



Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital

SmartMetropolis – Plataforma e Aplica es para Cidades Inteligentes

WP6 – An lise e Visualiza o de Dados

Relat rio de Atividades do Primeiro Trimestre do WP6 - An lise e Visualiza o de Dados

Natal-RN, Brasil
[Abril de 2016]

Equipe Técnica

Prof. Dr. Daniel Sabino Amorim de Araújo (Coordenador) – [IMD/UFRN]

Prof. Dr. Allan de Medeiros Martins - DEE/UFRN

Prof. Dr. Adrião Duarte Dória Neto - DCA/UFRN

Prof. Dr. Nélio Alessandro Azevedo Cacho - DIMAp/UFRN

Discentes

Camila Nascimento dos Santos - Mestrado (Pós-graduação em Eng. Elétrica e de Computação)

Felipe de Souza Pinheiro - Mestrado (Pós-graduação em Eng. Mecatrônica)

Leandro Albino Ferreira - Iniciação Científica (Engenharia de Computação)

Maximiliano Araújo da Silva Lopes - Doutorado (Pós-graduação em Eng. Elétrica e de Computação)

Mickael Raninson Carneiro Figueredo - Iniciação Científica (Ciência e Tecnologia)

Adelson Dias de Araújo Júnior - Iniciação Científica (Engenharia de Computação)

Sumário

1	Introdução	6
2	Metodologia	6
2.1	Mineração de Textos	7
2.2	Técnicas de Visualização	7
2.3	Processamento de Sinais	8
2.4	Processamento de Imagens	9
3	Andamento das Atividades e Resultados Obtidos	9
3.1	Mineração de Textos	10
3.2	Técnicas de Visualização	12
3.3	Processamento de Sinais	14
3.4	Processamento de Imagens	14
3.5	Redução de Dimensionalidade	14
4	Considerações Finais	14

Lista de Figuras

1	Sinais referentes a situação da mão.	8
2	Análise de sentimentos de tweets em língua inglesa durante a Copa do Mundo 2014. . .	11
3	Análise de sentimentos de tweets em língua portuguesa durante a Copa do Mundo 2014.	11
4	Tela do Módulo Dashboard Web	13
5	Arquitetura FIWARE Data/Context.	13

Lista de Tabelas

1 Introdução

O ‘WP6 - Análise e Visualização de Dados’, doravante denominado apenas por WP6, é parte integrante do projeto SmartMetropolis que tem como objetivo geral desenvolver soluções de tecnologia da informação que ajudem a resolver problemas das cidades e regiões metropolitanas atuais. Ou seja, o SmartMetropolis trata do desenvolvimento de soluções inteligentes para tratar problemas, principalmente, em pontos críticos de regiões urbanas e em crescimento, como segurança pública, saúde, educação, trânsito e qualidade de vida.

O WP6, por sua vez, é um dos seis grupos de trabalho que compõem o SmartMetropolis. Ele é responsável por realizar análises profundas nas massas de dados geradas por aplicações desenvolvidas pelos demais grupos de trabalho. Essa análise pode fazer parte do foco principal da aplicação ou ser secundária e tratar dos dados gerados por ela. Em ambos os casos, a análise tem como objetivo principal extrair informações úteis que estejam “embutidas” nos dados para ajudar na tomada de decisão.

As análises de dados são realizadas, principalmente, por técnicas das áreas de Aprendizado de Máquina e Estatística. No entanto, o grupo não fica restrito somente a essas áreas e também utilizam dentro do processo de análise a Visão Computacional, Teoria da Informação, Processamento de Linguagem Natural, entre outras.

O restante deste relatório é composto pela Seção 2, que detalha a forma que o grupo vem trabalhando e dá breves explicações sobre os tópicos abordados; a Seção 3, que expõe as atividades realizadas até o momento pelos membros do WP e, por fim, a Seção 4, que traz as principais condireções sobre o primeiro trimestre de trabalho do WP6.

2 Metodologia

Com o objetivo de organizar as atividades do WP6, decidimos por dividir as principais áreas de estudo entre os integrantes do grupo. Para isso, foram coletadas as áreas de atuação que teriam prioridade para análise de dados e atribuímos a cada membro discente do WP6. Dessa forma, cada integrante deve se tornar especialista em uma dessas áreas e conseguir canalizar as demandas para si, tornando o processo de desenvolvimento mais rápido. Os docentes, de uma maneira geral, estão envolvidos em todas as frentes de trabalho do WP.

