

Detecção de Fraudes: Uma Revisão Sistemática

Jean Avila Rangel · Maria Claudia
Figueiredo Pereira Emer · Adolfo
Gustavo Serra Seca Neto

Received: date / Accepted: date

Resumo Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords Fraud detection · Anomaly detection · Deception detection · Standard deviation detection · Detecção de fraude · Detecção de anomalia · Detecção de engano · Detecção de desvio padrão

1 Introdução

A detecção de fraudes é utilizada para resolver problemas variados, sendo geralmente utilizada para reduzir falhas de segurança em sistemas onde há gasto de recursos com usuários mal intencionados.

Segundo Fawcett and Provost [1997], há muitas técnicas para detecção de fraudes. As mais difundidas e utilizadas são as que produzem detecções por meio de regras pré estabelecidas e as que realizam a comparação entre valores de dados. Estas classificações geraram duas principais ramificações em pesquisas.

Uma das ramificações, a detecção de fraude por meio de regras, determina que uma fraude será detectada devido ao conhecimento que a equipe adquiriu observando fraudes recorrentes em outrora. A vantagem nesta técnica está

Jean Avila Rangel
Federal University of Technology - Paraná, 3165 Sete de Setembro Avenue, Curitiba, PR
80230-901, BRA

Maria Claudia Figueiredo Pereira Emer · Adolfo Gustavo Serra Seca Neto
Academic Department of Informatics, Federal University of Technology - Paraná, 3165 Sete
de Setembro Avenue, Curitiba, PR 80230-901, BRA
E-mail: mciemmer@gmail.com

Adolfo Gustavo Serra Seca Neto
E-mail: adolfo@dainf.ct.utfpr.edu.br

em predir como a fraude ocorre. A desvantagem está na possibilidade de descobrir somente fraudes já conhecidas.

O outro ramo na detecção de fraudes ocorre por meio de cálculos e comparação de valores. Como principal exemplo, um desvio no padrão comportamental de alguma variável pode ser identificado comparando seus dados com dados anteriores ou de outras variáveis similares. Sua principal vantagem é a utilização de algoritmos já conhecidos na literatura [Fawcett and Provost, 1997]. O problema apresentado pela técnica é a possibilidade do indivíduo fraudador conhecer e inserir dados fraudulentos de maneira em que o sistema não consiga o identificar como um usuário malicioso.

Fawcett and Provost [1997] realizaram uma das primeiras e principais abordagens no ramo de detecção de fraudes encontradas na literatura. Em seus estudos, analisaram e criaram um sistema para detecção de fraudes em sistemas de telecomunicação. Em seus estudos, constaram que indivíduos fraudadores estão constantemente mudando suas táticas para burlar sistemas, que devem adaptáveis às novas situações.

Chan et al [1999] realizaram um trabalho em detecção de fraude em cartões de crédito utilizando dados reais, fornecidos por companhias bancárias, e obtiveram um aumento nos seus resultados em comparação com técnicas de detecções de fraudes utilizadas anteriormente. Em suas constatações, indicaram que os métodos mais potentes para detectar fraudes são utilizados em cartões de crédito, porém as organizações não costumam revelar suas tecnologias de forma aberta, pois possíveis fraudadores podem se aproveitar do conhecimento e utilizar nos sistemas com má fé.

A realização da revisão sistemática proposta por este trabalho obtém motivação devido à crescente expansão do Estado da Arte no assunto. Em pesquisas realizadas em bases de dados por trabalhos realizados na área de detecção de fraude, notou-se a produção de muito material anteriormente ao ano de 2010, porém com um significativo aumento na porcentagem de publicações posteriores a 2010. Este fator indica o crescimento do interesse da comunidade científica em pesquisar o assunto.

O controle organizacional de uma empresa ou setor pode garantir a sua boa estabilidade. Soluções para detectar, prevenir e evitar fraudes podem se tornar uma ferramenta para auxiliar a economia de recursos [Chan et al, 1999].

Em todo contexto organizacional, cada ambiente possui um tipo de tratamento e armazenamento de dados. Portanto, ambientes podem possuir dados organizados de maneira estável por induzirem normativas e tecnologias padronizadas ou fazer exatamente o contrário, possuindo dados desorganizados com poucas conexões lógicas. Desta maneira, este estudo visa conhecer abordagens gerais e adaptáveis em detecção de fraudes.

A detecção de fraudes vem sendo utilizada há muito tempo para controle organizacional e econômico [Seyedhossein and Hashemi, 2010]. Em estudos realizados no estado da arte, houve uma grande apresentação sobre algoritmos para detectar fraudes em sistemas financeiros ou de cartão de crédito [Chan et al, 1999], [Chandola et al, 2009] e [Abdallah et al, 2016].

O problema computacional que ainda ocorre é a não detecção de todos os falsos negativos (onde ações fraudulentas não são detectadas), bem como a classificação equivocada de falsos positivos (ações legítimas caracterizadas, erroneamente, como fraudulentas). Identificar uma forma para reduzir a incidência de erros é uma importante contribuição para a área e a principal justificativa para este estudo.

