

Proposta de aplicação da fusão de dados e informações no apoio à prevenção de acidentes de trânsito nas rodovias federais brasileiras

Proposal of application of the data and information fusion to support the prevention of traffic accidents on Brazilian federal highways

Jordan Ferreira Saran¹, Ronnie Shida Marinho¹, Clayton Martins Pereira¹, Leonardo Castro Botega¹, José Eduardo Santarem Segundo²

(1) Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília – SP, {jordan.saran, ronnie.shida, clayton.martins, leonardo.botega}@unesp.br

(2) Universidade de São Paulo (USP), Faculdade de Filosofia, Ciências e Letras, Ribeirão Preto – SP, santarem@usp.br

Resumo:

A responsabilidade pela coleta e disponibilização de dados oficiais sobre acidentes, infrações e condições nas rodovias federais brasileiras está descentralizada em diferentes órgãos públicos do Poder Executivo Federal. Como efeito desta descentralização, não há uma unificação das bases de dados para o planejamento de ações de fiscalização e de prevenção de acidentes nessas vias. Por outro lado, os modelos de fusão de dados e informações surgem como forma de orientar os processos de desenvolvimento de sistemas para a aquisição, inferência, avaliação e representação de informações situacionais de alto nível. Nesse contexto surge o presente trabalho de pesquisa, que tem como objetivo auxiliar o planejamento de ações de prevenção de acidentes de trânsito em rodovias federais, ao propor a aplicação dos métodos da fusão de dados e informações, a partir de dados coletados de forma semiautomática de órgãos governamentais, e de informações inseridas manualmente pelos agentes envolvidos. Assim, ao utilizar a fusão de dados na aplicação proposta neste artigo, espera-se obter representações mais completas, precisas e enriquecidas com dados de múltiplas fontes de acidentes de trânsito. Almeja-se com isso minimizar problemas de acidente de trânsito nas rodovias federais brasileiras.

Palavras-chave: Prevenção de Acidentes de Trânsito; Fusão de Dados e Informações; Rodovias Federais Brasileiras.

Abstract:

The responsibility for the collection and availability of official data on accidents, infractions and conditions on the Brazilian federal highways is decentralized in different public agencies of the Federal executive branch. As a result of this decentralization, there is no unification of the databases for the planning of surveillance and accident prevention actions in these ways. On the other hand, data and information fusion models emerge as a way to guide the processes of systems development for the acquisition, inference, evaluation and representation of high-level situational information. In this context arises the present research work, which aims to assist the planning of actions to prevent traffic accidents on federal highways, through the application of methods of merging data and information, from data collected from Government agencies, and information entered manually by the agents involved. Thus, by using data fusion in the application proposed in this article, we expect to obtain more complete, accurate and enriched representations with data from multiple sources of traffic accidents. It aims to minimize traffic accident problems on Brazilian federal highways.

Keywords: Prevention of Traffic Accidents; Data and Information Fusion; Brazilian Federal Highways.

1 Introdução

De acordo com a Organização Pan-Americana de Saúde (OPAS, 2009), cerca de 1,35 milhão de pessoas no mundo morrem, a cada ano, em decorrência de acidentes de trânsito. Preocupados com estes números alarmantes, a Organização das Nações Unidas (ONU) em sua “Agenda 2030 para o Desenvolvimento Sustentável” fixou uma meta ambiciosa quanto à segurança no trânsito, que consiste em reduzir pela metade, até o ano de 2020, o número de mortos e

feridos por acidentes de trânsito em todo o planeta.

Dentre os principais fatores relacionados à ocorrência dos acidentes de trânsito estão: os erros humanos; o excesso de velocidade; a condução de veículos sob influência de álcool e outras substâncias; a distração ao dirigir; a falta de segurança da infraestrutura viária e dos veículos; a falta de obediência às normas e leis de trânsito; entre outros (OPAS, 2009).

Consoante aos fatores e, considerando especificamente as rodovias federais brasileiras, a responsabilidade pela coleta e disponibilização de dados oficiais sobre acidentes, infrações e condições nestas vias está descentralizada em diferentes órgãos públicos do Poder Executivo Federal.

