

ANALISE E MINERAÇÃO DE DADOS SOCIAIS COM A FERRAMENTA WEKA

Mardonio Moura Araújo

Elifranio Alves Cruz

FAMETRO – Faculdade Metropolitana da Grande Fortaleza

mardoniomaraújo@gmail.com

elifranio.cruz@fametro.com.br

ANALISE E DESCOBERTA DE DADOS COM A FERRAMENTA WEKA

RESUMO

Avanços tecnológicos recentes, barateamento de custos de componentes e sistemas e até mudanças de paradigmas permitiram a coleta de grandes bases de dados para diversas aplicações. Em alguns casos o enorme volume de dados brutos coletados é processado para extrair informações mais interessantes, mas que ainda são em grande volume. A utilização de ferramentas que permitam a captação, o gerenciamento e análise das informações, além do estabelecimento de estratégias internas e externas, são hoje fatores indispensáveis a qualquer empresa que deseje apresentar significativa participação no mercado. Foi analisado duas ferramentas de mineração, extração e visualização de dados, o WEKA e Matlab. Para obter resultados específicos de uma base de dados, para o tal foi utilizado três bancos de dados distintos. Serão apresentados conceitos de mineração de dados, qual sua importância, e exemplos de como aplicar, com base em resultados alcançados nos teste efetuados durante a pesquisa de iniciação científica.

Palavras-chave: Dados, Informação, Mineração de dados, Análises, Conhecimento;

INTRODUÇÃO

Este artigo tem por motivação um estudo de ferramentas computacionais para a aquisição de conhecimento (informação útil) nos dados em sistemas de informação, em especial, extrair informações de usuários em redes sociais e permitir tomadas de decisão por administradores e gestores. As aplicações desenvolvidas para a extração de informações são classificadas como sistemas de descoberta de conhecimento, também chamado de KDD

(*knowledge Discovery in Databases*), e poderão atender a gestores, políticos, diretores, empresários.

A análise sistemática a ser estudada tem como premissa a capacidade forte de integração em redes sociais. Um esboço de cenário operacional de mineração de dados em redes sociais é ilustrado na Figura 01.

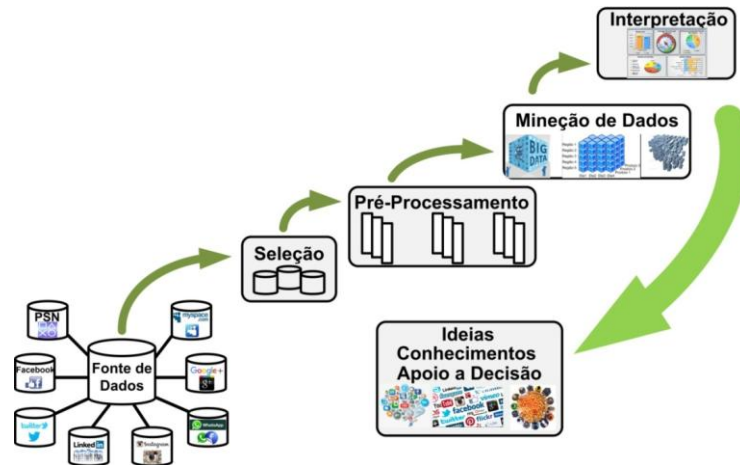


Figura 01. Protótipo Operacional para Mineração de dados em Redes Sociais [SILVA, 2016]

Na figura 01 é mostrada a aquisição de dados de possíveis diversas fontes geradora de dados em redes sociais. Os dados extraídos das fontes são selecionados de acordo com nível de interesse em análise, após a etapa operacional de seleção os dados são transformados e adaptados para formatos computacionais tratáveis e permitir compatibilidade com os algoritmos e técnicas mineração a serem aplicadas sobre os dados em redes sociais, feito esta etapa pode-se contribuir para gerar interpretações personalizadas e dar apoio a geração de conhecimento e tomadas de decisão [SILVA, 2016].

Com a popularização de redes sociais, a participação de usuários se comunicando, navegando ou mesmo acessando conteúdos e atividades publicados, tem possibilitado a geração de um histórico destas interações.

