

MINERAÇÃO DE DADOS EM FONTES ABERTAS: UMA FERRAMENTA DE APOIO AO PROCESSO DECISÓRIO

RAFAEL COSTA BARROS

Pós-graduado, lato sensu, em Gestão de Sistemas Táticos de Comando e Controle

RESUMO: A mineração de dados é fundamental no processo de tomada de decisão. Nos últimos 30 anos o processo de informatização cresceu exponencialmente. Estima-se que em 2020 o tráfego global de dados na *internet* será de 61,386 GBps. Esse crescimento teve como consequência o desenvolvimento da Ciência de Dados. Essa abrange uma grande gama de especialidades relacionadas com a exploração dos dados produzidos. Dentre essas áreas destacam-se: *Big Data*, *Machine Learning*, *Business Intelligence*, *Data Mining* (Mineração de Dados), *Knowledge Discovery in Databases* (KDD) e *Deep Learning*. Os padrões encontrados em um processo de mineração de dados têm sido utilizados nos mais variados ramos da sociedade: finanças, saúde, telecomunicações, direito, astronomia, biologia, engenharias e ciências da computação, como subsídio para tomada de decisão. Com a evolução do combate, temos um emprego crescente de satélites, aeronaves, *drones*, navios, sensores, radares, mídias sociais, *internet*, que geram um grande volume de dados, dessa forma a mineração de dados se estabelece como uma ferramenta muito útil, ao proporcionar uma análise oportuna dos dados gerados.

Palavras-chave: mineração de dados, processo decisório.

1 INTRODUÇÃO

A mineração apresenta-se como uma ferramenta essencial no processo decisório. O crescimento do volume de informação não é um fenômeno novo, é um processo contínuo ao longo dos anos. No entanto, com o avanço da tecnologia nas últimas décadas, o volume de informação gerado tem crescido cada vez mais rápido. A informação constitui a base do processo decisório. Dessa forma, o grau de êxito em uma decisão será proporcional à confiabilidade das informações obtidas, para a tomada de decisão.

A grande produção de informação tem sido explorada por diversas áreas do conhecimento, acompanhando essa tendência, as forças armadas de diversos países têm utilizado a mineração de dados em proveito do seu emprego, nos mais variados ambientes operacionais. O combate moderno tem acompanhado o avanço da tecnologia e, atualmente, os conflitos modernos caracterizam-se pelo conceito da Guerra Centrada em Redes (GCR), em que os agentes e os sistemas envolvidos no conflito utilizam meios informatizados.

No combate moderno é produzido um grande volume de dados através dos diversos sistemas utilizados, sendo fundamental a utilização de mecanismos que possam selecionar as informações úteis para determinada situação. Uma das soluções para esse fato é a utilização de mineração de dados, que

através da análise de padrões dos dados produzidos, consegue selecionar dados de interesse que podem gerar conhecimento através de uma posterior interpretação e avaliação.

2 RESULTADOS E DISCUSSÃO

A partir dos anos 1990, com o advento da internet, a produção de dados cresceu de forma acelerada. Na tabela abaixo, podemos verificar um estudo realizado pela Cisco Systems, em que podemos constatar o crescimento do volume de dados a partir dos anos de 1990, e esse estabelece, ainda, perspectivas para o ano de 2020.

TABELA 1 - Cisco VNI Previsão / Histórico - Internet

ANO	TRÁFEGO GLOBAL NA INTERNET
1992	100 GB por dia
1997	100 GB por hora
2002	100 GBps
2007	2.000 GBps
2015	20.235 GBps
2020	61.386 GBps

Fonte: Adaptado de Cisco VNI, 2016.

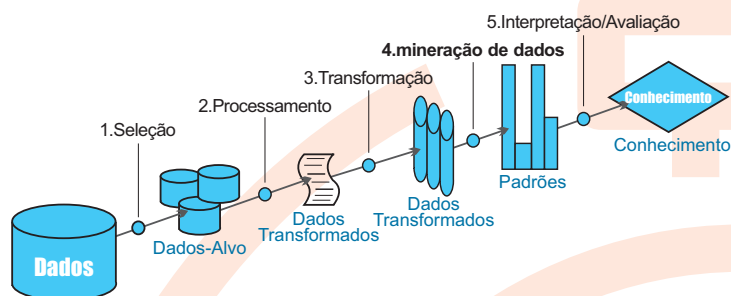
O grande volume de dados produzidos ao longo dos últimos 30 (trinta) anos estabeleceu uma vasta fonte de conhecimento existente na *internet*. Com o avanço da tecnologia, cada vez mais, os mais variados tipos de interações humanas estão sendo estabelecidos em redes informatizadas: atividades econômicas, culturais, sociais e políticas. Esse crescimento teve como consequência o desenvolvimento da Ciência de Dados.