Como efeito desta descentralização, não há uma unificação das bases de dados para o planejamento de ações de fiscalização e de prevenção de acidentes nas rodovias federais brasileiras. Neste sentido, diante deste complexo cenário que envolve as ocorrências de acidentes de trânsito, a formulação de políticas públicas e de campanhas educativas, e a adoção de melhorias na segurança das vias, necessitam do emprego de elementos de suporte ao conhecimento (DNIT, 2014).

Por outro lado, os modelos de fusão de dados e informações surgem como forma de orientar os processos de desenvolvimento de sistemas para a aquisição, inferência, avaliação e representação de informações situacionais de alto nível. Tais sistemas são alimentados por várias fontes de dados heterogêneas, na maioria dos casos de forma dinâmica, tal como seria no caso das rodovias federais brasileiras.

Nesse contexto surge o presente trabalho de pesquisa, que tem como objetivo auxiliar o planejamento de ações de prevenção de acidentes de trânsito em rodovias federais, por meio da proposta de aplicação dos métodos da fusão de dados e informações, a partir de dados coletados de forma semiautomática de órgãos governamentais, e de informações inseridas manualmente pelos agentes envolvidos.

A metodologia empregada neste trabalho, de natureza qualiquantitativa e de tipo descritivo, consiste primeiramente na realização de uma pesquisa bibliográfica, de forma a permitir uma breve revisão sobre os modelos de fusão de dados e informações, bem como estudar os trabalhos que seguem a mesma linha de pesquisa deste projeto. Por fim, foi realizada uma análise documental, onde foram obtidos fontes, dados e estatísticas sobre acidentes de trânsito no Brasil, além dos elementos necessários para a proposta de aplicação da fusão de dados e informações na prevenção de acidentes de trânsito nas rodovias federais brasileiras.

2 Acidentes de trânsito nas rodovias federais brasileiras

No Brasil, de acordo com dados do Ministério da Infraestrutura, a malha rodoviária federal possui atualmente 75.800 Km de extensão total, sendo que 65.400 Km são de rodovias pavimentadas e 10.400 Km de rodovias ou trechos não pavimentados. A maior parte desta malha está concentrada nas regiões sul e sudeste do país, sendo o estado de Minas Gerais o que possui a maior extensão de rodovias federais em sua área territorial.

De acordo com dados do IPEA (2015), nesta extensa malha rodoviária federal ocorrem aproximadamente 20% das mortes em acidentes de trânsito registradas anualmente no país, além de deixar cerca de 26 mil feridos graves por ano, com fortes impactos para o governo e para as famílias dos acidentados. Quanto à apuração das causas dos acidentes de trânsito em rodovias federais, estas são analisadas sob três fatores: 1) Fator Humano: subavaliação da probabilidade de acidente; desatenção; cansaço; deficiências (visual, auditiva, motora); consumo de álcool; consumo de drogas; excesso de velocidade; desrespeito à distância mínima entre veículos; ultrapassagem indevida; não-uso de cinto ou de capacete; imprudência de pedestres, ciclistas e motociclistas; 2) Fator Veículo: violência do choque; defeitos de manutenção; utilização incorreta; 3) Fator Infraestrutura: condições de conservação da via; mudanças do contexto da rodovia; evolução do tráfego; condições meteorológicas.

Atualmente, os dados utilizados para esta análise de acidentes de trânsito, sob os fatores humano e veículo, são obtidos exclusivamente da Polícia Rodoviária Federal (PRF) a partir das seguintes bases: 1) Acidentes de Trânsito em Rodovias Federais; e 2) Infrações de Trânsito em Rodovias Federais.

Já os dados que possibilitam a análise de acidentes de trânsito sob o fator infraestrutura pode ser obtidos a partir do Departamento Nacional de Infraestrutura de Transportes (DNIT), que fornece informações como dados de contagem de tráfego e estatísticas de Acidentes.

