

PoI: uma Aplicação de Detecção de Pontos de Interesse

Alice A. F. Menezes^{1,3}, Thais G. Almeida¹, Bernardo B. Gatto¹,
Eulanda M. dos Santos¹, Carlos Maurício S. Figueiredo², Eduardo F. Nakamura¹

¹Universidade Federal do Amazonas (UFAM) – Manaus, AM – Brasil

²Universidade do Estado do Amazonas (UEA) – Manaus, AM – Brasil

³Fundação Paulo Feitoza (FPF Tech) – Manaus, AM – Brasil

{alice.menezes, tga, bernardo, emsantos, nakamura}@icompu.fam.br,
cfigueiredo@uea.edu.br

Abstract. *With the emergence of social networks such as Twitter, Instagram and Foursquare, there was also the introduction of the concept of Participatory Sensor Networks (PSNs). Through this type of sensing, it is possible to collect general information of a particular location, using data provided by users and their mobile devices. Based on a data set of 333,286 check-ins, we present an application that uses Foursquare data to provide the Points of Interest (PoI) in Brazilian capitals, in other words, the most frequented venues, based on the number of check-ins. The aim is to provide the necessary information for tourism and advertising campaigns.*

Resumo. *Com o surgimento de redes sociais, como o Twitter, o Instagram e o Foursquare, houve também a introdução do conceito de Redes de Sensores Participativas (RSPs). Através deste tipo de sensoriamento, é possível coletar informações gerais de uma determinada localidade, utilizando dados disponibilizados por usuários e seus dispositivos móveis. Baseado em um conjunto de dados de 333.286 check-ins, apresentamos uma aplicação que, através de dados do Foursquare, fornece os Pontos de Interesse em capitais brasileiras, ou seja, os locais mais frequentados, baseado no número de check-ins do local. O objetivo é prover as informações necessárias para o turismo e para campanhas publicitárias.*

1. Introdução

Com a evolução dos celulares, hoje chamados de *smartphones*, tornou-se possível a detecção de dados relacionados aos usuários, como as suas opções e sua localização [Silva et al. 2014]. A introdução de sensores como GPS, acelerômetro, microfone, câmera e giroscópio nestes dispositivos [Lane et al. 2010], permitiu o sensoriamento de diversas áreas, em sua maioria urbanas. Esta rede de sensores, composta por todos os integrantes de uma sociedade que possuem dispositivos móveis, é denominada Rede de Sensores Participativa [Silva et al. 2013a].

Nas Redes de Sensores Participativas (RSPs), os usuários são entidades móveis autônomas capazes de sensoriar todo o ambiente no qual estão inseridos [Silva et al. 2014]. Assim, é possível adquirir informações relacionadas a uma determinada localidade através de dados de redes sociais e de dispositivos móveis utilizados por estes usuários.

A Figura 1 apresenta uma visão geral das RSPs, mostrando que os usuários agem como nós sensores, assim como as redes sociais agem como nós *sink* (sensores centralizadores de dados). Desta forma, a coleta de dados através de redes sociais permite identificar condições diversas como o trânsito, o clima, os eventos naturais, os aspectos econômicos, entre outros fatores.

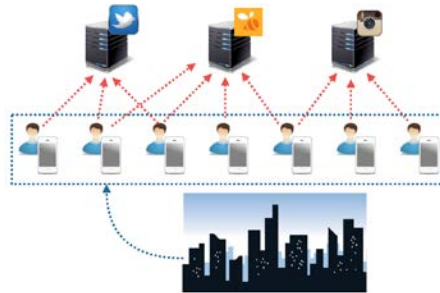


Figura 1. Visão geral das Redes de Sensores Participativas.

Outro aspecto importante das Redes de Sensores Participativas, é que os dados podem ser coletados para análise assim que o usuário os disponibiliza nas redes sociais. Isto permite a criação de soluções que provêm informações em tempo hábil para a tomada de decisões.

As principais contribuições deste trabalho são: um modelo de detecção de Pontos de Interesse, através dos dados da RSP do Foursquare, e um modelo de inferência de idade de usuários em redes sociais através de imagens. Vale ressaltar que o segundo modelo, é utilizado para prover informações auxiliares para o primeiro modelo.

Como resultado, apresentamos a aplicação PoI, cujo público-alvo é composto por turistas, empresários e órgãos governamentais que desejem identificar pontos turísticos e os locais mais frequentados por um público específico, através de dados como idade, sexo e categoria do local frequentado, tornando possível a realização de campanhas e divulgação de produtos.

Para a aplicação PoI, os Pontos de Interesse são os locais mais frequentados de uma região. Neste trabalho, fazemos a medição dos locais mais frequentados através da quantidade de *check-ins* do Foursquare que estes locais possuem em um determinado período, supondo que quanto maior é o número de *check-ins*, mais frequentado é o local.

As próximas seções deste trabalho estão divididas da seguinte forma: na seção 2 apresentamos os trabalhos relacionados, na seção 3 apresentamos o método proposto e a arquitetura da aplicação. Na seção 4, apresentamos a base de dados utilizada, os experimentos realizados e o resultados obtidos. Por fim, apresentamos as conclusões e as direções futuras para este trabalho.

2. Trabalhos Relacionados

Nesta seção é apresentado um resumo das soluções recentes para Pontos de Interesse e para soluções de aprendizagem de máquina para características visuais.

2.1. Pontos de Interesse

No trabalho proposto por [Lee and Sumiya 2010], os autores apresentam um modelo de pesquisa para detecção de eventos locais, utilizando experiências coletivas de uma

população através de dados compartilhados no Twitter.