A realização de reuniões periódicas do grupo e apresentação dos trabalhos faz com que, apesar de especialistas em uma das áreas, todos os membros tenham conhecimento sobre as atividades dos outros. Dessa forma, em casos de atividades que exijam mais de uma especialidade, os integrantes podem trabalhar em conjunto sem maiores problemas.

Diante disso, o grupo adotou as seguintes linhas de trabalho:

- Mineração de textos (Mickael Figueiredo)
- Técnicas de visualização (Leandro Ferreira)
- Processamento de sinais (Felipe Pinheiro)
- Processamento de imagens (Camila Santos)
- Redução de dimensionalidade (Maximiliano Lopes)

Nas subseções seguintes, as linhas de trabalho são descritas com um pouco mais de detalhe.

2.1 Mineração de Textos

O Processamento de Linguagem Natural (PLN) [1], em um sentido amplo, cobre qualquer tipo de manipulação de um computador sobre uma linguagem utilizada para a comunicação entre seres humanos. Sistemas de geração de linguagem natural convertem informação de bando de dados de computadores, por exemplo, em linguagem normalmente compreensível ao ser humano, e sistemas de compreensão de linguagem natural convertem ocorrências de linguagem humana em representações mais formais, mais facilmente manipuláveis por programas de computador.

O Processamento de Linguagem Natural, ou *Natural Language Process*, como é usado mais comumente na literatura possui uma série de aplicações e também limitações que estão sendo constantemente superadas. Atualmente, podemos citar aplicações do PLN em recuperação de informação, que lida com o armazenamento de documentos e a recuperação automática de informação associada a eles, tradução automática, geração automática de texto, geração de linguagem natural, interpretação da linguagem natural, simplificação de texto, correção ortográfica e reconhecimento vocal.

Em termos de aplicações associadas à *Smart Cities*, o PLN coloca à disposição uma grande ajuda quando se trata de processamento de Big Data. Atualmente um volume imenso de dados é gerado oriundo das redes sociais. Esses dados são de uma riqueza de informação impressionante. Dentro desse contexto, é possível produzir aplicações para explorar o conteúdo postado por usuários nas redes sociais como Twitter e Facebook. A análise de sentimentos é uma dessas aplicações e se baseia na ideia de classificar um dado texto em linguagem humana como positivo, negativo ou neutro em um determinado contexto.

2.2 Técnicas de Visualização

À medida que o volume e a variedade de dados, utilizadas nas mais diversas aplicações, se tornam cada vez maiores, mais necessário se torna a utilização de novas técnicas de visualização. Por sua vez, a criação de novas técnicas de visualização acabam incentivando o uso de mais análises de dados, pois mostram padrões anteriormente desconhecidos, fazendo com que exista uma relação circular destas técnicas com o crescimento das demandas de análises de dados. A apresentação das informações de forma clara, de modo a permitir aos usuários o consumo e extração de valor delas, é fundamental.

A linha de trabalho de técnicas de visualização de dados é uma área de aplicação de computação gráfica interativa extremamente importante no contexto de *big data* e cidades inteligentes, pois permite a análise visual de dados estruturados e não estruturados, capturados a partir de fontes internas e externas (como leituras de sensores em tempo real, arquivos históricos, atualizações sobre o tráfego, avisos de segurança pública e mídia social), e descobrir informações úteis e inteligentes, além de oportunidades até então escondidas, podendo levar à tomada de melhores decisões e com mais confiança. Além disso, melhores decisões podem significar maior eficiência operacional, redução de risco e redução de custos.

Como seres humanos, a grande maioria das informações que captamos é visual, o cérebro consegue interpretar melhor os dados que estão dispostos dessa forma. Por exemplo, é bem mais difícil e trabalhoso de se identificar tendências e correlações entre dados em uma planilha do que em gráficos e imagens. A partir daí vemos a importância dessa área: identificação de coisas que estão acontecendo, ou que irão acontecer, análise de padrões e correlações de dados antes invisíveis, através do uso de recursos gráficos e animação de dados, potencializando a apropriação da informação pelo consumidor. Grandes empresas e startups inovadoras estão investindo nesta área, pois estão vendo as novas necessidades dos usuários, que estão cada vez mais em busca de novas formas de acessar, visualizar e utilizar seus dados de maneira fácil, intuitiva e segura.

2.3 Processamento de Sinais

Dada a imensa quantidade de dados adquiridos pela tecnologia implantada nas cidades inteligentes, se faz necessário o processamento desses dados para compreender melhor o que significam.

