



## Mobile Intelligence - Um arcabouço para análise e visualização de dados tempo-espaciais originados por dispositivos móveis\*

Mobile Intelligence - A framework for analysis and visualization of time-spatial data sourced for mobile devices

Pasteur Ottoni de Miranda Junior<sup>1</sup>  
João Francisco de Abreu<sup>2</sup>

### Resumo

Dados tempo-espaciais são usualmente provenientes de dispositivos móveis equipados com sistemas de posicionamento global (GPS) e são gerados por carros, caminhões, ônibus, trens, aviões, navios, telefones celulares, dentre outros. Esses dados são úteis porque permitem localização e rastreamento. No entanto, uma grande quantidade de conhecimento pode ser extraída por meio de uma análise mais profunda deles, uma vez que tal conhecimento não é facilmente obtido por simples inspeção, pois esses bancos de dados são enormes. Este artigo apresenta um arcabouço para extração, processamento, análise e visualização de dados tempo-espaciais, de modo que se possa deles extrair inteligência. Este arcabouço é composto de ferramentas para extração e processamento de dados tempo-espaciais, uma infra-estrutura de bancos de dados multidimensionais usados em data warehouses com extensões espaciais que analisa dados em muitas dimensões e níveis de agregação, uma ferramenta para a execução de consultas sobre paradas e movimentos de dispositivos móveis e um conjunto de algoritmos de mineração de dados espaciais, para extrair padrões de trajetória, classificação, regras de associação e clustering de trajetórias.

**Palavras-chave:** Análise tempo-espacial. Mineração de dados. Data warehouse. Dispositivos móveis.

---

\* Artigo apresentado à revista Abakós.

<sup>1</sup> Departamento de Ciência da Computação - Pontifícia Universidade Católica de Minas Gerais,

<sup>2</sup> Pós-Graduação em Geografia - Tratamento da Informação Espacial - Pontifícia Universidade Católica de Minas Gerais,

### **Abstract**

Time-spatial data usually comes from mobile devices equipped with Global Positioning Systems (GPS) and are generated by cars, trucks, buses, trains, planes, ships, mobile phones, among others. These data are useful because they allow locating and tracking. However, a great amount of knowledge can be extracted through a deeper analysis of them, since that knowledge is not easily obtained by simple inspection, as these databases are huge. This paper presents a framework for extraction, processing, analysis and visualization of time-space data, so that it will be possible to extract intelligence from them. This framework consists of tools for extraction and processing of time-space data, an infrastructure of multidimensional databases used in data ware house with spatial extensions that analyses data on many dimensions and levels of aggregation, a tool for executing queries based on stops and moves of mobile devices and a set of algorithms for spatial data mining, to extract trajectory patterns, classification, association rules and clustering trajectories.

**Keywords:** Time-space analysis. Data mining. Data warehouse. Mobile devices

## 1 INTRODUÇÃO

Dados tempo-espaciais são aqueles representados por um conjunto de pontos localizados no tempo e no espaço. Esses dados são provenientes de dispositivos móveis, equipados ou não com sistemas GPS (*Global Positioning System*) e, atualmente, são gerados em grande volume. Tais dispositivos encontram-se instalados em automóveis de passeio, caminhões, ônibus, trens, aviões, navios, telefones celulares, animais, acoplados a sensores e sistemas de medição, dentre outros. Esses dados em si já são bastante úteis, na medida em que permitem localizar e rastrear. Entretanto, uma grande quantidade de conhecimento pode ser extraída por meio de uma análise mais profunda dos mesmos, conhecimento esse não facilmente obtido por simples inspeção, já que ocorrem em grande quantidade. Portanto, essa descoberta de conhecimento tem como objetivo a extração de conhecimento implícito, novo e útil de um banco de dados, especificamente, de um banco de dados que contenha dados tempo-espaciais provenientes de dispositivos móveis.