MINERAÇÃO DE DADOS

A mineração de dados ou data mining surgiu da junção de três áreas científicas que se relacionam são elas: Banco de dados, Estatística e Inteligência Artificial [KAWUU e CHUNG, 2015]. Ela oferece estratégias para análises de Big data, encontrando padrões e correlações entre os dados, que ajudam na tomada de decisões e desenvolver estratégias. Uma das ferramentas de apoio a esta atividade é ferramenta WEKA

Entender a natureza distribuída e conexa dos dados nestes ambientes, é algo fundamental para que se possa atingir metas de negócio, e garantir uma maior abrangência de

personalização e inovação de conhecimento sobre hábitos e padrões de interação humanas em redes sociais. A mineração de dados aplicada nas redes sociais busca obter resultados de comportamentos ou tendência social, usa como base de pesquisa as informações de cultura, costumes, localização. Observa-se na Figura 02, a análise de tendência aplicada sobre dados em redes sociais simulada no WEKA, com dados no formato .arff.

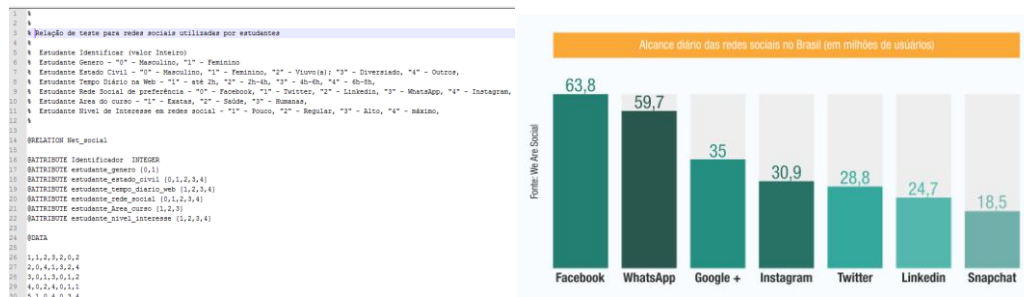


Figura 02. Base de dados .arff para análise no WEKA e gráfico de redes sociais

A Weka se consolidou como a ferramenta de mineração de dados mais utilizada em ambiente acadêmico. Ela é aplicada não apenas em pesquisas científicas, mas principalmente para fins didáticos. Foi exatamente a sua adequação para este último tipo de aplicação que a tornou popular. Professores responsáveis por ministrar cursos de mineração de dados ou business intelligence em universidades de todo o mundo, costumam empregar a Weka como instrumento de apoio para o ensino de conceitos básicos sobre data mining. Por meio de sua interface gráfica (conhecida como Weka Explorer) é possível conduzir processos de mineração de pequenas bases de dados, realizando a avaliação dos resultados obtidos e a comparação de algoritmos. Além disso, é possível executar tarefas relacionadas ao pré-processamento de dados como, por exemplo, a seleção e a transformação de atributos [WEKA, 2017]. O formato .ARFF (Attribute-Relation File Format), é o formato esperado pelos componentes do software Weka.

Na sequência são apresentados algumas estatísticas coletadas em redes sociais por meio de mineração de dados.

APLICAÇÃO NAS REDES SOCIAIS

Hoje, com a base informacional pública produzida em ambientes interacionais online, como as mídias sociais, temos a possibilidade de resgatar os rastros deixados pelos usuários, compreender um pouco mais alguns comportamentos, cenários e hábitos existentes para, com base nisso, realizar tomadas de decisão. É nesse contexto que a Mineração de Dados dialoga fortemente com as mídias sociais [Yakushev e Mityagin, 2014], permitindo que organizações possam obter insumos para a inteligência competitiva. Segundo a empresa

MAX2DIGITAL em seu blog, algumas informações relevantes sobre redes sociais já levantadas por empresas em 2017, por meio de mineração de dados e ferramentas de BI, são:

Dados Demográficos

- ✓ 75% dos usuários de internet do sexo masculino estão no Facebook, enquanto 83% correspondem às mulheres internautas que também estão na rede social de Mark Zuckerberg.
- ✓ 32% dos adolescentes consideram o Instagram a rede social mais importante.
- ✓ Mulheres na internet estão mais propensas a usar o Instagram do que os homens: 38% vs. 26%.
- ✓ 29% dos internautas com ensino superior completo usam o Twitter, comparando com 20% dos que têm ensino médio completo ou inferior a isso.
- ✓ 81% dos millennials verificam o Twitter pelo menos uma vez por dia.
- ✓ A maioria dos usuários do Instagram estão entre 18-29 anos.
- ✓ 7.22% da população total do planeta usa o Facebook.
- ✓ LinkedIn possui mais de 450 milhões de perfis cadastrados.
- ✓ Instagram Stories já conquistou 150 milhões de usuários ativos em pouco tempo de vida.
- ✓ YouTube alcança mais os públicos de 18-34 e 18-49 anos.