Para identificar padrões nos dados, uma das técnicas que podem ser utilizadas é a identificação de sistemas. Essa técnica identifica modelos de acordo com os dados em tempo real, avaliando qual o modelo regressivo que melhor representa o que está acontecendo e baseado em dados anteriores pode ser utilizada para determinar padrões nos acontecimentos e influenciar tomadas de decisões.

Utilizando um banco de dados para determinar o que caracteriza cada modelo, é possível comparar o modelo encontrado em tempo real e determinar ações para cada situação. Dessa maneira as medidas a serem tomadas como reação à cada situação podem ser feitas em tempo real.

Um exemplo de como a técnica pode ser aplicada é na identificação de modelos para movimentos da mão a partir da resposta elétrica da musculatura do antebraço. O procedimento para fazê-lo foi exatamente o mesmo citado anteriormente: coleta de dados, análise dos dados e aplicação do método.

A coleta dos dados foi feita com quatro pares de sensores cutâneos sobre a pele distribuídos de forma simétrica ao redor do antebraço, um circuito para amplificar o sinal analógico obtido e um conversor analógico/digital para uma análise computacional. O resultado dessa aplicação foi bem sucedida para os movimentos propostos: mão aberta, mão fechada, mão flexionada para a esquerda e mão flexionada para a direita. Podemos analisar de a identificação dos movimentos de maneira gráfica nos gráficos da Figura 1.

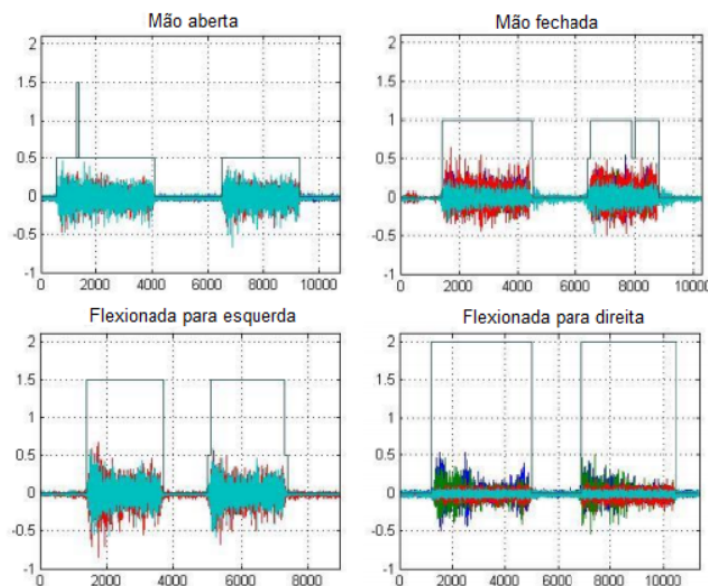


Figura 1: Sinais referentes a situação da mão.

Foram atribuídas amplitudes para melhor visualizar os resultados nos gráficos onde 0 está associado ao relaxamento, 0,5 a mão aberta, 1 a mão fechada, 1,5 a flexionada para esquerda e 2 a flexionada para direita. Analisando os gráficos podemos perceber que o identificação de sistemas obteve sucesso ao modelar sinais aparentemente aleatórios. Após identificados os padrões, é possível implementar para as mais diversas aplicações.

2.4 Processamento de Imagens

As cidades modernas são descritas pela rápida mudança de cenário e o grande volume de dados disponíveis sobre as mesmas. Um dos principais tipos de dados são as imagens. A análise manual rápida desses dados é muitas vezes complicada ou até mesmo impossível. Uma prática muito difundida é o desenvolvimento do sistema de análise para cada problema separado (por exemplo identificação de carros, identificação de números ou QR-codes) que conduz a um certo número de dificuldades.

Por este motivo, neste trimestre foram estudados métodos de processamento de imagens, com o intuito de facilitar a análise das imagens provenientes das zonas urbanas. O processamento digital de imagem (PDI) é uma área da eletrônica em que imagens são convertidas em matrizes de números inteiros, sendo que cada elemento desta matriz é composto por um elemento fundamental: o pixel (abreviação, do Inglês: *picture element*). A partir desta matriz de pixels, que representa a imagem, diversos tipos de processamento digital podem ser implementados por algoritmos computacionais. A aplicação destes algoritmos realiza as transformações necessárias para que se possa, por exemplo, obter uma imagem com os realces pretendidos ou extrair atributos e informações pertinentes [4]

Sendo assim, para que uma imagem possa ser manipulada por computador, segundo [9], é necessário que essa imagem seja discretizada ou digitalizada, tanto espacialmente quanto em sua amplitude. Com isso, uma imagem pode ser discretizada através de um processo que envolve dois passos: (i) amostragem, no qual são definidos os valores para as coordenadas x e y da matriz de pontos, e (ii) quantização, que consiste em escolher um valor múltiplo relacionado à banda de frequência da intensidade para cada ponto da imagem.