O modelo proposto consiste em 3 etapas:

1. Coleta de dados geo-localizados através de um sistema de monitoramento de redes sociais para *microblogging*;
2. Identificação de regiões de interesse (RoIs) dos usuários do Twitter, através de medidas de regularidades do comportamento populacional;
3. Detecção de eventos geo-sociais, através da comparação entre medidas de regularidades.

Os autores utilizaram o algoritmo de aprendizagem de máquina K-means para agrupar os *tweets* geo-localizados obtidos através do sistema de monitoramento, e assim, identificar as RoIs. Vale a pena ressaltar que diferentemente deste trabalho, onde os autores definiram RoIs como sendo regiões compostas por aglomerações de usuários, o trabalho proposto neste artigo assume que Pontos de Interesse consistem em localidades (e.g., restaurantes, parques, teatros) onde um número expressivo de usuários realizou *check-in*.

Como experimento, os autores fizeram uma lista de 15 eventos festivos que iriam ocorrer no Japão, para verificar se o modelo proposto conseguiria determinar a existência desses eventos. Como resultado, foram coletados 9 eventos festivos (60% de acurácia). Porém, através do modelo, também foi possível detectar a ocorrência de eventos naturais que ocorreram no mesmo período.

Já no trabalho proposto por [Silva et al. 2013b], os autores utilizaram a RSP do Instagram, visando mostrar os desafios e as oportunidades que emergem do sensoriamento participativo realizados através dos usuários desta rede.

Para os experimentos, foram coletadas 2,3 milhões de fotos do Instagram utilizando o Twitter, o que mostra a abrangência da rede. Entre 30 de junho e 31 de julho de 2012, foram coletados 2.272.556 *tweets* contendo fotos georeferenciadas, postadas por 482.629 usuários. Cada *tweet* é composto de coordenadas (latitude e longitude) e o horário de compartilhamento da foto. Através dos dados coletados, foi possível construir uma aplicação que identifica os Regiões de Interesse em uma localidade.

O método utiliza um algoritmo de agrupamento para identificar Regiões de Interesse a partir das fotos do Instagram. Com isto, a popularidade de uma região baseia-se no número de diferentes usuários que compartilharam uma foto na mesma. Vale ressaltar que o nosso trabalho, diferentemente do trabalho descrito anteriormente, utiliza dados provenientes do Foursquare para identificar Pontos de Interesse em uma localidade, baseando-se apenas na quantidade de *check-ins* do local.

Em outro trabalho, a abordagem apresentada por [Frias-Martinez et al. 2012], utiliza dados geo-localizados que levam em consideração a privacidade do usuário. O método proposto, identifica automaticamente as atividades mais comuns entre os usuários de uma área específica, através da padronização dos dados provenientes do Twitter.

O método também identifica Pontos de Interesse automaticamente, sendo que estes possuem uma grande quantidade de *tweets* relacionados a eles. Para estudo de caso, foram utilizados dados da cidade de Nova Iorque em um período de 49 dias.

2.2. Características Visuais

Em [Jou et al. 2015], é proposto um modelo onde conjuntos de conceitos visuais (imagens onde o seu contexto é conhecido) são utilizados para enriquecer inferências envolvendo análise de polaridade. Mais detalhadamente, é argumentado que cada cultura e idioma são únicos. Logo, explorando a singularidade de cada cultura e idioma é possível entender os efeitos e relações existentes entre os sentimentos e as emoções semânticas através de informações contidas em multimídias sociais. Este trabalho explora os conceitos propostos por [Borth et al. 2013], onde estruturas semânticas chamadas pares adjetivo-substantivo são aplicados em um *framework* de análise de polaridade. Entretanto, em [Borth et al. 2013], esses conceitos são aplicados em um contexto multilíngue, onde apenas análise semântica é realizada. Em [Jou et al. 2015], o elemento visual é adicionado, e é formalizado um *framework* dependente de idioma para a descoberta automática de construções adjetivo-substantivo.

O processo de construção do *framework* multilíngue de sentimentos visuais proposto começa com a mineração de imagens e meta-dados com base em palavras-chaves. *Tag's* de imagens (T1, ..., T5) são rotuladas com partes do discurso (análise sintática, por exemplo. artigo, numeral, verbo) e adjetivos e substantivos são usados para formar combinações de pares de candidatos substantivo-adjetivo (PSA), enquanto outros são ignorados. Finalmente, os candidatos PSA's são filtrados com base em vários critérios (incluindo características semânticas de cada linguagem e eliminação de sentimentos neutros), que ajudam a remover pares incorretos, formando um *framework* final com a diversidade e abrangência necessárias para sua generalização.

Localização geográfica através de imagens é um problema de visão computacional onde, através de uma imagem (em geral, imagens de monumentos ou estátuas populares) deve-se inferir a localização geográfica de onde esta imagem foi obtida. Problemas de iluminação, mal posicionamento da câmera, oclusão entre prédios, árvores e outros monumentos, tornam esse problema extremamente desafiador. Além disso, entender quais características em uma imagem devem ser exploradas para uma maior representação (quais características são mais importantes naquele monumento) e discriminação (quais características diferenciam aquele monumento dos outros monumentos) são problemas ainda em aberto em visão computacional. O *framework* proposto por [Jin Kim et al. 2015] é uma solução para o problema de reconhecer um lugar representado por uma imagem de consulta usando um grande banco de dados de imagens rotuladas. Essa abordagem faz uso de extração de características visuais e predição através de um banco de classificadores baseados em *Support Vector Machines* (Máquina de Vetores Suporte).