Ainda sob o fator infraestrutura, os dados geográficos, meteorológicos e pluviométricos podem ser obtidos a partir do Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN).

Quanto às ações de fiscalização e às campanhas de prevenção de acidentes, dado a enorme extensão da malha rodoviária federal, que é distribuída em um país de dimensões continentais como o Brasil, estas acabam sendo pontuais (locais com maior incidência de acidentes) e restritas aos centros urbanos das vias mais importantes. Muitas vezes estas ações são planejadas sem contar com um sistema de apoio à tomada de decisão que unifique as informações provenientes desta diversidade de bases de dados descentralizadas.

Assim, este artigo pretende apresentar uma representação que minimize tais lacunas apresentadas. Para tal, propõe a aplicação de um processo de fusão de dados e informações, pelo qual seja possível integrar, por meio de técnicas apropriadas, diferentes dados oriundos de órgãos como PRF, DNIT e CEMADEN, com o objetivo de disponibilizar uma ferramenta para suporte às ações de fiscalização e às campanhas de prevenção de acidentes nas rodovias federais brasileiras.

3 Fusão de dados e informações

Em um primeiro instante, a definição de Fusão de Dados era dada como um conjunto de métodos para lidar com dados e informações de uma ou mais fontes, realizando associações, correlações e combinações múltiplas entre elas, com o intuito de obter estimativas mais precisas sobre um determinado objeto e possíveis relações entre estes, sendo o processo um ciclo contínuo de estimativas e validações, assim como a aplicabilidade de fontes externas adicionais ou alterações no processo (HALL e JORDAN, 2010).

Entretanto, para Blasch (2013) a Fusão de Dados e Informações é aplicada como sistema para apoiar a avaliação de situações e a tomada de decisão em sistemas complexos. Além disso, possibilita ainda a redução da dimensionalidade dos dados, a agregação de valor à informação, o aumento da representatividade e a produção de

subsídios para a construção do conhecimento sobre situações.

Em razão do crescimento do fluxo de dados e informações, da computação e da tecnologia de sensores, a área de fusão vem se tornando um importante objeto de estudo interdisciplinar. Inúmeras técnicas para combinar dados têm sido exploradas em diferentes áreas tais como: inteligência artificial, processamento de sinais, teoria de controle, entre outras.

No entanto, como forma de auxiliar no avanço da área de fusão, ao realizar a aderência a outras áreas, o modelo *Joint Directors of Laboratories (JDL)* foi o primeiro e mais clássico a surgir com o intuito de normalizar e difundir o conhecimento sobre as bases dos sistemas de Fusão de Dados. Devido sua extrema facilidade de compreensão e padronização para diversos problemas, auxiliando também na avaliação da relevância da solução, o modelo *JDL* é o mais conhecido e amplamente aplicado para demonstrar a utilidade da Fusão de Dados como suporte à tomada de decisão (LLIANS, 2004; HALL e MACMULLEN, 2004).

Este modelo é composto por diferentes módulos e funções que dão suporte e possibilitam a fusão, dentre os quais: preparação e processamento das informações; avaliação de objetos e situações; refinamento das informações e análise de impacto e risco; importação de dados por meio de fontes externas.

O modelo *JDL* apresenta níveis de processos para acomodar e executar os módulos e funções citados acima, sendo eles: Pré-processamento (Nível 0), Identificação e Refinamento de Objetos (Nível 1), Elaboração e Refinamento da Situação (Nível 2), Definição e Refinamento de Risco (Nível 3) e Refinamento do Processo (Nível 4) e Interface Humano-Computador (Nível 5) (STEINBERG et al., 1999). Mais adiante, na seção 5, o modelo *JDL* será descrito de forma mais detalhada.

Assim, ao aplicar a fusão de dados proposta neste artigo, espera-se obter representações mais apuradas dos acidentes de trânsito em rodovias federais brasileiras, a partir do uso de dados disponibilizados pelos seguintes órgãos: PRF, DNIT e CEMADEN. Almeja-se com isso minimizar a ocorrência de

acidentes de trânsito nestas rodovias. A seguir são apresentados alguns trabalhos que, de alguma forma, relacionam-se com a proposta deste artigo.