Alguns autores possuem conceitos diferentes quanto à taxonomia do processo de um sistema PDI. Entretanto, seja uma classificação mais detalhada ou não do processo, todos seguem uma mesma linha que, neste caso, segue a classificação [4]: formação e aquisição da imagem, pré-processamento, segmentação, representação e descrição e reconhecimento e interpretação.

O nosso estudo, se focou no estágio de segmentação, na qual, o seu objetivo é obter, a partir de uma imagem digitalizada pré-processada, um conjunto de primitivas ou segmentos significativos, que contém a informação semântica relativa à imagem de origem. Usualmente as primitivas usadas são naturais, em geral contornos e regiões. Os diferentes tratamentos utilizados levam à noção de contornos, regiões ou a uma noção mista.

A segmentação é considerada, dentre todas as etapas do PDI, a mais crítica do tratamento da informação. É nesta etapa que são definidas as regiões de interesse para processamento e análise posteriores. Como consequência, quaisquer erros ou distorções presentes nesta etapa refletem-se nas demais, de forma a produzir ao final do processo resultados não desejados, podendo contribuir de forma negativa para a eficiência de todo o processamento, premissa essa também verdadeira para o caso inverso [3].

3 Andamento das Atividades e Resultados Obtidos

De acordo com o cronograma estabelecido inicialmente, o WP6 possui como primeiro entregável o seguinte item:

“Estudo sobre as principais técnicas de análise de grandes volumes de dados (Fase 1).”

Sendo assim, todos dos membros passaram o primeiro trimestre do projeto pesquisando e estudando a literatura de referência para sua linhas de trabalho. No entanto, alguns discentes já estavam inseridos em alguma pesquisa em andamento e utilizaram o conhecimento prévio para produzir avanços significativos. Tais avanços serão detalhados por linha de trabalho, conforme feito na Seção 2.

3.1 Mineração de Textos

O trabalho foi dedicado a um analisador de sentimento que fosse capaz de trabalhar em duas línguas: inglês e português. Para isso, foi necessário buscar uma fonte de dados para testes de classificação de polaridade. Em nosso caso, o interesse é na análise de textos de tweets. A captura dos tweets é feita utilizando uma API em Java (Twitter4j¹) para integrar a plataforma com um twitter service. Usando a função *Stream* da API, é possível coletar dados de acordo com um conjunto de filtros que podem ser definidos. A partir daí, gera-se um conjunto de dados, armazenando-se: nome do usuário, identificador, idade, latitude e longitude do post, data e hora da postagem, cidade e corpo do tweet. Alguns destes dados, como a cidade do usuário, só são obtidos quando o usuário permite seu acesso no seu perfil do Twitter.

O próximo passo depois de identificar o idioma é retirar algumas construções comumente utilizadas por usuários nas redes sociais. Em geral, as ferramentas de processamento de linguagem natural dependem de um texto escrito na forma mais formal possível. Por este motivo foram adicionados dois níveis para fazer a retirada (limpeza), responsáveis por retirar abreviações, URLs, imagens e *emojicons* dos dados.

Na sequência é realizada a polarização das postagens. A polarização consiste em identificar se uma postagem tem sentimento positivo, negativo ou neutro sobre determinado assunto. Na versão atual da plataforma são suportados os polarizadores em inglês e português. Para a polarização do inglês foi utilizada a biblioteca Stanford CoreNLP [8]. Para a polarização das postagens em português foi criado um componente de análise de sentimento que atendesse a estrutura de execução definida pela infraestrutura de processamento em tempo real. A base desse componente é o SentiLex [12], um léxico de sentimento para o português, constituído por 7.014 lemas e 82.347 formas flexionadas. As classes de palavras presentes nesse contexto são adjetivos, nomes, verbos e expressões idiomáticas.

Em seguida, o Lemantizador é utilizado de acordo com o idioma para retirar a inflexão das palavras empregadas, gerando elementos linguísticos genéricos. A ferramenta *Lemmatizer for Portuguese* [1] foi utilizada para efetuar o processamento textual no lematizador em português enquanto que o Stanford CoreNLP foi utilizado para lematizar as postagens em inglês. Por fim, os resultados são analisados e responsáveis por manter atualizados os contadores de termos (palavras, hashtags) mais utilizados no twitter, sendo possível fazer um conceito similar ao top trends.