Para incentivar a comunidade científica a buscar soluções para esse problema, é proposto um *benchmark*, denominado MediaEval Multimedia Benchmark [Larson et al. 2015]. Nesse *benchmark*, é apresentada uma tarefa que exige que os participantes desenvolvam algoritmos que predizem automaticamente a localização geográfica de multimídia sociais. No trabalho de [Ballan et al. 2010], os desafios de se explorar multimídias sociais para interferência de localização geográfica são abordados. Em muitos casos, multimídias sociais tais como fotos e vídeos, possuem meta-informações (títulos, descrições, comentários) que, quando analisadas de forma adequada, podem indicar de maneira muito precisa a localização geográfica da imagem ou vídeo.

As multimídias sociais também estão associadas com etiquetas de contribuições

(*likes*, *dislikes*) dos usuários e comentários que podem ser explorados através de análise de sentimento. Muitas vezes, os itens de multimídia enviados por usuários estão conectados em uma rede social, neste caso, informações como conexões sociais e pontos de vista (opiniões) também podem estar disponíveis. No caso de vídeos, o sinal multimídia envolve padrões temporais que podem ser exploradas através de algoritmos de visão computacional. Mídias que envolvem vídeos possuem, em geral, um canal visual e um canal de áudio. Informação de áudio possui conteúdo falado e/ou ambiental, onde o contexto possui grande correlação com a localização geográfica. No trabalho de [Choi et al. 2012] é proposta uma solução para dado um banco de dados de imagens/vídeos de treinamento com suas respectivas coordenadas geográficas e dados textuais, estimar a localização geográfica de um vídeo (contendo meta-informações, tais como mensagens e pontos de vista).

3. Método Proposto

Nesta seção é descrito o método proposto, que contém mineração de dados em redes sociais, extração de características visuais a partir da foto do perfil do usuário e uma aplicação *Web* para o processamento e exibição de informações, relacionadas aos Pontos de Interesse de uma localidade.

3.1. Mineração de Dados e Características Visuais

O método proposto neste trabalho, consiste na coleta de dados de usuários e *check-ins* do Foursquare, através do Twitter. Posteriormente, através da API do Foursquare¹, são extraídas informações dos locais onde os *check-ins* foram realizados, além de informações do usuário que realizou o *check-in*. Neste sentido, como os dados relacionados a idade do usuário não podem ser obtidos da rede social, foram coletadas imagens do perfil destes usuários, para inferir a idade dos mesmos através de técnicas de aprendizagem de máquina.

O Twitter² é uma rede social para *microblogging*, na qual os usuários podem enviar e receber atualizações através de mensagens de até 140 caracteres, chamadas de *tweets*. Esta rede social, lançada em Julho de 2006, possui cerca de 320 milhões de usuários ativos mensalmente³. Atualmente, o Twitter possui duas APIs de coleta de dados, sendo elas:

- **Streaming API:** coleta de dados em tempo real, sem limite de *tweets* na coleta;
- **Search API:** coleta de dados históricos de até uma semana, com um limite de 2.700 *tweets* a cada 15 minutos.

Estas APIs, utilizam *queries* de busca para a coleta de dados. Para a solução proposta, a *query* utilizada foi "swarmapp.com", pela qual é possível obter *links* de *check-ins* do Foursquare.

O Foursquare⁴ é uma rede social baseada em localização, cujo principal serviço é o *check-in*, um evento que é sensoriado quando a hora e o local de um usuário são

¹<https://developer.foursquare.com/>

²<https://www.twitter.com>

³<https://about.twitter.com/pt/company>

⁴<https://www.foursquare.com>

coletados [Silva et al. 2014]. Os locais cadastrados no Foursquare são associados a uma das 10 categorias atualmente existentes na rede social, das quais podem ser observados alguns exemplos na Tabela 1.

Tabela 1. Exemplos de locais existentes nas categorias do Foursquare.

Categoria do Foursquare	Exemplos de locais
<i>Arts and Entertainment</i>	Cinemas, museus e casinos.
<i>College and University</i>	Escolas, laboratórios universitários e centros de estudo.
<i>Food</i>	Restaurantes, cafeterias e padarias.
<i>Nightlife Spot</i>	Bares, clubes de <i>rock</i> e boates.
<i>Outdoors and Recreation</i>	Parques, academias e praias.
<i>Professional and Other Places</i>	Fábricas, auditórios e centros médicos.
<i>Shop and Service</i>	Lojas, salões de beleza e supermercados.
<i>Travel and Transport</i>	Aeroportos, estradas e hotéis.

Através do Foursquare, também é possível extrair informações de um usuário e associá-lo ao local no qual o *check-in* foi realizado. Para a solução proposta, extraímos o sexo e a foto de perfil do usuário, para a inferência da idade do mesmo. Neste caso, aplicamos técnicas de processamento de imagens e de visão computacional, para que os padrões existentes nas imagens pudessem ser extraídos e analisados.

Como as imagens podem encontrar-se parcialmente corrompidas por ruídos, oclusão, má qualidade de câmeras e iluminação não adequada, o processamento de imagens é aplicado para realçar características visuais desejadas, tais como contornos, texturas, formas ou cores. Esse processo de aperfeiçoamento produz imagens mais adequadas para análise. Por exemplo, é possível remover o ruído, nitidez e brilho de uma imagem, tornando possível a identificação das principais características.