Uma trajetória é um conjunto de pontos referenciados no tempo e no espaço, gerados por dispositivos móveis (um veículo com um GPS, por exemplo) em sequência específica. A análise de trajetórias permite inferir uma série de informações, dentre as quais podemos citar:

- Padrões de movimento de pessoas em cidades em determinado período e horário;
- Identificação de movimentos típicos, que possam denotar padrões de comportamento de motoristas;
- Gerenciamento de frotas de veículos, através da identificação de tempos de entrega, tempos de percorrimto de determinados trechos, localização de um veículo em relação aos demais;
- Identificação de áreas de congestionamento;
- Identificação de tráfego de veículos em áreas de maior probabilidade de ocorrência de acidentes ou de roubos;
- Associar hábitos de movimentação de motoristas a hábitos de consumo;
- Análise de proximidade de veículos a pontos de interesse (postos de combustível, hospitais, lojas, etc);
- Identificação de perfis de risco a partir da determinação do comportamento tempo-espacial de motoristas.

Uma vez que dados tempo-espaciais são gerados em grande quantidade e são provenientes de múltiplas fontes, o que se observa, entretanto, é que analisá-los em sua forma bruta, consome tempo e muitas vezes inviabiliza tomadas rápidas de decisão. Portanto, é necessário que tais dados estejam estruturados de maneira a tornar tal análise mais eficiente. Para tal, os

denominados *data warehouses* surgem como uma solução bastante interessante, por trabalharem com dados em uma forma não normalizada – os chamados modelos multidimensionais – tornam consultas e análises muito mais rápidas. Uma vez tratados, dados tempo-espaciais podem ser analisados de maneira mais eficiente, por meio de:

- Técnicas de mineração de dados, que permitem a extração de conhecimento implícito na base de dados não obtido por técnicas convencionais de pesquisa e busca.
- Técnicas de análise espacial, que usam a informação espacial contida nos dados para extrair conhecimento.
- Técnicas de OLAP (*On Line Analytical Processing*), que permitam a visualização dos dados sob várias perspectivas.

## 2 DADOS TEMPO-ESPACIAIS - DEFINIÇÕES

### 2.1 Padrões de trajetória (t-patterns)

Padrão de trajetória é um conjunto de trajetórias que compartilham a propriedade de visitar a mesma sequência de lugares em tempos de deslocamento similares. Portanto, padrões de trajetória são descrições concisas de comportamentos frequentes (ABREU; MIRANDA-JR., 2007). Diante disso, dois conceitos são fundamentais: o de regiões de interesse e o de tempo de viagem típico de objetos movendo-se entre tais regiões.

### 2.2 Regiões de Interesse

Regiões de interesse são regiões conhecidas a priori, por onde dispositivos móveis trafegam. Podem ser, por exemplo, parques, logradouros, praças, bairros, edificações, dentre outros. Essas regiões de interesse introduzem conteúdo semântico a informações de tempo e espaço provenientes de dispositivos móveis.

### 2.3 Sequências temporalmente anotadas (TAS)

Sequências temporalmente anotadas (ABREU; MIRANDA-JR., 2007) (TAS) são sequências que incluem informação sobre tempos de transição entre suas regiões de interesse:

$$T = s_0 \xrightarrow{\alpha_1} s_1 \xrightarrow{\alpha_2} \dots \xrightarrow{\alpha_n} s_n$$

Também é representada como a dupla  $T = (S, A)$  de uma sequência  $S = \langle s_0, \dots, s_n \rangle$  com anotações temporais  $A = \langle \alpha_1 \dots, \alpha_n \rangle$ .

## 2.4 Sequências Espaço-Temporais (Sequências ST)

Uma sequência ST ou trajetória é uma sequência de triplas  $S \langle (x_0, y_0, t_0), \dots, (x_k, y_k, t_k) \rangle$  onde  $t_i (i = 0..k)$  são tempos,  $\forall_{0 \leq k, t_i < t_j < t_{i+1}}$  e  $(x_i, y_i)$  são coordenadas.

## 3 ESTRUTURA GERAL DO ARCABOUÇO

O arcabouço proposto está esquematizado na Figura 1. Ele é constituído pelos seguintes elementos básicos:

- Funções para visualização e tratamento das informações armazenadas nas bases de dados: funções para consultas *Online Analytical Processing* (OLAP) à base multidimensional, funções para mineração e análise espacial de dados e funções para realização de *queries* eficientes sobre a base de paradas e movimentos.
- Bases de dados: uma base primitiva, constituída por dados provenientes de dispositivos móveis, e duas bases dela derivadas, uma multidimensional e outra para armazenamento de paradas e movimentos de dispositivos móveis dentro de e entre regiões de interesse;
- Extração de dados: rotinas que permitem, periodicamente, a extração de informações da base de dados primitiva para as bases de dados derivadas.