O objeto de estudo do trabalho consiste em encontrar técnicas de detecção de fraude, preferencialmente aplicáveis a contextos de cartões de crédito, e que apresentem bons resultados e desempenho. A eficácia e utilidade dos algoritmos computacionais serão analisadas e consideradas para realizar estudos futuros onde serão aplicadas concomitantemente.

1.1 Objetivos

1.1.1 *Objetivo Geral*

Este trabalho possui como objetivo geral realizar uma revisão sistemática em trabalhos de revisões em técnicas voltadas para detecção de fraudes. Desta forma, sistemas desenvolvidos para inspecionar desvio de padrões, comportamentos indevidos ou fraudes serão identificados e estudados através dos trabalhos publicados no Estado da Arte.

1.1.2 *Objetivos Específicos*

Para encaixar o objeto que está sendo estudado com o problema apresentado, pretende-se - após realizada a pesquisa entre trabalhos na área - encontrar uma ou um conjunto de técnicas apropriadas para resolver o problema computacional apresentado.

Paralelamente com o objetivo geral, espera-se encontrar as carências que os trabalhos estudados demonstram. As falhas cometidas por autores passados sobre o assunto podem gerar novas pesquisas e um melhor aprofundamento sobre o tema.

Também se espera realizar uma revisão sistemática considerando os trabalhos que foram publicados após as ultimas revisões sistemáticas serem escritas.

Todas as técnicas para detecção de fraudes que forem encontradas e forem consideradas relevantes, serão apresentadas e descritas.

2 Referencial Teórico

3 Metodologia

Ao primeiro momento do trabalho de revisão sistemática, o planejamento da estratégia de pesquisa norteia a pesquisa em publicações na área abordada.

A revisão da literatura foi baseada no trabalho apresentado por Kitchenham and Charters [2007], onde os elementos de pesquisa são definidos para realizar e relatar a revisão.

Foram selecionadas as bibliotecas digitais abaixo para realizar a pesquisa. O motivo da escolha foi a grande quantidade de material publicado com relevância científica.

- ACM Digital Library¹
- IEEE Xplore Digital Library²
- ScienceDirect³
- SpringerLink⁴

A tarefa para selecionar os artigos pode ser visualizada no diagrama da Figura 1, onde as etapas foram divididas em início, desenvolvimento e conclusão. Na parte inicial, o tema é definido e publicações naquele tema são pesquisadas.

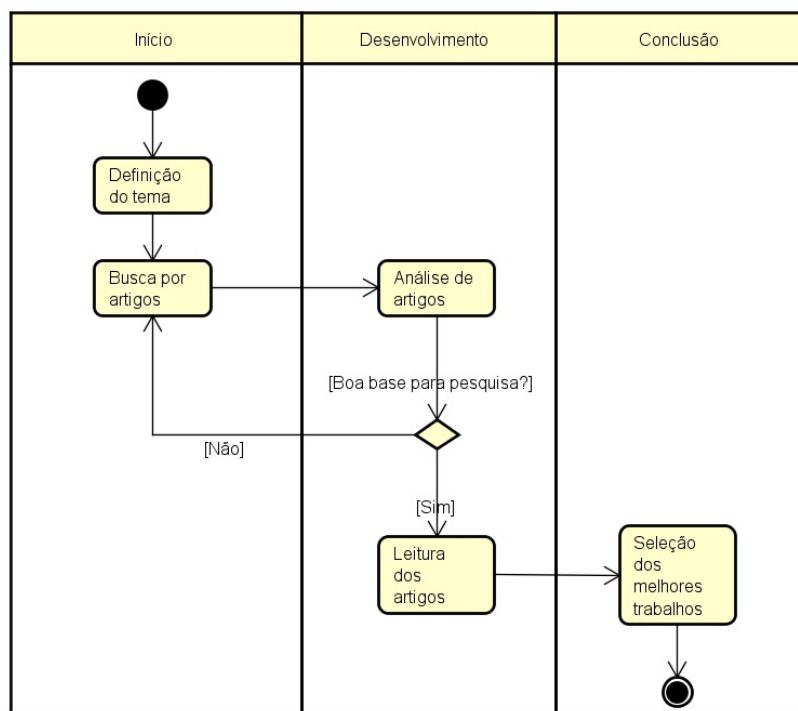


Figura 1 Diagrama de sequência indicando as etapas da pesquisa.

¹ <http://dl.acm.org/>

² <http://ieeexplore.ieee.org/>

³ <http://www.sciencedirect.com/>

⁴ <http://link.springer.com/>

Como critérios de seleção e ordenação para a pesquisa, a revisão escolheu os artigos com maior número de citações e relevância para a área. Realizando uma exceção temporal, os principais artigos que tratam do tema de detecção de fraudes foram incluídos no estudo, pois eram encontrados como referência na maioria dos trabalhos atuais.