3.2. Arquitetura da Solução

A arquitetura proposta neste trabalho é composta por um servidor *web*, uma aplicação *web* e um gerenciador de banco de dados, como ilustrado na Figura 2. Desta forma, as requisições dos usuários são enviadas ao servidor para que este se comunique com as camadas da aplicação. Logo em seguida, a aplicação solicita dados do gerenciador de banco de dados. Este, por sua vez, retorna os dados solicitados para a aplicação, a fim de que a mesma direcione o resultado para o servidor. Por fim, o servidor envia uma resposta para o *browser* dos usuários.

3.2.1. Servidor *web*

O servidor *web* utilizado foi o **Apache HTTP Server** devido ao fato de seu código ser aberto, apresentar uma ótima performance e confiabilidade, possuir uma vasta documentação e uma comunidade extremamente ativa.

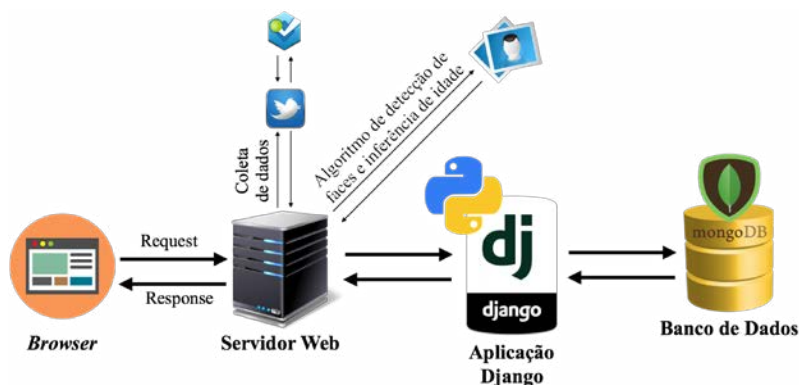


Figura 2. Visão geral da arquitetura da aplicação.

3.2.2. Aplicação web

Dentre as inúmeras opções de linguagens e ferramentas para desenvolvimento *web* no mercado, a aplicação proposta neste artigo utiliza um *web framework* de alto nível, escrito na linguagem de programação **Python**, conhecido como **Django**.

O *framework* Django proporciona um desenvolvimento *web* ágil, sem que os elementos de *design* e performance sejam negligenciados. Isto é possível, devido a inúmeras ferramentas e padrões de projetos que o compõe. As principais ferramentas do Django são [Santana and Galesi 2010]:

- **ORM - Object-Relational Mapper (Mapeador objeto-relacional)**: por meio desta ferramenta é possível abstrair a comunicação entre a camada de modelos de dados (classes de domínio) e o banco de dados. Ou seja, ela permite o mapeamento automático de modelos de dados escritos utilizando o paradigma Orientado a Objetos para o paradigma relacional;
- **URL Dispatcher (Despachador de URL)**: por meio desta ferramenta é possível gerenciar qual função será executada de acordo com a URL requisitada. Além disso, o URL Dispatcher provê mecanismos que facilitam a internacionalização do site e permite apresentar URLs simples e amigáveis ao usuário;
- **Sistema de templates**: fornece uma linguagem própria para geração de *templates*. No contexto de aplicações desenvolvidas em Django, um *template* consiste em uma *string* que, quando combinada com elementos de páginas *web* (*tags* HTML, por exemplo), gera uma saída consistente. Deste modo, é possível garantir que blocos de código e de representação não se misturem;
- **Sistema de administração**: consiste em uma interface de administração que possibilita gerenciar os objetos das classes de domínio do *site*. Desta forma, não é necessário implementar operações **CRUD** (*Create, Read, Update, Delete*) para cada uma das classes do domínio.

A aplicação *web* proposta neste trabalho apresenta como uma de suas principais características o fato de ser responsiva. Isto é, o *layout* da aplicação é capaz de se adaptar a diversos tipos de dispositivos, como pode ser observado na Figura 3. Para desenvolver o caráter responsivo, foram utilizadas, em conjunto com o Django, as seguintes tecnologias: **HTML5** e **Bootstrap**.

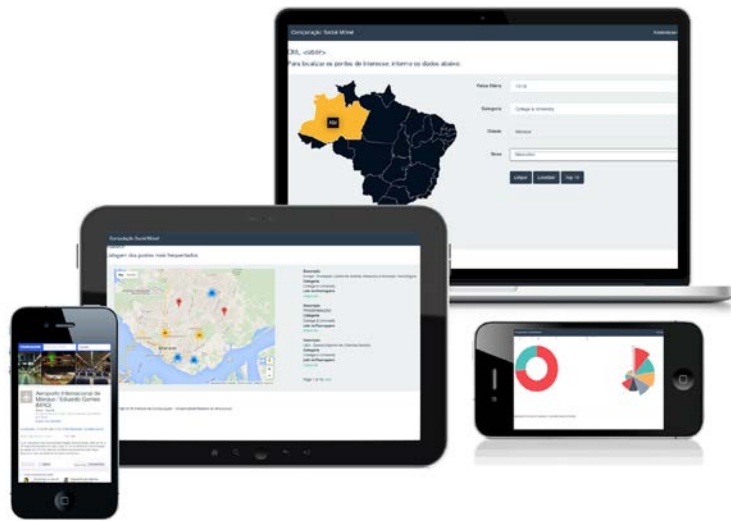


Figura 3. Dispositivos que podem utilizar a aplicação devido a sua responsividade.