Dados Estatísticos

- ✓ Facebook continua sendo a rede social mais usada, com 1,86 bilhão de usuários ativos por mês. 76% dos usuários visitaram o site diariamente em 2016.
- ✓ A média de tempo gasto por usuário no LinkedIn é de 17 minutos por mês.
- ✓ 51% dos usuários de Instagram acessam a rede diariamente e 35% afirma usar a plataforma várias vezes por dia.
- ✓ Quase 80% do tempo gasto nas redes sociais acontece via mobile (viva os dispositivos móveis, principalmente o celular).
- ✓ Katy Perry tem o maior número de seguidores no Twitter, com seus 94.65 milhões.
- ✓ Em média, 400 milhões de snaps são compartilhados no Snapchat por dia e quase 9 mil fotos são compartilhadas todo segundo.
- ✓ Apesar de uma leve queda no número de usuários depois do lançamento do Instagram Stories, o Snapchat possui 150 milhões de usuários ativos diariamente, sendo 73% millennials, 71% jovens com menos de 25 anos e 70% mulheres.
- ✓ Apenas 10 mil vídeos do YouTube geraram mais de 1 bilhão de visualizações..
- ✓ Mais da metade de todas as visualizações do YouTube vêm de dispositivos móveis.

Estatísticas de negócio

- ✓ Instagram ganha U\$595 milhões de dólares em receita de publicidade móvel por ano, um número cada vez maior.
- ✓ Apesar da notícia de demissões e executivos deixando a empresa, a receita do Twitter é de 8% YOY (taxa homóloga em relação a dois anos).
- ✓ Mais de 50 milhões de empresas usam o Facebook Business.
- ✓ 2 milhões de empresas usam o Facebook para fazer propaganda.
- ✓ A receita total do Facebook cresceu 56% em 2016, enquanto as receitas publicitárias cresceram 59%.
- ✓ 93% dos usuários do Pinterest usam a plataforma para planejar ou fazer compras.
- ✓ 39% dos usuários do LinkedIn pagam, mensalmente, pela conta premium.
- ✓ Pinterest impulsiona 25% de todo tráfego de referência de sites de varejo.
- ✓ Mais de 56% dos adultos online usam mais de uma rede social.

Estatísticas de conteúdo

- ✓ Tuites com imagens recebem 18% mais cliques do que os envios sem conteúdo visual.
- ✓ Há 100 milhões de catálogos de comida e 146 milhões de moda no Pinterest.
- ✓ No LinkedIn, 98% dos posts com imagens recebem mais comentários e os posts com links conquistam 200% mais engajamento.
- ✓ Há por volta de 81 milhões de contas fake no Facebook.
- ✓ No Twitter, 5% dos usuários são falsos.
- ✓ 100 milhões de horas são gastas por usuários vendo vídeos no Facebook diariamente.
- ✓ Mais de um milhão de usuários do LinkedIn publicaram conteúdos extensos, com 160 mil posts publicados semanalmente.
- ✓ Foram upadas mais de 19.7 milhões de apresentações no SlideShare no LinkedIn.
- ✓ 88% das empresas com mais de 100 funcionários usam Twitter com foco em marketing.
- ✓ O vídeo com o maior número de visualizações é “Charlie bit my finger”, com mais de 845 milhões.
- ✓ Pizza é a comida mais publicada no Instagram, seguida pelo bife e pelo sushi.
- ✓ Os blogs continuam em crescimento, com mais de 409 milhões de pessoas visualizando mais de 23.6 bilhões de páginas por mês apenas no WordPress.

A partir do momento em que empresa/marcas podem ‘ouvir’ na web o que as pessoas falam, desejam, pedem, reclamam, elas, também, serão mais assertivas no momento em que precisarem agir (de forma reativa ou pró-ativa). Se podemos compreender hábitos e comportamentos em diferentes ambientes online, podemos melhorar não apenas a mensagem

que chega até o público de interesse, mas, também, o modo como interagimos e nos relacionamos com esse público. Práticas como o Monitoramento, Pesquisas (Surveys, Enquetes etc.), *Web Analytics*, Mensuração, Netnografia, entre outras, tomam para si técnicas de *Data Mining* para coletar os dados que serão lapidados através da interpretação [SILVA, 2016].

