

MAC0499 - Proposta de Trabalho

Análise de Redes Sociais com Uso de Aprendizado de Máquina para Prever o Tráfego de Veículos em Zonas Urbanas

Lucas de Carvalho Dias
Orientador: Roberto M. Cesar Jr

28 de abril de 2017

1 Contextualização

As grandes cidades são compostas de milhões de elementos que se movimentam e interagem entre si de forma dinâmica. Tal conjunto gera uma gigantesca máquina complexa, caótica e heterogênea. E dessa mesma forma são seus dados. Um massivo conjunto de dados que se movimentam rapidamente por toda a cidade, carregados de informações sobre o que acontece a todo momento (incluindo meta-dados). Tais dados são de uma natureza complexa para que seres humanos os dissequem em procura de padrões que forneçam alguma previsibilidade sobre o que está por vir. Felizmente, com o nível de tecnologia atual, já é possível lidar com esse volume de dados usando técnicas computacionais. Essa abordagem é conhecida como informática urbana (*urban informatics*), usada para extrair, a partir dos volumosos dados, compreensão sobre aspectos inerentes ao funcionamento das cidades, tais como mobilidade urbana, violência, economia, etc. De fato, essa já é uma estratégia de interesse de muitas prefeituras de cidades. Kitchin [4], em um artigo sobre cidades inteligentes (*Smart Cities*), cita o exemplo do Centro de Operações da prefeitura do Rio de Janeiro, uma parceria do governo municipal e da IBM que tem como tarefa investigar dados obtidos de trinta agências (dados de trânsito, serviços de emergência, previsão do tempo, etc.) Tais dados são analisados, com uso de algoritmos, por analistas para compreender aspectos da vida na cidade e construir modelos de previsão. A área de informática urbana tem se desenvolvido intensamente nos últimos anos. Técnicas de ciência de dados estão sendo criadas para analisar problemas como violência urbana [2], trânsito intenso [6], acidentes de trânsito [3] e gasto energético [5]. O presente projeto se encaixa nesse contexto. Utilizando técnicas de mineração de dados, informações provenientes de redes sociais, aplicativos de trânsito e fontes de notícia na internet serão extraídas, filtradas e formatadas. Ao aplicar os métodos de aprendizado de máquina nessas informações, serão obtidas conclusões so-

bre como o trânsito é afetado por diferentes eventos que ocorrem na cidade (eventos de origem humana ou natural). Essas conclusões também serão empregadas com as técnicas de aprendizado de máquina, a fim de fazer previsões sobre como eventos futuros interferirão no fluxo do trânsito. O presente projeto é particularmente influenciado pelo trabalho de Di Wang et al. [7], no qual foi desenvolvida uma metodologia semi-supervisionada para fazer previsões sobre o trânsito usando a rede social Twitter. Adicionalmente, o presente projeto propõe usar dados de mais de uma rede social e também de outras fontes da internet (como dados meteorológicos).

2 Motivação

Na modernidade, um dos principais problemas que assola a vida dos moradores das grandes cidades é o congestionamento nas vias principais. A análise de dados de redes sociais se tornou um tópico de pesquisa bastante rico e complexo, que pode ser usada para extrair conhecimento sobre diversos comportamentos sociais [1]. Como alguns dados de redes sociais possuem a informação do local de onde foram publicados, eles podem ser empregados a fim de se obter informações sobre a mobilidade na cidade. Também, é sabido que chuvas influenciam bastante na qualidade do trânsito, de forma que dados meteorológicos podem fornecer informações úteis sobre a mobilidade. Motivado por esses fatos, este projeto objetiva estudar como dados de redes sociais e dados meteorológicos expressam informações sobre o estado do trânsito. Para isso, são usadas técnicas de *data mining* e aprendizado de máquina (*machine learning*) para aprender padrões sobre dados e, posteriormente, usá-los para fazer previsões a partir de novos dados.

3 Objetivo

O objetivo deste projeto é estudar e desenvolver técnicas de mineração de textos (*text mining*) e aprendizado de máquina para analisar a dinâmica da movimentação de veículos em ambientes urbanos. Para isso, serão utilizados dados obtidos de redes sociais e de outras tipos fontes online, tais como portais de notícias.

4 Plano e Cronograma

– Janeiro, Fevereiro e Março

- Definir redes sociais (por exemplo, Waze) e fontes de notícia na internet que forneçam uma boa cobertura dos acontecimentos na

cidade. Ainda, nessas fontes de dados, deve ser possível minerar informações e navegar automaticamente com uso de um robô Também procurar pelas aplicações que forneçam informações abrangentes sobre o fluxo de veículos em áreas urbanas, e que nas quais também seja possível utilizar um robô para minerar e navegar pelos dados. **(CONCLUÍDO)**

- Definição das bibliotecas de Python que serão usadas para a extração dos dados e as que serão usadas para aprendizagem de máquina. **(CONCLUÍDO)**

- Implementar o robô que irá navegar pelas redes sociais e portais de notícias para captar seus dados. **(CONCLUÍDO)**

– **Abril, Maio e Junho**

- Implementar o robô que irá captar os dados das aplicações de trânsito. **(CONCLUÍDO)**
- Coletar os Dados com os robôs e organizá-los. (Em Andamento)

– **Julho, Agosto e Setembro**

- Desenvolver métodos de aprendizado de máquina para analisar os dados obtidos.
- Aplicar nos dados obtidos os métodos desenvolvidos.

– **Outubro e Novembro**

- Desenvolver métodos para explorar as conclusões obtidas a fim de fazer previsões sobre eventos futuros. Também, quando ocorridos esses eventos, as previsões serão analisadas acerca de o quanto se aproximaram da realidade.
- Escrever monografia

Referências

- [1] Samuel Barbosa, Dan Cosley, Amit Sharma, and Roberto M. Cesar, Jr. Averaging gone wrong: Using time-aware analyses to better understand behavior. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 829–841, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.
- [2] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*, pages 427–434. ACM, 2014.
- [3] Miao Chong, Ajith Abraham, and Marcin Paprzycki. Traffic accident analysis using machine learning paradigms. *Informatica*, 29(1), 2005.
- [4] Rob Kitchin. The real-time city? big data and smart urbanism. *Geo-Journal*, 79(1):1–14, 2014.

- [5] Cynthia Rudin, David Waltz, Roger N Anderson, Albert Boulanger, Ansaf Salieb-Aouissi, Maggie Chow, Haimonti Dutta, Philip N Gross, Bert Huang, Steve Ierome, et al. Machine learning for the new york city power grid. *IEEE transactions on pattern analysis and machine intelligence*, 34(2):328–345, 2012.
- [6] Shiliang Sun, Changshui Zhang, and Guoqiang Yu. A bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):124–132, 2006.
- [7] Di Wang, Ahmad Al-Rubaie, John Davies, and Sandra Stinčić Clarke. Real time road traffic monitoring alert based on incremental learning from tweets. In *Evolving and Autonomous Learning Systems (EALS)*, 2014 *IEEE Symposium on*, pages 50–57. IEEE, 2014.