As seções subsequentes apresentam as principais funções supracitadas.

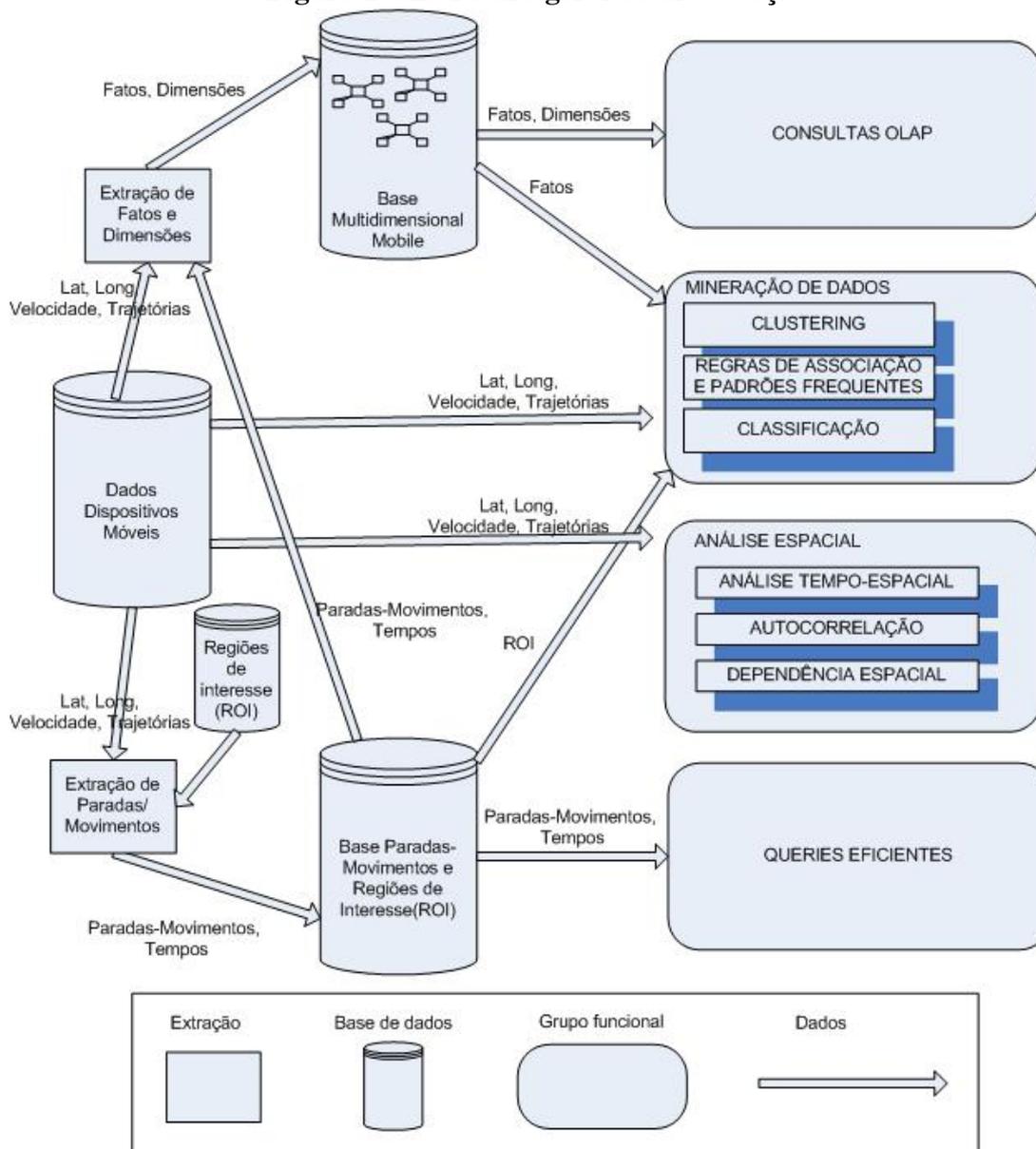
## 4 MINERAÇÃO DE DADOS TEMPO-ESPACIAIS

### 4.1 Regras de Associação

Algoritmos para obtenção de regras de associação extraem regras de comportamento ao longo do tempo. Tais regras são denominadas regras de associação, pois descrevem relacionamentos lógicos do tipo causa-e-efeito entre itens em um conjunto de dados.