Uma série de aplicações podem ser construídas sobre essa ferramenta desenvolvida. Utilizando um banco de dados oriundo do tema da Copa do Mundo Fifa 2014, podemos testar nossa ferramenta. Os resultados são apresentados nos gráficos a seguir:

A partir dos gráficos, pode-se perceber que de tweets oriundos da língua inglesa, a copa do mundo obteve uma repercussão bem mais positiva durante a copa do mundo. Partindo para uma análise dos tweets em português, podemos ver que a quantidade de textos negativos predomina durante maior parte da copa. Após alguns estudos, podemos perceber que é oriundo da onda de protestos ocorrencio no Brasil na mesma época. Entratando o cenário se inverte em dias como a derrota da Espanha para a Holanda e um caso curioso da Copa do mundo: a mordida de Luís Suarez.

Adicionalmente a pesquisa mencionada anteriormente, inciou-se um estudo de classificação utilizando Aprendizado de Máquina e Mineração de Dados de uma ferramenta disponibilizada pela Universidade de Waikato: O Weka [5]. Com essa ferramenta, composta por uma gama de métodos e aplicações de processamento de linguagem Natural, podemos criar um classificador para relacionar postagens em redes sociais com crimes conhecidos pela polícia.

¹Disponível em: <http://twitter4j.org/en/index.html>

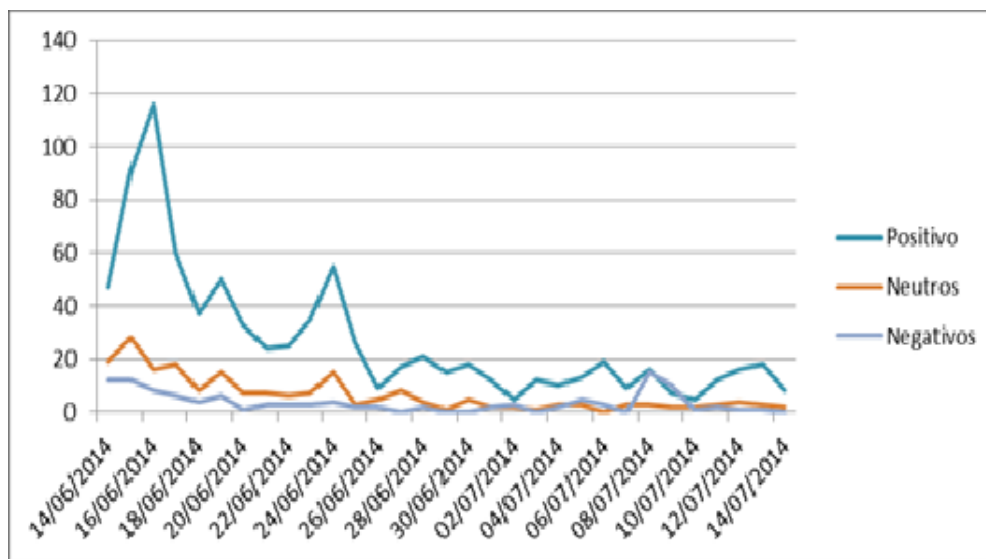


Figura 2: Análise de sentimentos de tweets em língua inglesa durante a Copa do Mundo 2014.