O HTML (*Hypertext Language Markup*) é uma linguagem utilizada para publicação de conteúdo em navegadores *web*. Desde 2004, a sua quinta especificação (HTML5) vem sendo desenvolvida. Como novidades, o HTML5 apresenta: novas *tags* (delimitadores de conteúdo) que auxiliam no processo de divisão da página (cabeçalhos, seções, rodápes, etc) de forma intuitiva, um recurso chamado *canvas* que possibilita a renderização de desenhos em tempo de execução, uma API que permite a visualização e manipulação de mapas, entre outros recursos [Poyatos 2014]. Já o Bootstrap é um *framework front-end* amplamente utilizado pela indústria. Em sua estrutura interna, o *framework* possui arquivos CSS (*Cascading Style Sheets*) e Javascript que, quando combinados, permitem construir interfaces com poderosos recursos visuais [Rodrigues 2015].

3.2.3. MongoDB

Devido ao fato de bancos de dados relacionais apresentarem problemas de escalabilidade à medida que o volume de dados armazenados aumenta rapidamente, um novo padrão de banco de dados foi desenvolvido com o objetivo de representar dados sem utilizar inúmeras restrições entre eles [Simões 2014]. Este novo padrão é conhecido como **noSQL** (*Not Only SQL*) [Medeiros 2014].

O noSQL utiliza como modelo de dados o **Modelo Relacional Não Normalizado** (MRNN). Desta forma, os dados são persistidos em estruturas desnormalizadas, o que possibilita recuperar dados sem a utilização de operações de junções (JOINS), uma vez que as estruturas de armazenamento já se encontram aglutinadas. Dentre os principais sistemas gerenciadores de banco de dados que utilizam noSQL está o **MongoDB** [Membrey et al. 2011].

O MongoDB é um Sistema Gerenciador de Banco de Dados (SGBD) baseado em documentos, ou seja, os dados são armazenados em documentos ao invés de serem armazenados em tabelas. Estes documentos são estruturados utilizando uma forma binária do JSON (*Javascript Object Notation*) conhecida como BSON (*Binary Json*). Isto

ocorre principalmente porque o BSON suporta mais formatos de dados e a performance de indexação destes dados é muito maior [Foundation 2016, Banker 2011].

A aplicação proposta nesse artigo opera com um grande volume de dados extraídos de redes sociais e, por ser destinada a *web*, necessita de um alto grau de escalabilidade. Além disso, é necessário que as estruturas que armazenam os dados sejam extremamente flexíveis, como por exemplo, o tempo de processamento gasto para incluir mais um campo na estrutura deve ser mínimo. A partir das necessidades, foi escolhido o MongoDB como Sistema Gerenciador de Banco de Dados.

4. Experimentos e Resultados

Nesta seção, apresentamos os experimentos e resultados relacionados a detecção de pontos de interesse e a inferência de idade dos usuários, utilizada para auxiliar na aquisição de informações necessárias para a aplicação proposta.

4.1. Detecção de Pontos de Interesse

Os dados da aplicação são provenientes do Foursquare e foram coletados entre 14/12/2015 e 29/12/2015. A coleta dos dados foi realizada através do Twitter, utilizando a *query* de busca "swarmapp.com". Para este período foram coletados 333.286 *tweets* provenientes de capitais brasileiras, contendo URLs do Swarm⁵, aplicação responsável por registrar os *check-ins* do Foursquare.

Após a coleta dos dados, foi realizada a limpeza dos mesmos. Esta etapa se fez necessária devido ao fato de que alguns dados possuíam anomalias, o que invalidava a sua utilização pelos modelos presentes na solução deste trabalho. Dentre as anomalias, podemos destacar: *check-ins* cujas URLs foram coletadas, mas durante a extração de dados viu-se que os mesmos foram excluídos, e *check-ins* cujas contas dos usuários foram excluídas ou tornaram-se privadas.

Com a limpeza e a extração dos dados, verificamos que o Rio de Janeiro era a capital brasileira com a maior quantidade de *check-ins*, seguida por Porto Alegre e São Paulo (Figura 4). Além disso, verificamos que as 5 categorias do Foursquare com mais *check-ins* nesse período eram *Shop and Service*, *Food*, *Outdoors & Recreation*, *Professional & Other Places* e *Residence*, nesta ordem. Isto pode ser justificado pelo fato de que os dados são referentes a uma época festiva, o Natal.

Através dos *check-ins* também é possível recuperar informações a respeito do usuário que realizou o mesmo. Assim, verificamos que, neste período, cerca de 75% dos *check-ins* foram realizados por homens. Também coletamos as fotos dos perfis dos usuários a fim de que fossem utilizadas para a inferência da idade dos mesmos, conforme é descrito na subseção 4.2. Após a inferência de idade, tornou-se possível listar os locais mais frequentados de acordo com a faixa etária dos usuários.

Por fim, verificamos que cerca de 50% dos *check-ins* ocorreram no horário comercial (08:00 às 17:00), enquanto 42% ocorreram durante a noite e a madrugada.