Os algoritmos mais utilizados para obtenção de regras de associação são o Apriori e o AprioriTID. Bogorny (2006), Agrawal e Srikant (1994), partem desses algoritmos para propor uma abordagem para obtenção de regras de associação em dados espaciais. Uma regra de associação espacial é uma regra na forma  $X \rightarrow Y$  onde X e Y são conjuntos de predicados e alguns destes são espaciais [Koperski, 1995]. Nesse caso, cada linha é uma instância de uma classe e as

**Figura 1 – Estrutura geral do arcabouço.**



colunas são predicados (BOGORNÝ, 2006). Cada predicado pode estar associado a um atributo não espacial ou a uma relação espacial. Predicados espaciais representam relações espaciais entre entidades geográficas. Em mineração de regras de associação espaciais, interessam-nos as relações possíveis, ou seja, relacionamentos proibidos ou obrigatórios são descartados. O algoritmo proposto por Bogorný (2006) elimina dependências entre objetos geográficos e gera conjuntos frequentes de regras de associação de maneira significativamente mais eficiente que os convencionais.

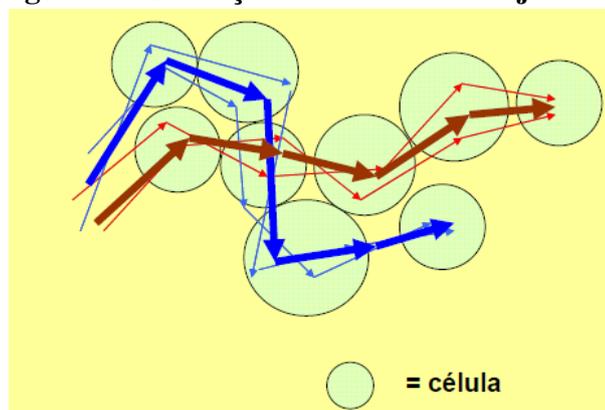
A título de exemplo, seja a seguinte regra de associação espacial:  $trajetoria(X) \wedge trajetoria(Y) \wedge intersecciona(X, Y) \rightarrow cruzamentoderisco(X, Y)$ . Se essa relação de causa-e-efeito for muito frequente, o local onde existe o cruzamento pode estar especialmente sujeito a colisões.

## 4.2 Clustering

Em essência, o processo de obtenção de clusters de trajetórias consiste em se agruparem trajetórias tempo-espaciais similares, ou seja, que a cada instante de tempo estejam aproximadamente no mesmo local, de forma a se determinarem padrões de comportamento similares (Figura 2).

Algoritmos tradicionais de *clustering*, como *k-means* (WITTEN; FRANK, 2005), PAM (KAUFMAN; ROUSSEEUW, 1990) ou CLARA (NG; HAN, 2002), devem ser adaptados de forma a incluírem a informação espacial. Uma possível adaptação consiste em se considerar a avaliação da dissimilaridade como sendo constituída por duas componentes ponderadas, uma para o espaço de atributos (por exemplo, velocidade média, tempo total, número de paradas) e a outra para a distância geográfica, considerada, para o caso de distância entre trajetórias, como a soma da distância entre objetos, em cada instante.

**Figura 2 – Obtenção de clusters de trajetórias.**



Fonte: (BOGORNÝ, 2006).

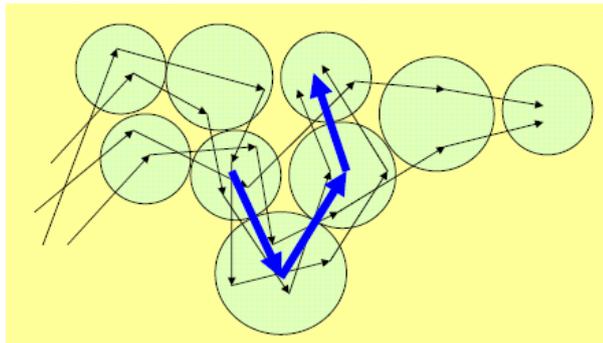
## 4.3 Padrões Sequenciais

Esta categoria de algoritmos identifica padrões mais frequentes de caminhos percorridos ou tendências (Figura 3). Esses padrões podem ser agrupados (*clustering*) ou classificados. Dessa forma, a partir dos mesmos podem-se realizar previsões, como as de densidade de dispositivos, alcance de lugares, trajetórias futuras e eventos (GEOPKDD, 2006).