4 Trabalhos relacionados à aplicação da fusão de dados e informações na prevenção de acidentes de trânsito

Diversos esforços têm sido realizados com o intuito de amenizar as causas dos acidentes de trânsito e beneficiar a segurança rodoviária e veicular. Ryder et al. (2017) desenvolveram um sistema de apoio à decisão para veículos, que faz uso de alertas aos condutores informando o risco do trajeto a ser percorrido com base no histórico de acidentes do local. Para isso, no momento em que o condutor se aproxima de um local onde já tenha sido registrada uma ocorrência de trânsito, automaticamente esse condutor é notificado pelo sistema para que diminua a velocidade e tenha prudência naquele trecho.

Diferentemente do trabalho descrito acima, o trabalho proposto por Abulatif (2018) apresenta um processo de integração de diversas fontes de dados relacionadas à acidentes de trânsito, de modo que os dados gerados por tal processo sejam utilizados para subsidiar a elaboração de ações de segurança viária em cinco capitais brasileiras: Belo Horizonte/MG, Campo Grande/MS, Curitiba/PR, Palmas/TO e Teresina/PI. Para tanto, foi realizada a integração de dados do Sistema de Informação sobre Mortalidade (SIM) e do Sistema de Informação Hospitalares (SIH), ambos mantidos pelo Ministério da Saúde.

Em uma abordagem diferente, Sohn et al. (2003) fizeram uso da análise de dados, por meio de algoritmos de agrupamento e fusão de dados, para verificar o nível de gravidade dos acidentes de trânsito em rodovias da Coreia do Sul. Para tal, utilizaram dados como a largura da via, a velocidade antes do acidente, o formato do carro, além da verificação da existência de dispositivos de proteção nos veículos.

Além da utilização de algoritmos de agrupamento e fusão de dados para a descrição de acidentes de trânsito, pode-se citar o trabalho de Sun et al. (2018), onde os autores tiveram a ideia de prever a ocorrência de acidentes de trânsito por meio da

conscientização de incidentes e anomalias. Para tal, utilizam dados reais de uma estrada chamada Kunshi, onde são coletados registros como: taxa de fluxo, velocidade média, entre outras informações. Mediante esses dados é proposto um método com base nos vizinhos mais próximos, chamado *k-Nearest Neighbors (kNN)*, muito utilizado para classificação de dados em aprendizado de máquina.

Costa et al. (2014) propuseram a aplicação de mineração de dados nos boletins de ocorrências registrados nas rodovias federais brasileiras pela Polícia Rodoviária Federal no ano de 2012, com o intuito de identificar associações entre variáveis relacionadas aos acidentes de trânsito neste tipo de rodovia. Para isso, utilizaram algoritmos de aprendizado de máquina visando a extração de modelos e padrões.

Com base no contexto e nos trabalhos supracitados, a seguir é descrita a proposta deste artigo, que visa apresentar um processo fusão de dados e informações para ser aplicado na prevenção de acidentes de trânsito nas rodovias federais brasileiras.

5 A Fusão de Dados e Informações no contexto da prevenção de acidentes de trânsito nas rodovias federais brasileiras

Para cumprir com o objetivo deste trabalho de pesquisa, é apresentada a seguir uma proposta de aplicação da fusão de dados e informações na prevenção de acidentes de trânsito no Brasil. Cabe destacar que essa proposta seguirá os princípios do modelo *JDL*, mencionado anteriormente, mas, no entanto, não seguirá todos os níveis que o modelo dispõe, uma vez que o modelo permite esse tipo de adaptação (STEINBERG et al., 1999).

O objetivo do processo de fusão aqui descrito é proporcionar uma melhor representação, organização e confiabilidade das informações, de forma a auxiliar no planejamento e tomadas de decisão voltadas para prevenção de acidentes de trânsito nas rodovias federais brasileiras, dado o alto nível de variação, inconsistência e inconfiabilidade das informações que são disponibilizadas pelos órgãos diretamente envolvidos nesta questão.

