

Descoberta de Conhecimento em Banco de Dados Relacionados à Violência Contra a Mulher

Thales V. Maciel¹, Luiza M. F. Vinholes¹

¹Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSul)
Campus Bagé – Av. Leonel de Moura Brizola, 2501 – 96408400 – Bagé – RS – Brasil

thalesmaciel@ifsul.edu.br, vinholesluiza@gmail.com

Abstract. *The present paper aims at a method for predicting local rates for crimes against women based on the hypothesis that distinct crime rates have impact on each other. It is proposed that data from SSP-RS and IBGE data are integrated and go through further preprocessing, such as cleaning and discretization, before being analyzed by the Apriori algorithm in order to discover association rules. Preliminary results lead to further experimentation to evaluate potential for helping prevent violence against women.*

1. Introdução

O serviço Ligue 180 (Central de Atendimento à Mulher) foi criado pela Secretaria de Políticas para as Mulheres da Presidência da República (SPM-PR), em 2005, para servir como canal direto de orientação gratuita sobre direitos e serviços públicos para a população feminina em todo país (OMV, 2016).

Segundo o relatório próprio de atividade de 2015 (OMV, 2016), o serviço realizou 749024 atendimentos no ano, o que representou aumento de 54% nas denúncias em relação ao ano anterior. O relatório evidenciou que as denúncias mais comuns são as referentes a ameaças sofridas, seguidas de lesões corporais dolosas, estupro e crimes violentos letais intencionais.

Dados de crimes de violência contra a mulher têm sido material de pesquisa em estudos de computação aplicada, como, por exemplo, em Bansal e Bhambhu (2013), onde dois algoritmos tiveram seus desempenhos comparados na descoberta de regras de associação sobre um conjunto de dados não detalhado e sem objetivos específicos para a aplicação da mineração de dados; e Silva (2014), onde foi avaliada a utilidade da aplicação de uma ferramenta específica de business intelligence (BI) sobre um conjunto de dados não especificado.

O problema de pesquisa abordado no presente trabalho é “como se relacionam os indicadores de ocorrências de crimes contra a mulher?”. A hipótese adotada é a de que tais indicadores impactam na proporção direta ou inversa entre si. O objetivo do trabalho é a especificação de um método para previsão de panoramas criminais locais e prover embasamento para planejamento estratégico no emprego de medidas de prevenção dos tipos de violência contra a mulher.

2. Solução Proposta

Para tal, procedeu-se à descoberta de conhecimento em bancos de dados (DCBD) conforme processo descrito em Maciel et al. (2015): definido pelas etapas de pré-processamento, onde o conjunto de dados bruto é limpo, integrado, redimensionado e transformado; processamento, onde é aplicado, sobre o conjunto de dados, um algoritmo de mineração de dados para descoberta de padrões, em ambiente computacional, e; pós-processamento, onde os padrões descobertos são analisados no tocante a novidade, relevância e aplicabilidade em âmbito científico ou de domínio de negócio.

No estudo, foram integrados dois conjuntos de dados. O primeiro deles foi referente à quantitativos das ocorrências de crimes contra a mulher em 496 municípios do Estado do Rio Grande do Sul (RS) em cada mês do ano de 2015, obtido junto à Secretaria de Segurança Pública do Estado do Rio Grande do Sul (SSP-RS). O segundo conjunto de dados contou com o quantitativo de população feminina em 496 municípios

do Estado do Rio Grande do Sul no ano de 2015 e foi obtido junto ao Instituto Brasileiro de Geografia e Estatística (IBGE).

Esta integração ocorreu com vistas da oportunidade de cálculo da razão entre o quantitativo de crimes contra a mulher e o quantitativo de população feminina (QPF) em cada cidade. Desta forma, pretende-se evitar a possível irrelevância estatística de municípios de menor população quando comparados àqueles de população maior.

Após estes cálculos, foram removidos atributos do conjunto de dados a fim de manter apenas aqueles referentes às razões entre quantitativo de ameaças e QPF, lesões corporais e QPF, estupros e QPF, feminicídio e QPF e tentativa de feminicídio e QPF, totalizando 5952 instâncias com os 5 atributos descritos.

Ainda em pré-processamento, optou-se pela discretização dos atributos numéricos referentes às razões entre os quantitativos criminais e os QPF conforme escala categorizada em baixo, intermediário e alto. Isto se fez necessário pela natureza da tarefa de processamento proposta no estudo, a descoberta de regras de associação, que requer atributos do tipo categórico, estritamente (Bansal e Bhambhu, 2013).

Contudo, o método de discretização pela segmentação igualitária do intervalo numérico se mostrou inadequado ao estudo, devido a que produziu alto desbalanceamento às classes identificadas. Desta forma, optou-se pela discretização através de tarefas de clustering sobre cada atributo individualmente (Gupta et al., 2010), que foi capaz de descobrir, por aprendizado não supervisionado, os intervalos para as categorias baixo, intermediário e alto para cada atributo.

O algoritmo Apriori tem o propósito da descoberta de regras de associação, que identificam combinações frequentes nos valores de atributos dentre as instâncias de um conjunto de dados (Bansal e Bhambhu, 2013). Quando aplicado sobre o conjunto de dados pré-processado, o Apriori revelou associação entre altos e baixos índices de criminalidade contra a mulher.

Os experimentos preliminares mostraram que: em 61% dos períodos em que ocorrências de lesões foram baixas, ameaças também foram baixas; em 52% dos períodos em que ocorrências de estupros foram altas, ameaças foram baixas e; em 50% dos períodos em que ocorrências de lesões corporais foram altas, ameaças foram baixas.

3. Considerações Finais

Experimentos preliminares de descoberta de associações entre os índices de ocorrências de crimes contra a mulher foram capazes de encontrar padrões de combinações frequentes nos dados processados. Isto propõe que a hipótese estudada é verdadeira e mostra o potencial de evolução do estudo para uma ferramenta de previsão de panoramas criminais locais e prover embasamento para planejamento estratégico no emprego de medidas de prevenção dos tipos de violência contra a mulher. Experimentos futuros envolvem a utilização de regressão na estimativa de valores numéricos.

Referências

- Bansal, D. and Bhambhu, L. (2013). Usage of Apriori algorithm of data mining as an application to grievous crimes against women. *IJCTT* 2013.
- Gupta, A., Mehrotra K. e Mohan C. (2010). "A clustering-based discretization for supervised learning." *Statistics & probability letters* 80.9-10.
- Maciel, T., Seus, V., Machado, K. and Borges, E. (2015). Mineração de dados em triagem de risco de saúde. *Revista Brasileira de Computação Aplicada*, 7(2), 26-40.
- Observatório da Mulher contra a Violência (OMV). (2016) Panorama da Violência contra as Mulheres no Brasil: indicadores nacionais e estaduais. Brasília. Senado Federal.
- Silva, A., Silva, A., Viana, T. e Melo, M. (2014). Uso Da Ferramenta Pentaho BiServer: Analisando A Violencia Contra Mulher no Estado Do Rio Grande Do Sul. In: IX CONNEPI.