[5,9; 3,0; 5,1; 1,8]	2,190
	[5,4; 3,0; 4,5; 1,5]	2,300
	[6,1; 3,0; 4,6; 1,4]	2,351
	[6,0; 3,0; 4,8; 1,8]	2,351
Distância de Manhattan	[5,6; 3,0; 4,5; 1,5]	2,500
	[5,9; 3,0; 5,1; 1,8]	2,500
	[6,0; 3,4; 4,5; 1,6]	2,600
	[5,6; 3,0; 4,1; 1,3]	2,700
	[5,5; 2,6; 4,4; 1,2]	2,799
Distância de Minkowski	[5,7; 2,8; 4,5; 1,3]	0,900
	[6,1; 2,8; 4,7; 1,2]	0,900
	[6,4; 2,9; 4,3; 1,3]	0,900
	[6,2; 2,9; 4,3; 1,3]	0,900
	[5,6; 2,7; 4,2; 1,3]	1,000

Tabela 3 - Resultados obtidos com k=5.

Resultados

- Os 3 primeiros casos recuperados não diferem quando o k assume valor igual a 3 ou igual a 5.
- Métricas diferentes recuperaram casos iguais, porém em diferentes prioridades.
 - *Caso 1 – [6.0, 3.4, 4.5, 1.6]:*
 - 1º mais similar pela Distância Euclidiana Ponderada.
 - 3º mais similar pela Distância de *Manhattan*.
 - *Caso 2 – [6.1, 2.8, 4.7, 1.2]:*
 - 1º mais similar pela Distância Euclidiana.
 - 2º mais similar pela Distância de Minkowski.

Resultados

Tabela 4 - Médias de dissimilaridade.

Métrica	k = 3	k = 5	Média (k=3 e k=5)
Distância Euclidiana	1,431	1,471	1,451
Distância Euclidiana Ponderada	2,163	2,238	2,201
Distância de Manhattan	2,533	2,620	2,577
Distância de Minkowski	0,900	0,920	0,910

Considerações Finais

- O método RBC é uma ferramenta versátil e pode ser aplicada para diversas funcionalidades, podendo ser moldada de acordo com as necessidades de quem a desenvolve.
- A recuperação de casos é uma etapa importante em todo método RBC, pois esta etapa pode determinar se o problema enfrentado terá uma boa solução ou não.
- As métricas de dissimilaridade aplicadas no problema também impactam no desempenho do método.
 - *Neste artigo a diferença entre a melhor e pior técnica foi de 1,667 pontos de dissimilaridade.*

Referências

- Aamodt, A.; Plaza, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. In: AICOM – Artificial Intelligence Communications. Vol. 7. Ed. IOS Press, EUA, 1994.
- Borges, A., P. Uma contribuição para geração de políticas de ações para condução de trens de carga usando Raciocínio Baseado em Casos. 2015. 220 f. Dissertação (Doutorado) – Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná. Curitiba, 2015.
- Bricce, B. R., Silva, P. P., Martins, H. P., Silva, E. G. . Aplicação do algoritmo knn para controle de movimentos de npcs em um ambiente dinâmico (jogo). REGRAD - Revista Eletrônica de Graduação do UNIVEM - ISSN 1984-7866, [S.l.], v. 9, n. 1, p. 18-32, Aug. 2016. ISSN 1984-7866.

Referências

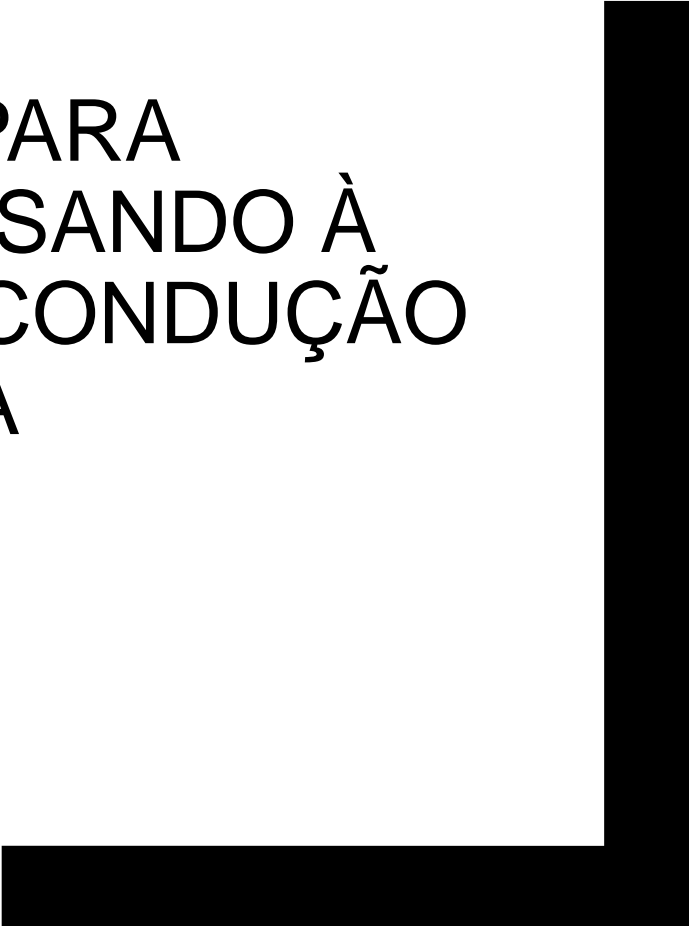

- Cole, R., M.. Clustering with Genetic Algorithms, M. Sc., Department of Computer Science, University of Western Australia, Australia, 1998.
- Corchado, J. M., Lees, B., Aiken, J.. Hybrid instance-based system for predicting ocean temperature. International Journal of Computational Intelligence and Applications, v. 1, n. 1, p. 35-52, 2001.
- Kolodner, J. Reconstructive memory: a computer model. Cognitive Science 7(4), 1983 apud KOLODNER 1993.
- Linden, R.. Técnicas de Agrupamento. Revista de Sistemas de Informação de FSMA. n. 4, pp. 18-36, 2009.

Referências

- Abel, M.. Um estudo sobre Raciocínio Baseado em Casos. 1996. 41 f. Dissertação (Pós-Graduação em Ciência da Computação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 1996.
- Mitchell, T. Machine Learning. McGraw Hill, 1997.
- Riesbeck, R. S. Inside Case-Based Reasoning. Lawrence Erlbaum, 1989.
- Vasconcelos, G. Q.. Estudo comparativo de algoritmo de junção por similaridade em memória. 2016. 57 f. Dissertação (Graduação) – Curso de Bacharelado em Ciência da Computação, Universidade Estadual de Londrina, Londrina, 2016.

Referências

- Zia, S. S., Akhtar, P., Mughal, T. J., & Mala, I.. Case Retrieval Phase of Case-Based Reasoning Technique for Medical Diagnosis. *World Applied Sciences Journal*, 32(3), 451-458. 2014.



ANÁLISE DE MÉTRICAS PARA RECUPERAÇÃO DE CASOS VISANDO À APLICAÇÃO NO PROBLEMA DE CONDUÇÃO DE TRENS DE CARGA

Aluno: Matheus Streisky
Orientador: Dr. André Pinz Borges