Nos parágrafos seguintes são descritos os passos de cada nível do processo, com

base no modelo *JDL*, que será empregado nesta proposta de aplicação da fusão de dados e informações. O processo de fusão aqui proposto utilizará apenas os níveis de pré-processamento, identificação de objetos, construção de situações e interação humano-computador.

Como ponto de partida, o processo de fusão de dados e informações necessita de algum tipo de insumo para que sejam processados os dados e informações. No contexto de acidentes de trânsito deste trabalho de pesquisa, são utilizados três tipos de fontes, sendo elas: 1) DNIT, fornecendo informações do tráfego das rodovias federais; 2) PRF, fornecendo informações sobre a quantidade de acidente e infrações nas vias; 3) CEMADEN, fornecendo informações de dados geográficos e das condições meteorológicas das vias.

Com esses três tipos de fontes iniciais, é possível obter um volume de informações satisfatórios para aplicação do processo de fusão, e assim, seguir para o nível 0 (de acordo com o modelo *JDL*). No entanto, pode haver mais fontes além das mencionadas anteriormente caso seja necessário. Cabe destacar que tais dados são fornecidos de forma estática pelos órgãos e entidades responsáveis, não entrando assim este trabalho na questão dos dados fornecidos de forma dinâmica e em tempo real.

A etapa de pré-processamento, conhecida como nível 0 do modelo *JDL*, se caracteriza como nível de entrada dos dados e informações, e de preparação dos mesmos para os níveis de fusão seguintes (HALL e JORDAN, 2010).

No contexto deste trabalho de pesquisa, as informações advindas das fontes mencionadas anteriormente, recebem alguns tratamentos como limpeza, normalização e padronização das informações, com objetivo de organizar e facilitar na identificação, análise e recuperação das mesmas, e assim, facilitar a aplicação das técnicas do nível 1 (identificação de objetos) do modelo de fusão.

No nível 1, o processo de análise e associação de dados de múltiplas fontes tem como intuito definir, da maneira mais confiável possível, quais as entidades estão envolvidas no processo, como por exemplo, associar características de infrações de trânsito aos

acidentes em uma determinada rodovia, e por sua vez, às informações climáticas.

Tal processo de associação oferece a possibilidade de se descobrir uma concentração grande de um dado tipo de situação, como por exemplo, o volume de acidentes de trânsito em um trecho específico de uma rodovia federal, ou até mesmo classificar uma região da rodovia com alto risco de acidentes provocados por excesso de velocidade. Todos os pontos descritos anteriormente se caracterizam com base em técnicas computacionais aliadas ao conhecimento de especialistas neste tipo de domínio.

Ao realizar a identificação dos objetos das múltiplas fontes de dados, o nível 2 elabora uma interpretação contextual dos dados, ou seja, analisa como os objetos encontrados no nível 1 estão relacionados com o ambiente em que estão empregados (HALL e JORDAN, 2010).

Um exemplo disso, é ao identificar em quais rodovias de pista simples o nível de acidentes por excesso de velocidade ocorre devido ao volume de tráfego de veículos de grande porte (caminhões ou ônibus), na qual uma possível medida de prevenção seria incluir mais faixas adicionais a esse tipo de via ou colocar radares em localizações estratégicas para averiguar, a médio ou longo prazo, se o índice de acidentes irá diminuir.

Já os níveis 3 e 4 do modelo *JDL* não serão utilizados nesta proposta, devido ao alto nível de complexidade e falta de padronização das próprias fontes, onde suas informações em certos casos são divergentes e acarretam interpretações erradas da situação.

Por fim, no nível 5, o processo de fusão busca apresentar as informações ao usuário de forma mais clara e compreensível, para auxiliar na tomada de decisão. As informações devem ser representadas, por meio de indicadores de situações, tais como: nível de acidentes de trânsito, características envolvendo situações nível excessivo de excesso de velocidade em quilômetros específicos, entre outras informações.