ANALISE DOS RESULTADOS

O WEKA é uma das ferramenta de mineração de dados com grande potencial, possui uma abordagem moderna e interativa que capacita aos usuários a acessar, descobrir, visualizar, fazer análises de uma grande base de dados.

Com uma vasta opções de ferramentas para análises que estão em constantes atualizações a partir de relatórios para modelagem preditiva os usuários podem trabalhar com o WEKA fazendo extrações e apresentações de dados sem necessitar do setor de TI da empresa. A ferramenta Matlab tem como ponto forte a forma de apresentação dos resultados e cálculos matemáticos, ele contém variadas formas de visualização através de inúmeras opções de gráficos e comandos, com um ambiente amigável ao usuário ele permite a junção e manipulação de várias bases de dados através de seus filtros, ideal para um analista de negócios. A articulação de novas variáveis em um determinado ambiente fornece um passo importante para o posterior descobrimento de novos conhecimentos sobre o processo de disseminação e evolução social no Brasil.

CONSIDERAÇÕES FINAIS

A mineração de dados vem auxiliar as corporações em suas análises de negócios, que devido as mudanças constantes das tendências do mercado, surge a necessidade de montar estratégias para se manter firme.

Os softwares de atuação de mineração de dados (*Data Mining*), são utilizados em qualquer área de atuação no qual se busque por resultados para tomadas de decisões, inclusive nas redes sócias, que é onde as pessoas se mantém conectadas frequentemente e que é uma grande rede de relacionamento.

A facilidade de usar as ferramentas de extração e manipulação de dados fazem com que o uso não se limite apenas a profissionais de TI, aumentando ainda mais a busca por utilização das ferramentas.

REFERÊNCIAS

- [CABENA, 1997] CABENA, P. et al.; “*Discovering Data Mining from concept to implementation*”. New Jersey: Prentice-Hall, 1997, 195p.
- [D’Avanzo e Pilato, 2015] E. D’Avanzo, G. Pilato. *Mining social network users opinions’ to aid buyers’ shopping decisions*. Computers in Human Behavior. Elsevier Journal 51. 2015.
- [ELMARI e NAVATHE, 2010] ELMARI, R.; NAVATHE, S. B.; Sistemas de Banco de Dados. 6ª Edição. Editora Pearson. (pp 698 – 731). 2010.
- [KAWUU e CHUNG, 2015] kawuu W. Lin, Sheng-Hao Chung. *A fast and resource efficient mining algorithm for discovering frequent patterns in distributed computing environments*. Future Generation Computer Systems. Elsevier Journal 52. 2015.
- [MAX2DIGITAL, 2017] Portal max2digital. Disponível em <http://blog.max2digital.com.br/redes-sociais/40-estatisticas-de-redes-sociais/>. Acessado em Setembro de 2017.
- [MICROSOFT, 2016] MICROSOFT, site oficial. <https://msdn.microsoft.com/pt-br/library/ms174949.aspx>
- [Nettleton, 2013] David F. Nettleton. *Data mining of social networks represented as graphs*. Computer Science Review. Elsevier 7. 2013.
- [Podobnik e Lovrek, 2015] V. Podobnik, I. Lovrek. *Implicit Social Networking: Discovery of Hidden Relationships, Roles and Communities among Consumers*. 19th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. 2015.
- [RAVI e RAVI, 2015] Kumar Ravi, Vadlamani Ravi. *A survey on opinion mining and sentiment analysis: Tasks, approaches and applications*. Knowledge-Based Systems. Elsevier Journal 89. 2015.
- [RUSSEL, 2011] Matthew A. Russel. *Mining the Social Web*. O’Reilly, 1ª edição, 2011.
- [SILVA, 2016] SILVA, A. C.; CRUZ, E. A.; Análise e mineração de dados sociais com ferramenta tableau e pentaho. XII Semana Acadêmica: Conexão Fаметro 2016. ISSN: 2357-8645.
- [TABLEAU, 2016] TABLEAU, site oficial 1. Acesso em 08/16. <https://www.tableau.com/pt-br/stories/workbook/get-twitter-analytics-you-need-about-your-own-brand-or-build-interactive-iral-content>
- [WEKA, 2017] UNIVERSITY OF WAIKATO. Weka 3.8 – Machine Learning Software in Java. Disponível no site da University of Waikato (2017). URL: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- [YAKUSHEV e MITYAGIN, 2014] A. Yakushev and S. Mityagin. Social networks mining for analysis and modeling drugs usage. 14th International Conference on Computational Science. Procedia Computer Science. ICCS 2014.