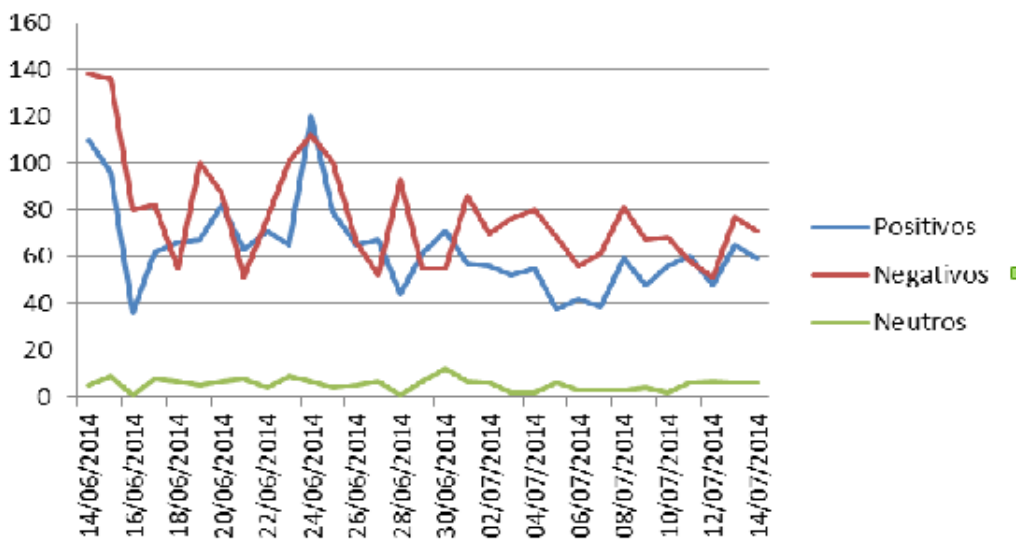


Figura 3: Análise de sentimentos de tweets em língua portuguesa durante a Copa do Mundo 2014.

Para a criação de um classificador utilizando o Weka, seria necessário uma base de dados de qualidade, e com uma boa quantidade de dados para o aprendizado de máquina. Quanto mais dados, e quanto maior a qualidade deles, teríamos um classificador mais preciso o possível.

Foram utilizados dados das ocorrências da cidade de Natal de um determinado período de tempo. Cada ocorrência possui um código que define seu tipo. Então, a ideia seria criar um classificador que fosse capaz de pegar um texto, e coloca-lo dentro de algum desses códigos com a maior precisão possível. Para que isso fosse possível, foram testados alguns filtros para retirar uma amostra significativa dos dados e classificadores, como o *Support Vector Machines* [2] e Naive Bayes [2].

As aplicações de um classificador deste tipo são inúmeras. Está ferramenta poderia ser utilizada como um indificador de crimes em redes sociais, e tendo as geolocalização desses textos, poderíamos assim ter noção de áreas da cidade com maior incidência de um determinado tipo de crime. Porém, a ferramenta utilizada para isso, no caso, o Weka, demanda um processamento muito grande da máquina que está suportando. O objetivo agora, seria utilizar novos métodos para atingir um precisão tão boa quanto o Weka, em outra ferramenta de exigência menor.

3.2 Técnicas de Visualização

Com relação ao que já se foi implementado, deve-se descrever que integrado ao WorkPackage 2, de Aplicações, trabalhou-se com um módulo de visualização dos dados espaço-temporais das viaturas policiais. Os conjuntos de dados trabalhados envolvem as trajetórias das viaturas durante cerca de uma semana e o objetivo do trabalho que foi realizado foi implantar um dashboard que utilize dos recursos do OpenStreetMap² para abrir arquivos KML ou localizações comuns. Esse módulo web dashboard, então, tem o objetivo de disponibilizar para visualização as trajetórias das viaturas, as ocorrências pela cidade (dados recolhidos pelo Centro Integrado de Operações de Segurança Pública, CIOSP), o rastreo das viaturas no tempo desejado, e ainda os polígonos AISPS, que são regiões de ação da polícia.

Assim, com o foco do dashboard delimitado na Figura 5, trabalhar com essas trajetórias parece ser a análise mais importante. Essas trajetórias, tuplas definidas $\langle id, traj_i, d, lon, lat, t \rangle$ em que id corresponde ao número da viatura, $traj_i, d$ o identificador da trajetória, lon e lat a posição georreferenciada e t o tempo com data, são basicamente dados espaço-temporais trabalhados sobre uma base de dados de Objetos Móveis (*Moving Objects Database*, MOD), o HERMES, que é utilizado para dar apoio às consultas espaçotemporais. HERMES[11] é um MOD que integra funções que atendem serviços baseados em localização (Location-Based Services), construído por cima do banco geoespacial PostGIS, e que suporta aplicações dinâmicas de tempo real sobre esses objetos.

Com relação aos detalhes de implementação, o dashboard web desenvolvido utiliza a biblioteca Leaflet³ e o plugin Omnivore⁴ do OpenStreetMap, desenvolvida em Javascript, que facilita o processo de manipulação do mapa exibido. Ainda, as tecnologias utilizadas na implementação foram JavaScript e HTML, mas queremos migrar o sistema para JSF para uma facilidade futura na integração dos módulos já prontos que serão fornecidos pela polícia, assim muito código que foi feito em JavaScript será executado em Java puro.