Assumindo que os locais mais frequentados de uma região são os locais com o maior número de *check-ins*, identificamos através da aplicação PoI os locais mais frequentados nas capitais brasileiras e suas adjacências, de acordo com cada categoria do

⁵<https://www.swarmapp.com/>

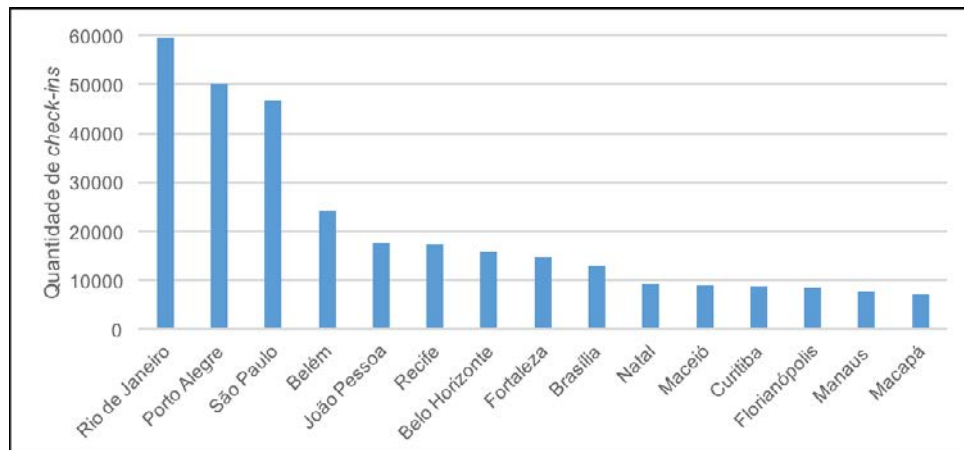


Figura 4. Quantidade de *check-ins* do Foursquare em 15 capitais brasileiras, extraídos no período de 14/12/2015 a 29/12/2015.

Foursquare. Como exemplo, apresentamos a Figura 5, na qual é possível identificar Pontos de Interesse em diferentes categorias na cidade de São Paulo e outras regiões próximas, nos dias 20/12/2015 e 21/12/2015. Para este exemplo, apresentamos a Tabela 2, na qual é possível verificar os 3 locais mais frequentados em 4 categorias, ou seja, os 3 principais Pontos de Interesse destas categorias.

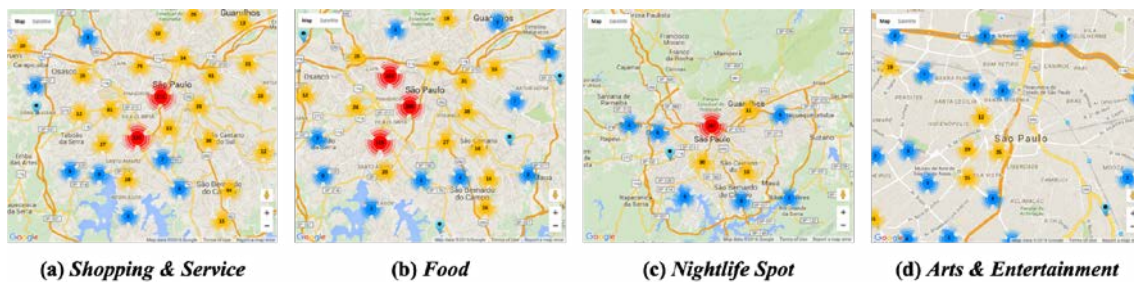


Figura 5. Pontos de Interesse em diferentes categorias na cidade de São Paulo e regiões próximas, nos dias 20/12/2015 e 21/12/2015.

Além das categorias do Foursquare, também é possível determinar Pontos de Interesse utilizando os dados dos usuários. Um exemplo disso, seria verificar os locais mais frequentados por homens ou mulheres em uma região.

Com os exemplos apresentados nesta subseção, verificamos que a aplicação PoI, através das redes sociais, pode prover informações necessárias para diversos fins, dentre os quais podemos destacar:

- Verificação de pontos de interesse de acordo com parâmetros previamente estabelecidos (categoria do local, idade e sexo do usuário, etc.). Isto pode auxiliar em campanhas publicitárias que queiram atingir um determinado público-alvo, por exemplo;
- Listagem dos locais mais frequentados em um região para o auxílio de turistas.

4.2. Inferência de Idade dos Usuários

Estimar a idade de pessoas através de imagens é um problema desafiador, pois envolve fatores que, em geral, geram estruturas não-lineares que são difíceis de serem modeladas

Tabela 2. Lista dos 3 locais mais frequentados nos dias 20/12/2015 e 21/12/2015, divididos por categoria referentes à cidade de São Paulo.

Categoria do Foursquare	Quantidade de <i>check-ins</i>	Lugares mais frequentados
<i>Shop & Service</i>	1331	Shopping Metrô Boulevard Tatuapé, Linda Unha, Havaianas.
<i>Food</i>	1249	Cris Doces e Salgados, Central de Massas, Lanches Escadão.
<i>Nightlife Spot</i>	386	Ilhabela Irish Pub, Barcearia, Coconut Brasil.
<i>Arts & Entertainment</i>	363	Estúdio Iquiririm, Cinemark, Cinépolis.

e classificadas por algoritmos de aprendizagem de máquina modernos. Na estimativa de idade através de imagens, utilizam-se técnicas de extração de características que são necessárias para discriminar e representar o processo de envelhecimento e aparência das faces em imagens.

A tarefa de estimativa de idade por imagens herda as características dos problemas de reconhecimento de objetos através de imagens, tais como o processo de detecção da face, a localização das características faciais, a formulação vetor de características e classificação. Dependendo da aplicação, o resultado da classificação pode ser uma idade exata ou uma faixa etária.

Outro fator importante, que pertence ao problema estimativa da idade, é o intervalo de idades consideradas. Este parâmetro, é um aspecto importante do problema, com diferentes características de envelhecimento que aparecem em diferentes grupos etários. Portanto, um modelo treinado para lidar com uma determinada faixa etária pode não ser aplicável a outras faixas etárias. Em redes sociais, essas imagens encontram-se em ambiente ainda mais desafiador, onde oclusão, mal posicionamento da câmera e iluminação podem tornar a extração de características das imagens difícil.