A Ciência de Dados abrange diversas áreas relacionadas aos dados disponíveis na *internet*, dentre elas: *Big Data*, *Business Intelligence*, *Machine Learning*, *Data Mining* (Mineração de Dados), sendo fundamental a distinção entre essas áreas.

O *Knowledge Discovery in Databases* (KDD), descoberta de conhecimentos em base de dados refere-se ao processo completo de obtenção de conhecimento

desde a coleta de dados, a seleção, a transformação, a mineração de dados e a interpretação dos dados, com a finalidade de utilizar esse conhecimento obtido, para uma tomada de decisão. O KDD é o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e, finalmente, compreensíveis em dados (Fayyad, Piatetsky-Shapiro e Smyth, 1996, p.40).

FIGURA 1 - Descoberta do Conhecimento em Base de Dados



Fonte: Adaptado de (Fayyad, Piatetsky-Shapiro e Smyth, 1996).

A Fig.1 representa o processo de descoberta de conhecimento, em que um determinado volume de dados é submetido a diversas etapas. Após os dados serem coletados, selecionados, processados e transformados inicia-se a fase de mineração de dados, em que serão utilizados padrões, escritos em variadas linguagens de programação, para encontrar dados que atendam ao interesse proposto. Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 41) a etapa de mineração de dados pode ser de dois tipos: verificação e descoberta. O primeiro consiste em verificação simples de um determinado padrão pré-estabelecido. O segundo possibilita a descoberta de variados padrões.

Os padrões encontrados em um processo de mineração de dados têm sido utilizados nos mais variados ramos da sociedade, como subsídio para tomada de decisão.

De acordo com Han, J Kamber, M Pei, J, (2012, p. 613):

Enormes quantidades de dados de comunicação humana são produzidas diariamente. Tal comunicação existe em muitas formas, incluindo notícias, blogs, artigos, páginas da *web*, discussões *on-line*, revisões de produtos, *twitters*, mensagens, propagandas e comunicações, na *internet* e em vários tipos de redes sociais. Com isso, a mineração de dados nas Ciências Sociais tornou-se cada vez mais popular. Além disso, os comentários de usuários ou leitores sobre produtos, discursos e artigos podem ser analisados para deduzir opiniões e sentimentos gerais sobre os pontos de vista daqueles na sociedade. Os resultados da análise podem ser utilizados para prever tendências, melhorar o trabalho e ajudar na tomada de decisões.

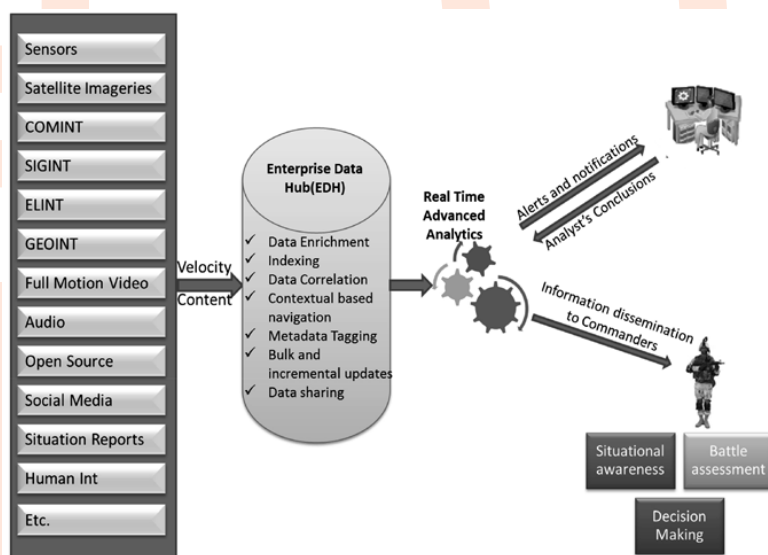
O combate no século XXI tem como característica o massivo emprego de sistemas informatizados, devido ao surgimento de novas ameaças. O inimigo tornou-se um elemento difuso, diferentemente dos conflitos do

século XX, em que eram bem definidos, exigindo uma grande utilização de ferramentas para monitoramento: satélites, aeronaves, *drones*, navios, sensores, radares, mídias sociais, *internet*. Os conflitos modernos são marcados por emprego de *streaming* de vídeo. Todas essas ferramentas geram um grande volume de dados, que necessitam ser utilizados oportunamente. Sendo essencial a utilização de dispositivos que consigam analisar esses dados.