Outro critério de seleção foi a proximidade com o tema destacado na introdução deste trabalho, onde a detecção de fraudes deveria ser a principal abordagem do artigo encontrado.

Foram excluídos da revisão os artigos que fugiam do tema proposto, não possuíam citações ou relevância nas bases de dados ou eram anteriores ao ano de 2006. Os critérios de exclusão estão elencados abaixo. Foram excluídos da seleção os artigos contendo assuntos relacionados a detecção de fraudes em:

- Imagens, vídeos ou impressões físicas;
- Contextos biológicos, como análises de DNA e moléculas;
- Componentes eletrônicos;
- Dispositivos de segurança eletrônica relacionados a vírus, *firewalls* e *cyber* ataques.

Para a busca de publicações, foram considerados somente trabalhos que realizaram uma revisão no Estado da Arte em detecção de fraude, na língua inglesa e publicados no período entre 2006 e 2016. As palavras chave *fraud detection survey* foram utilizadas para a busca, que obteve artigos selecionados com a leitura na respectiva ordem: título, resumo e artigo completo. Foram considerados somente trabalhos em que nenhum novo método era apresentado ou testado. Portanto, somente revisões na literatura da área foram incluídas.

Na primeira fase da pesquisa, foi considerada a utilização das palavras chave a seguir, que foram eliminadas da metodologia:

- *fraud detection AND state of art OR systematic review OR meta analysis;*
- *anomaly detection AND survey OR state of art OR systematic review OR meta analysis;*
- *outlier detection AND survey OR state of art OR systematic review OR meta analysis;*
- *deception detection AND survey OR state of art OR systematic review OR meta analysis.*

Contudo, os resultados contendo os temas *anomaly*, *outlier* e *deception detection* retornaram trabalhos duplicados a pesquisa com as palavras chave *fraud detection survey* ou fora do escopo da pesquisa, onde a grande parte foi eliminado da seleção através dos métodos de exclusão citados anteriormente. Por esta razão, as palavras chave não foram consideradas para a pesquisa de artigos.

No início da revisão sistemática, foram encontrados 101 artigos com potencial para estruturar este trabalho. Destes, dois não estavam disponíveis para consulta. Após a leitura do título dos 99 artigos, 49 foram excluídos. Posteriormente, seguindo os critérios de exclusão, foi realizada a leitura do resumo dos artigos. Destes, 23 trabalhos foram selecionados para realizar a revisão sistemática proposta por este trabalho.

4 Estado da Arte

5 Discussão

6 Conclusão

7 Cronograma

Conforme visto na Figura 2, o escopo do trabalho foi delimitado pelas tarefas: Definir tema; Definir objetivos, justificativa e motivação; Escolher o local para publicação; Elaborar o método de pesquisa; Revisar a literatura e selecionar trabalhos; Produzir artigo científico; e Publicar o artigo finalizado.

As tarefas foram divididas entre os meses de junho a setembro e receberam a classificação: Concluída ("OK", em verde); Em andamento ("Andam.", em amarelo); e Não iniciada ("Não Inic.", em vermelho).

Tarefa \ Mês	Jun	Jul	Ago	Set
Definir tema.	OK			
Definir objetivos, justificativa e motivação.	OK			
Escolher o local para publicação.	Ok			
Elaborar o método de pesquisa.	OK			
Revisar a literatura e selecionar trabalhos.		Andam.	Andam.	
Produzir artigo científico.		Andam.	Andam.	Não Inic.
Publicar artigo finalizado.				Não Inic.

Figura 2 Cronograma do planejamento do projeto.

8 Projeto de Pesquisa do Mestrado

Na finalização da revisão sistemática, espera-se a publicação dos resultados obtidos na revista Data Mining and Knowledge Discovery, a qual possui a classificação de qualidade Qualis B1, elencada pela CAPES em sua última análise.

Referências

- Abdallah A, Maarof MA, Zainal A (2016) Fraud detection system: A survey. Journal of Network and Computer Applications 68:90 – 113, DOI <http://dx.doi.org/10.1016/j.jnca.2016.04.007>, URL <http://www.sciencedirect.com/science/article/pii/S1084804516300571>
- Chan PK, Fan W, Prodromidis AL, Stolfo SJ (1999) Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and their Applications 14(6):67–74, DOI 10.1109/5254.809570

- Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. *ACM Comput Surv* 41(3):15:1–15:58, DOI 10.1145/1541880.1541882, URL <http://doi.acm.org/10.1145/1541880.1541882>
- Fawcett T, Provost F (1997) Adaptive fraud detection. *Data Mining and Knowledge Discovery* 1(3):291–316, DOI 10.1023/A:1009700419189, URL <http://dx.doi.org/10.1023/A:1009700419189>
- Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering
- Seyedhossein L, Hashemi MR (2010) Mining information from credit card time series for timelier fraud detection. In: *Telecommunications (IST), 2010 5th International Symposium on*, pp 619–624, DOI 10.1109/ISTEL.2010.5734099