Para resolver os desafios listados acima, desenvolvemos um método baseado em classificação de conjuntos de imagens. Nesse método, a estimativa de idade de cada pessoa é obtida por meio de várias instâncias de padrões de um mesmo indivíduo, ao invés de apenas uma instância.

Com o rápido desenvolvimento de tecnologias de processamento e armazenamento de vídeos, sequências de imagens de um mesmo indivíduo são facilmente obtidas em redes sociais. Conseqüentemente, a tarefa de estimativa de idade pode ser aprimorada se for utilizado um conjunto de imagens para a extração de características. Por exemplo, cada amostra é um conjunto de imagens faciais com diferentes poses, iluminações e expressões. Utilizando um conjunto de imagens, a estimativa de idade pode alcançar maior precisão, pois conjuntos de imagens geralmente contêm mais informações do que uma única imagem. Ainda que esse conjunto de imagens obtido das redes sociais de um mesmo indivíduo tenha baixa qualidade, pode-se explorar a relação temporal entre essas

imagens, aprimorando a precisão do classificador.

Considerando que um conjunto de imagens de um mesmo indivíduo foi coletada, têm-se os seguintes passos para estimar a idade deste; 1) Detecção da face do indivíduo na imagem, 2) Extração de características da face, tais como região dos olhos, nariz e lábios, 3) Seleção de características, 4) Classificação do conjunto de imagens.

Para detecção de faces, utilizamos o método desenvolvido por [Viola and Jones 2004], que extrai características baseados no Haar-like e utiliza o classificador Ada-boost em cascata. Para a extração de características utilizou-se os seguintes métodos: Histograma de Gradientes Orientados (HOG), LBP e Gabor. Estes métodos possuem características complementares, tais como textura, cor e forma. A seleção de características é feita por meio da Análise dos Componentes Principais [Jolliffe 2002] e um classificador baseado nos vizinhos mais próximos [Cunningham and Delany 2007], que determina a faixa etária a qual pertence o conjunto de imagens dado.

O principal problema em se criar uma base de dados para a tarefa de estimar a idade através de imagens, é a ausência do ground-truth (informação sobre a idade real das pessoas nas imagens). No entanto, é possível inferir visualmente um rótulo para cada imagem manualmente. Para tanto, é necessário que um grupo de pessoas determinem a idade aproximada de cada imagem da base de dados. Nos experimentos apresentados a seguir, fizemos uso de 10 pessoas que avaliaram cada uma das imagens da base de dados. O valor final de cada rótulo foi obtido por meio da moda do conjunto de avaliações de cada imagem (Tabela 3).

Tabela 3. Base de dados obtida através de imagens de redes sociais.

Faixa etária	Moda	Variância	Homens	Mulheres	Total
12-17	14	2.9	8	11	19
18-23	20	2.7	11	29	40
24-28	25	1.8	9	12	21
29-35	32	3.1	7	10	17
36-41	38	2.5	7	6	13
42+	43	3.4	3	7	10

A base de dados confeccionada, é composta por 120 conjuntos de fotos de perfis de usuários. Cada conjunto possui entre 7 e 12 imagens, que foram extraídas de redes sociais, tais como o Foursquare. As imagens variam em relação ao fundo (*background*), resolução, orientação e iluminação. Para efeitos de simplificação, consideramos que todas as imagens em um mesmo conjunto são pertencentes a um único indivíduo, e que foram obtidas em um intervalo de tempo razoavelmente próximo. Quando há mais do que uma face na imagem, supomos que a maior imagem pertence ao proprietário do perfil.

O objetivo desse experimento, é determinar se é possível utilizar algoritmos de aprendizagem supervisionada para estimar a idade das pessoas em redes sociais. A fim de

normalizar as imagens, utilizamos o algoritmo proposto por [Viola and Jones 2004] para retirar apenas as imagens das faces de cada foto, sendo que todas as imagens de faces foram redimensionadas para 40x40 pixels. Em seguida utilizamos os seguintes descritores de características: Histograma de Gradientes Orientados (HOG), LBP e Gabor. A dimensão de cada vetor de características foi reduzida, utilizando-se a Análise dos Componentes Principais. Os parâmetros empregados encontram-se na Tabela 4.

Tabela 4. Principais parâmetros utilizados no experimento.

	Dimensão do vetor D1	Dimensão do vetor D2
HOG (Vh)	900	37
LBP (Vl)	1024	43
Gabor (Vg)	1092	51

Para a execução dos testes, a Análise dos Componentes Principais foi fixada para extrair 95% da energia acumulada em cada um dos vetores de características. Para classificação, fez-se uso do vetor concatenado $V_f = V_h || V_l || V_g$, onde $A = B || C$ significa que A é formado pela concatenação de B com C. O esquema de validação empregado, é baseado no *leave-one-out*, onde um conjunto é separado para teste e o restante é utilizado para treino.

A matriz de confusão apresentada na Tabela 5, contém as informações sobre a taxa de classificação dos conjuntos em relação a quantidade de classes. A acurácia total do método pode ser definida então por $79.05\% ((73.68\% + 87.5\% + 85.71\% + 70.58\% + 76.92\% + 80\%)/6)$. Levando em consideração que as imagens utilizadas para formar a base de dados foram obtidas em ambiente não controlado, o classificador possui uma boa acurácia. Além disso, a base de dados empregada no experimento não está balanceada em relação ao número de conjuntos por classe, por esse motivo tem-se a tendência de se classificar um conjunto na classe que possui o maior número de conjuntos.