De acordo com Haridas, M, (2015, p.72):

Hoje, os dados de máquina são gerados pelo movimento de navios, aeronaves e veículos, satélites no espaço, *drones*, veículos aéreos não tripulados, aeronaves de reconhecimento, sensores e radares de vigilância de campo de batalha. Os dados gerados por humanos incluem dados de sites de redes sociais, como *YouTube*, *Facebook*, *Tweeter*, etc. Os dados comerciais são gerados a partir de todas as transações de comércio eletrônico. Todos esses dados possuem inteligência, que não podemos desperdiçar. Análise de *Big Data* será utilizada para a coleta de informações em um futuro próximo, uma vez que os insumos para inteligência nacional e militar são obtidos continuamente durante a paz e a guerra, essa quantidade cresce exponencialmente durante as crises e as guerras. A análise humana deste volume de informação e dados de inteligência está muito além da capacidade física. Portanto, a grande inteligência baseada em análise de dados fornecerá o resultado necessário para a tomada de decisões e a condução das operações.

FIGURA 2 - Concepção do sistema de coleta de informações de aplicativos baseados em *Big Data*



Fonte: Adaptado de (HARIDAS, M. 2015).

3 CONCLUSÃO

Com o avanço da tecnologia, os processos envolvendo os diversos setores da sociedade tornaram-se informatizados, gerando um grande volume de dados. A mineração de dados em fontes abertas é uma ferramenta essencial em relação a produção de conhecimento em apoio ao processo decisório, pois vivemos em mundo que está amplamente interligado

através dos sistemas informatizados.

Esses produzem um grande volume de dados, que devem ser analisados de forma oportuna. A mineração de dados é uma ferramenta que tem sido utilizada nos mais variados setores da sociedade. Tem sido uma ferramenta muito útil nos conflitos modernos.

O processo de evolução do combate atual está diretamente ligado ao processo de evolução tecnológica. Forças Armadas de diversos países tem empregado equipamentos com alto valor tecnológico agregado e que geram uma expressiva quantidade de dados.

Com as perspectivas de crescimento do volume de dados em combate, é fundamental que as Forças Armadas possuam *expertises* para analisá-los em tempo oportuno, garantindo vantagem em uma situação de emprego.

e Pós-graduado pela Escola de Comunicações, *lato sensu*, em Gestão de Sistemas Táticos de Comando e Controle. Atualmente, exerce a função de instrutor da Escola de Comunicações e pode ser contactado pelo email rafaelcostabarross@gmail.com.

DATA MINING IN OPEN SOURCES: A TOOL TO SUPPORT THE DECISION-MAKING PROCESS

ABSTRACT

Data Mining is fundamental in the decision-making process. In the last 30 years the informatization process has grown exponentially. It is estimated that by 2020 global data traffic on the internet will be 61,386 GBps. This growth resulted in the development of Data Science. This covers a wide range of specialties related to the exploitation of the data produced. Among these areas are: Big Data, Machine Learning, Business Intelligence, Data Mining, Knowledge Discovery in Databases (KDD) and Deep Learning. The patterns found in a data mining process have been used in the most varied branches of society: finance, health, telecommunications, law, astronomy, biology, engineering and computer science, as a basis for decision making. With the evolution of combat, we have a growing employment of satellites, aircraft, drones, ships, sensors, radars, social media, internet, which generate a large amount of data, with this data mining is established as a very useful tool, by providing timely analysis of the data generated.

Keywords: data mining, decision-making.

REFERÊNCIAS

The Zettabyte Era — Trends and Analysis – Cisco Disponível em <<http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html>>. Acesso em: 27 maio 2017.

Fayyad, U. Piatetsky-Shapiro, G., Smyth P., **From Data Mining to Knowledge Discovery in Databases**, 1996, In: Advances in Knowledge Discovery and Data Mining, pp. 1–34.

Han, J Kamber, M Pei, J, **Data Mining Concepts and Techniques**, Waltham, Elsevier, 3ª Ed, 2012.

HARIDAS, M, **Redefining Military Intelligence Using Big Data Analytics**, Claws: Scholar Warrior, 2015.

O autor é graduado pela Academia Militar das Agulhas Negras