## 4.4 Classificação

Algoritmos incluídos nesta categoria extraem regras de comportamento ao longo do tempo (Figura 4), de forma que um comportamento futuro possa ser inferido (BOGORNÝ, 2006). Nesta categoria, incluem-se aplicações de previsão e análise de tráfego de veículos. Foi desenvolvida, com base no arcabouço, uma ferramenta que utiliza dados de velocidade média, fluxo

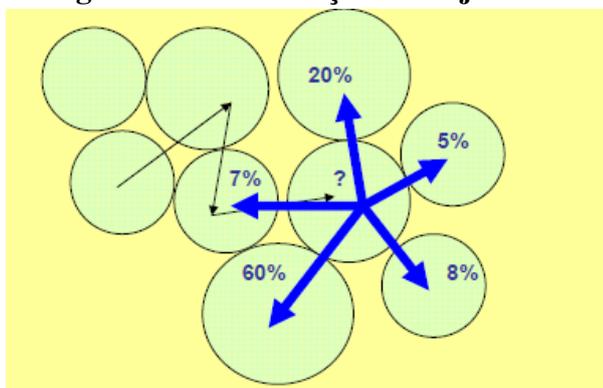
**Figura 3 – Obtenção de padrões de trajetórias - Fonte: Bogorny, 2006**



Fonte: (BOGORNY, 2006).

e densidade média de veículos em regiões de interesse para inferir a qualidade do tráfego. Essa ferramenta utiliza Redes Neurais Artificiais para realizar a classificação do tráfego em três categorias (ruim, intermediário e bom), com base nas três variáveis supra citadas.

**Figura 4 – Classificação de trajetórias.**



Fonte: (BOGORNY, 2006).

## 5 ANÁLISE ESPACIAL DE DADOS TEMPO-ESPACIAIS

### 5.1 Análise Tempo-Espacial

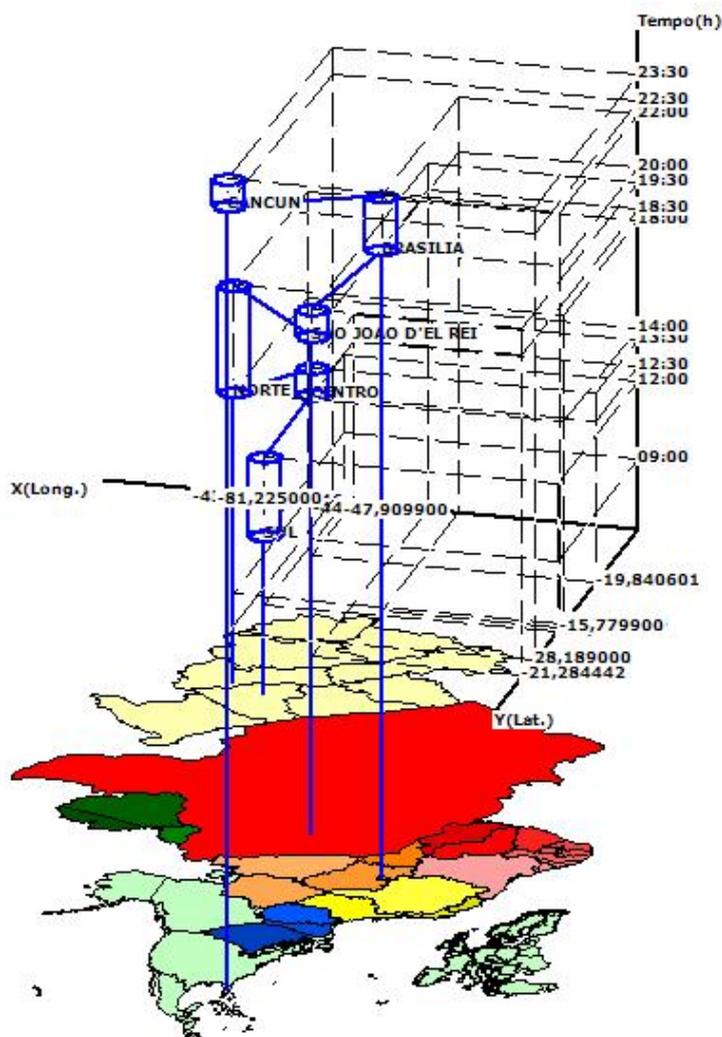
O geógrafo Törsten Hägerstrand, na década de 1960, notabilizou-se por integrar o tempo ao comportamento espacial, através de um diagrama tempo-espacial. Abreu e Miranda-Jr. (2007) definem o diagrama tempo-espacial multi-escopo, que representa as transições de um indivíduo em termos das dimensões tempo e espaço em até quatro escopos de interação, cada um deles a uma escala espacial distinta: escopo local (próximo), escopo estadual, escopo nacional e escopo internacional.