Tabela 5. Matriz de confusão.

		Classes de Teste					
		12-17	18-23	24-28	29-35	36-41	42+
Classes de Treino	12-17	73.68	-	-	-	7.69	-
	18-23	15.78	87.5	14.28	17.64	7.69	10
	24-28	5.26	7.5	85.71	5.88	7.69	10
	29-35	5.26	2.5	-	70.58	-	-
	36-41	-	-	-	-	76.92	-
	42+	-	2.5	-	5.88	-	80

5. Conclusão e Trabalhos Futuros

Neste trabalho, apresentamos uma aplicação *web*, cujo objetivo é auxiliar empresários, turistas e órgãos governamentais na análise de um perímetro urbano. Desta forma, os mesmos podem identificar locais propícios para o turismo, a realização de campanhas e propagandas, e a divulgação de produtos.

Como exemplo, podemos citar as campanhas de saúde, que buscam atingir um público-alvo específico de acordo com os locais frequentados e a faixa etária. Com a aplicação PoI, o tipo de local pode ser identificado através das categorias do Foursquare, e a idade pode ser inferida pela foto dos perfis dos usuários que frequentam os Pontos de Interesse.

Dentre os desafios deste trabalho, podemos destacar a proveniência de dados de redes sociais, pois algumas cidades estudadas possuíam menos de 1.000 *check-ins* no período de uma semana. Além disso, as imagens obtidas para a inferência de idade foram obtidas de um ambiente não controlado.

Como trabalho futuro, planejamos aprimorar o método de inferência da idade dos usuários, por meio de redes sociais. Para isto, pretendemos não só utilizar apenas as imagens, mas também outros dados que podem ser coletados de perfis, como por exemplo, o sexo, a forma de escrita (culto ou coloquial), a rotina, os locais mais frequentados e os círculos de amizade.

6. Referências

- Ballan, L., Bertini, M., Del Bimbo, A., Meoni, M., and Serra, G. (2010). Tag suggestion and localization in user-generated videos based on social knowledge. In *Proceedings of second ACM SIGMM workshop on Social media*, pages 3–8. ACM.
- Banker, K. (2011). *MongoDB in action*. Manning Publications Co.
- Borth, D., Ji, R., Chen, T., Breuel, T., and Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM.
- Choi, J., Friedland, G., Ekambaram, V., and Ramchandran, K. (2012). Multimodal location estimation of consumer media: Dealing with sparse training data. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 43–48. IEEE.
- Cunningham, P. and Delany, S. J. (2007). k-nearest neighbour classifiers. *Multiple Classifier Systems*, pages 1–17.
- Frias-Martinez, V., Soto, V., Hohwald, H., and Frias-Martinez, E. (2012). Characterizing urban landscapes using geolocated tweets. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 239–248. IEEE.
- Foundation, D. S. (2016). Django documentation. url = <https://docs.djangoproject.com/en/1.9/>. [Acessado em Janeiro/2016].
- Jin Kim, H., Dunn, E., and Frahm, J.-M. (2015). Predicting good features for image geo-localization using per-bundle vlad. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1170–1178.

- Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.
- Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., and Chang, S.-F. (2015). Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 159–168. ACM.
- Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. (2010). A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150.
- Larson, M., Kelm, P., Rae, A., Hauff, C., Thomee, B., Trevisiol, M., Choi, J., Van Laere, O., Schockaert, S., Jones, G. J., et al. (2015). The benchmark as a research catalyst: Charting the progress of geo-prediction for social multimedia. In *Multimodal Location Estimation of Videos and Images*, pages 5–40. Springer.
- Lee, R. and Sumiya, K. (2010). Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks*, pages 1–10. ACM.
- Medeiros, H. (2014). Introdução ao mongodb. url = <http://www.devmedia.com.br/introducao-ao-mongodb/30792/#ixzz3yRHHTBfC>. [Acessado em Janeiro/2016].
- Membrey, P., Plugge, E., and Hawkins, D. (2011). *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing*. Apress.
- Poyatos, H. (2014). Programando em html5. url = <http://www.devmedia.com.br/programando-em-html5/31040/#ixzz3ySFtOR3r>. [Acessado em Janeiro/2016].
- Rodrigues, J. (2015). Como criar uma aplicação responsiva com bootstrap. url = <http://www.devmedia.com.br/como-criar-uma-aplicacao-responsiva-com-bootstrap/33185>. [Acessado em Janeiro/2016].
- Santana, O. and Galesi, T. (2010). Python e django: desenvolvimento ágil de aplicações web. *São Paulo: Novatec*.
- Silva, T. H., de Melo, P. O. V., Viana, A. C., Almeida, J. M., Salles, J., and Loureiro, A. A. (2013a). Traffic condition is more than colored lines on a map: Characterization of waze alerts. In *Social Informatics*, pages 309–318. Springer.
- Silva, T. H., Vaz de Melo, P., Almeida, J. M., and Loureiro, A. A. (2013b). Uma fotografia do instagram: Caracterização e aplicação. *XXXII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*.
- Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., and Loureiro, A. A. (2014). Revealing the city that we cannot see. *ACM Transactions on Internet Technology (TOIT)*, 14(4):26.
- Simões, A. (2014). Comparando o nosql ao modelo relacional. url = <http://www.devmedia.com.br/comparando-o-nosql-ao-modelo-relacional/30917/#ixzz3yRM5amNk>. [Acessado em Janeiro/2016].
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.