Então, esse módulo de visualização oferece o suporte sobre esses dados dos GPS das viaturas ao passo que oferece ao gestor uma situação visual do patrulhamento pela cidade. A importância dele se desenvolve a partir de novas funcionalidades que podem ser acopladas. É pensado na implementação de mapas de calor sobre as trajetórias e as ocorrências, bem como futuramente um módulo de predição de

²OpenStreetMap, <http://www.openstreetmap.org/>

³Leaflet, an open-source JavaScript library for mobile-friendly interactive maps, <http://leafletjs.com/>

⁴Omnivore, universal format parser for Leaflet & Mapbox.js, <https://github.com/mapbox/leaflet-omnivore/>

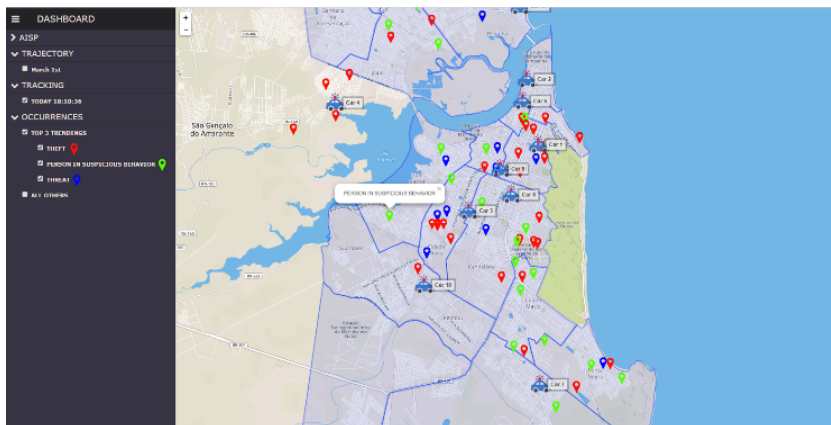


Figura 4: Tela do Módulo Dashboard Web

crimes a ser acoplado que trace uma rota de pontos base a serem visitados pela patrulha. Assim, regiões mais perigosas podem ser identificadas visualmente e atenuadas com um trabalho inteligente.

No contexto da literatura que foi estudada na linha de trabalho de técnicas de visualização, também temos a plataforma middleware FIWARE. Esta fornece uma série de serviços e tecnologias de análise e visualização de dados e funcionalidades de gerenciamento para dar suporte ao desenvolvimento de aplicações avançadas. Dentre esses serviços, temos o *Data/Context Management*, que visa facilitar o desenvolvimento e o provisionamento de aplicações inovadoras que requerem gestão, desempenho, processamento e exploração de dados em tempo real e em grande escala. A figura a seguir retrata a arquitetura FIWARE Data/Context Management, bem como os módulos que fazem parte dele.

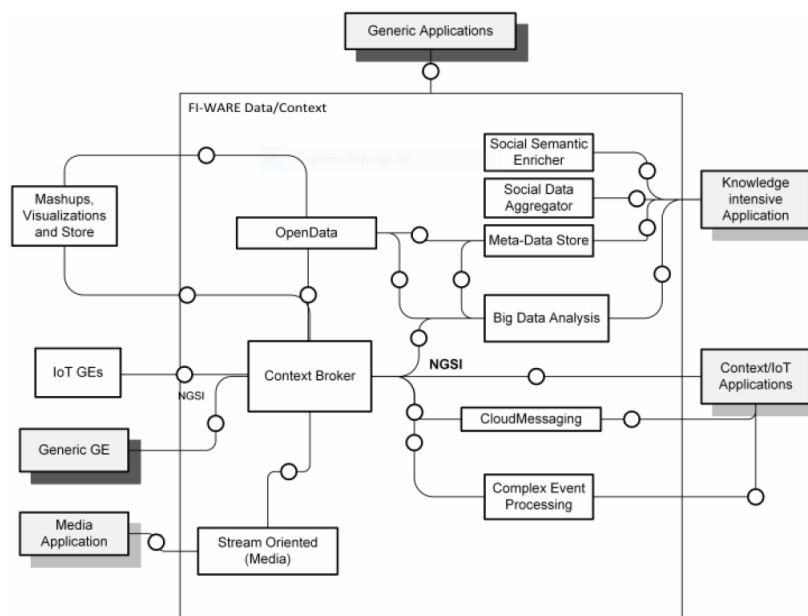


Figura 5: Arquitetura FIWARE Data/Context.

As principais funcionalidades do FIWARE Data/Context permitem a utilização do ecossistema hadoop, agilidade no compartilhamento, busca e utilização de dados, tratamento de recursos multimídia, entre outros.

3.3 Processamento de Sinais

No primeiro trimestre foi estudado como utilizar a técnica de identificação de sistemas no conceito de cidades inteligentes, visto que a aplicação para identificar padrões em dados aparentemente aleatórios foi um sucesso.