6 Resultados Esperados e Considerações Finais

Ao realizar o processo de fusão de dados e informações proposto neste trabalho, é

esperada a obtenção de informações mais claras e coerentes, com alto nível de confiabilidade e consistência, para possibilitar o máximo de entendimento das situações identificadas, e quais ações de prevenção podem ser adotadas, bem como os fatores de relevância que podem ser integrados para melhoria de todo o processo de fusão.

Como trabalhos futuros, pretende-se a aplicação do processo de fusão proposto neste artigo em um caso prático, com o intuito de validar se os resultados obtidos a partir deste processo auxiliam na representação das informações referentes à prevenção de acidentes de trânsito em uma rodovia federal brasileira.

Referências

ABULATIF, L. I. Processo de integração de dados: um modelo de gestão da informação para múltiplas bases de dados de acidentes de trânsito no Brasil. **Epidemiologia e Serviços de Saúde**. Vol. 27, e2017160, 2018.

BLASCH, E. et al. Revisiting the JDL model for information Exploitation. In: Proceedings of the 16th **International Conference on Information Fusion**. IEEE, p. 129-136, 2013.

BOTEGA, L. C, et. al. Quality-aware human-driven information fusion model. In **International Conference on Information Fusion**. IEEE Computer Society. p. 1-10, Xian, 2017.

COSTA, J, J., BERNARDINI, F. C., VITERBO J. F. A mineração de dados e a qualidade de conhecimentos extraídos dos boletins de ocorrência das rodovias federais brasileiras. **AtoZ: novas práticas em informação e conhecimento**. Vol. 3, n. 2, p. 139-157, 2014.

Departamento Nacional de Infraestrutura de Transportes (DNIT). **Operações rodoviárias**. 2014. Disponível em: <<https://www.dnit.gov.br/rodovias/operacoes-rodoviaria>>.

HALL, D. L., MCMULLEN, S. A. H. **Mathematical Techniques in Multisensor Data Fusion**. [SI]: Artech House, 2004.

HALL, D. L., JORDAN, J., **Human-centered information fusion**. [SI]: Artech House, 2010.

INSTITUTO NACIONAL DE PESQUISAS ECONÔMICAS APLICADAS (IPEA). **Acidentes de trânsito nas rodovias federais brasileiras: caracterização, tendências e custos para a sociedade**. Relatório de pesquisa. Brasília. 2015. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/relatoriopesquisa/150922_relatorio_acidentes_transito.pdf>.

LLINAS, J., BOWMAN, C., ROGOVA, G., STEINBERG, A., WALTZ, E., WHITE, F. Revisiting the JDL data fusion model II. In: **Space and Naval Warfare Systems Command**. San Diego, CA. 2004.

MINISTÉRIO DA INFRAESTRUTURA. **Programa Rodovida**. Disponível em: <<http://infraestrutura.gov.br/rodovida.html>>.

Organização Pan-Americana de Saúde (OPAS). **Folha Informativa - Acidentes de Trânsito**. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5147:acidente-s-de-transito-folha-informativa&Itemid=779>.

RYDER, B., GAHR, B., EGOLF, P., DAHLINGER, A. Preventing Traffic Accidents with In-Vehicle Decision Support Systems - The Impact of Accident Hotspot Warnings on Driver Behavior. **Decision Support Systems**. vol. 99, pp. 64-74, 2017.

SOHN, S. Y., LEE, S. H. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. **Safety Science**. Vol. 41, n. 1, p. 1-14, 2003.

STEINBERG, A, N., BOWMAN, C. L., WHITE, F. E., Revisions to the JDL data fusion model. In: **Proceedings of SPIE**. Vol. 3719, n.1, p. 430-441, 1999.

SUN, B. et al. Anomaly-Aware Traffic Prediction Based on Automated Conditional Information Fusion. In: **2018 21st International Conference on Information Fusion (FUSION)**. IEEE, p. 2283-2291, 2018.