No próximo trimestre serão aplicadas as técnicas estudadas para dados reais como um forma de simular situações reais e, assim, implementar em aplicação real.

3.4 Processamento de Imagens

Ao longo deste primeiro trimestre, foi realizado pesquisas sobre processamento de imagens, dando foco a algumas técnicas de segmentação de imagens. Dessa maneira, a perspectiva para o próximo semestre é continuar essas pesquisas, como também a implementação de algumas destas técnicas, para comparação de seus resultados.

3.5 Redução de Dimensionalidade

Nas pesquisas e trabalhos desenvolvidos ao longo deste primeiro trimestre foi dado foco às diversas técnicas para a redução da dimensionalidade de dados e conceitos voltados ao *Information Theoretical Learning* (ITL), já que a ideia é utilizar ITL junto às técnicas de redução para obter resultados mais expressivos.

Dentre as técnicas para a redução de dimensionalidade estudadas estão algumas lineares, como Principal Component Analysis (PCA)[7] e o MDS, e as não-lineares, tais como o Isomap, *Stochastic Neighbor Embedding* (SNE) [6] e o t-SNE [13].

Nesse sentido, foi desenvolvido um estudo para comparar diversas técnicas de redução de dimensionalidade baseadas em Informação Mútua [10], bem como a criação de uma técnica de combinação desses métodos. Os resultados foram submetidos a *International Conference on Artificial Neural Networks* (ICANN) e ainda estamos aguardando o resultados da análise do artigo.

A perspectiva para o próximo trimestre é dar continuidade as pesquisas, além de iniciar o desenvolvimento de algumas das técnicas estudadas para a validação das mesmas e, se possível, melhorar o desempenho das mesmas usando ITL.

4 Considerações Finais

Durante o primeiro trimestre, o WP6 trabalhou basicamente no estudo da literatura de referência relacionada as cinco frentes de trabalho designadas para os discentes: Mineração de Textos, Visualização de Dados, Processamento de Imagens, Processamento de Sinais e Redução de Dimensionalidade.

O WP6 tem como objetivo principal atender as demandas de análise de dados geradas pelos demais WPs. Como todos estão em fase inicial, tais demandas ainda não estão ocorrendo. No entanto, o estudo e o desenvolvimento em paralelo de métodos que possam atender as principais aplicações relacionadas a Cidades Inteligentes deve ajudar, não só a atender futuras demandas, como a contribuir com o trabalho dos demais membros do projeto.

Vale ressaltar que o WP6 teve dificuldades em selecionar bolsistas em nível de mestrado. Os dois discentes em nível de mestrado que estão no projeto só iniciaram suas atividades com um mês de atraso e, por isso, não tiveram muito tempo para aprofundar suas pesquisas.

No próximo trimestre, todos os integrantes já estarão mais familiarizados com o contexto de cidades inteligentes, o que torna possível o desenvolvimento de produtos mais específicos para esse contexto. Ao mesmo tempo, as demandas dos demais WPs devem se tornar mais intensas, o que implica em uma comunicação mais eficiente entre os integrantes do projeto que para sejam atendidas dentro das metas esperadas.

Referências

- [1] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python*. O’Reilly Media, Inc., 1st edition, 2009.
- [2] A. CARVALHO. *INTELIGENCIA ARTIFICIAL: UMA ABORDAGEM DE APRENDIZADO DE MAQUINA*. LTC, 2012.
- [3] I. Esquef. Técnicas de entropia em processamento de imagens. dissertação de mestrado. Master’s thesis, Rio de Janeiro, Brasil, 2002.
- [4] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009.
- [6] G. E. Hinton and S. T. Roweis. Stochastic neighbor embedding. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 857–864. MIT Press, 2003.
- [7] I. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.
- [8] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014.
- [9] F. Mascarenhas, N.; Velasco. *Processamento Digital de imagens*. Kapelusz, Buenos Aires, Argentina, 1st edition, 1989.
- [10] X. V. Nguyen, J. Chan, S. Romano, and J. Bailey. Effective global approaches for mutual information based feature selection. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’14*, pages 512–521, New York, NY, USA, 2014. ACM.
- [11] N. Pelekis, E. Frentzos, N. Giatrakos, and Y. Theodoridis. HERMES: A trajectory DB engine for mobility-centric applications. *IJKBO*, 5(2):19–41, 2015.
- [12] M. J. Silva, P. Carvalho, and L. Sarmiento. Building a sentiment lexicon for social judgement mining. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language, PROPOR’12*, pages 218–228, Berlin, Heidelberg, 2012. Springer-Verlag.
- [13] L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, Nov. 2008.