Três eixos de coordenadas são utilizados para representar a informação, um para tempo e os outros dois para latitude e longitude. Cada escopo é representado em uma camada diferente do diagrama, de forma que todas as transições podem ser simultaneamente visualizadas. A Figura 5 apresenta o aspecto geral do diagrama. Os cilindros em azul representam os tempos

gastos em cada atividade, ou seja, a base do cilindro corresponde ao início da atividade e o topo ao fim (que podem ser lidos no eixo vertical). As linhas que partem da base dos cilindros para os mapas nas camadas indicam o local exato onde ocorreu a transação. As linhas azuis conectando os cilindros representam a sequência em que as atividades ocorreram. Neste exemplo, as camadas representam os mapas de Belo Horizonte, do Estado de Minas Gerais, do Brasil e das Américas do Norte/Central e Europa, correspondentes ao escopo das transações que o indivíduo realizou.

É evidente a utilidade desse diagrama na representação de dados tempo-espaciais provenientes de trajetórias geradas por dispositivos móveis. A visualização pode ser realizada em diversos níveis, simultaneamente, aumentando significativamente a capacidade de análise. Além disso, podem ser graficamente visualizadas as paradas e os movimentos, por meio dos cilindros verticais.

**Figura 5 – Aspecto Geral do Diagrama Tempo-Espacial Multi-Escopo.**



Fonte: (ABREU; MIRANDA-JR., 2007)

## 5.2 Dependência espacial

Proposta originalmente por Tobler (1970), a dependência espacial enuncia que “tudo se relaciona com tudo, mas coisas próximas são mais relacionadas que coisas distantes”. Ou seja, a dependência espacial entre informações torna-se mais fraca na medida em que os dados estão mais distantes. Em essência, formaliza um conceito do senso comum: o de que valores semelhantes de atributos, geograficamente próximos uns dos outros, são um forte indicador de que eles estão relacionados entre si e, principalmente, de que provavelmente existe uma relação de causa-e-efeito entre eles e os eventos, fatos ou fenômenos que os produzem.

A dependência espacial calca-se fundamentalmente no conceito de distância. A definição matemática de distância entre dois pontos é formalizada pelo conceito de espaço métrico, ou seja, um espaço onde há uma distância métrica, definida da seguinte forma:

Dados  $(E; d)$  onde  $d$  é uma distância métrica em um espaço métrico  $E$ ,  $d$  é definida como a função  $d : E \times E \rightarrow R^+$  que associa dois elementos  $(a, b)$  de  $E$  a um valor real de tal forma que:

$$\begin{aligned}d(a, b) &\geq 0 \\d(a, b) &= 0 \Leftrightarrow a = b \\d(a, b) &= d(b, a) \quad (\textit{simetria}) \\d(a, b) &\leq d(a, b) + d(b, c) \quad (\textit{desigualdade triangular})\end{aligned}\tag{1}$$

A generalização da expressão da distância entre dois pontos em um espaço métrico é dada pela distância de Minkowski, definida da seguinte forma:

Dados dois pontos  $P = (x_1, x_2, \dots, x_n)$  e  $Q = (y_1, y_2, \dots, y_n)$  tem-se

$$d(P, Q) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

Alguns casos notáveis dessa distância: *Distância euclidiana* ( $p = 2$ ):

$$d(P, Q) = \left( \sum_{i=1}^n |x_i - y_i|^2 \right)^{1/2}$$

*Distância táxi* ( $p = 1$ ), utilizada para obtenção de distâncias em reticulados:

$$d(P, Q) = \left( \sum_{i=1}^n |x_i - y_i| \right)$$

O denominado *modelo gravitacional*, baseado na expressão da força gravitacional de Newton, é uma aplicação imediata da dependência espacial. Nesse modelo, a força gravitacional é substituída pelo conceito de *grau de interação* entre dois pontos. Quando utilizado em análise espacial de dispositivos móveis, as massas podem ser substituídas pela velocidade, densidade

ou fluxo médio de dispositivos em uma região de interesse. Assim pode-se, por exemplo, inferir o grau de interação entre regiões de interesse, útil na análise de tráfego de veículos que fluem de uma região para outra.

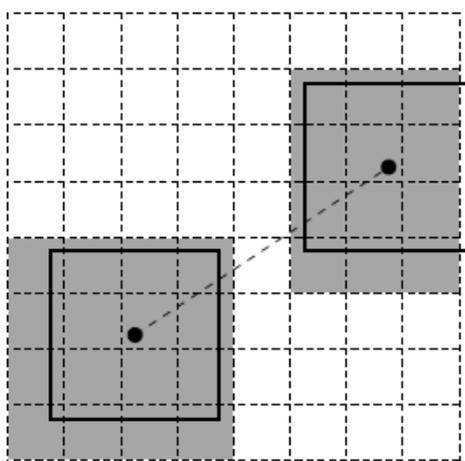
### 5.2.1 *Obtenção de Regiões de Interesse*

Quando regiões de interesse não são conhecidas a priori, algumas heurísticas para identificá-las automaticamente podem ser utilizadas. Essas heurísticas, que se baseiam principalmente no conceito de dependência espacial, são os seguintes (GEOPKDD, 2006):

- Selecionar sobre uma base de lugares candidatos (por exemplo, restaurantes, museus, lojas, etc) um subconjunto que satisfaça a determinado critério (por exemplo, todos os restaurantes e lojas próximos a uma estrada);
- Automaticamente computar lugares candidatos através da análise de trajetórias, por exemplo, selecionar todas as regiões quadradas que foram visitadas por ao menos 10% dos objetos;
- Mesclar as duas abordagens anteriores, por exemplo, selecionar todas as estradas onde mais de 50% das trajetórias que as cruzam mudam sua direção.

Denominam-se pontos populares, àqueles visitados por muitos indivíduos. Para computar pontos populares eficientemente, pode-se discretizar o espaço de trabalho em uma grade regular com células de tamanho pequeno. Então, a densidade das células é computada tomando-se cada trajetória e incrementando a densidade de todas as células que contêm quaisquer de seus pontos, ou seja, todas as células que interceptam a vizinhança de qualquer ponto da trajetória. A Figura 6 exibe um exemplo de uma trajetória de dois pontos em uma vizinhança quadrada e uma grade com tamanho das células igual a 1/3 da vizinhança e onde as células com maior densidade são coloridas de cinza.

**Figura 6 – Densidade de células.**



Fonte: (GEOPKDD, 2006).

### 5.3 Autocorrelação

A autocorrelação expressa quanto o valor observado de um atributo numa região é dependente dos valores deste mesmo atributo nas localizações vizinhas. O índice de Moran I é o mais utilizado para estimar a força da correlação entre valores observados em função da distância que os separam. Este índice é dado pela expressão:

$$I = \frac{\sum_{t=1}^n \sum_{f=1}^n w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{t=1}^n (Z_i - \bar{Z})^2}$$

onde

$Z_i$  = o valor normalizado do atributo considerado na área  $i$

$\bar{Z}$  = é o valor médio do atributo na região sob estudo

$n$  = número de áreas

$w_{ij}$  = peso da área  $i$  em relação à área  $j$ . Se forem vizinhas, seu valor pode ser 1, caso contrário, 0. Este peso pode ser também a distância entre as áreas.

Valores positivos (entre 0 e +1) do índice de Moran indicam uma correlação direta, valores negativos (entre -1 e 0) indicam correlação inversa.

Outro importante indicador de autocorrelação é o LISA (*Local Indicator of Spatial Association*) (ANSELIN, 1995), que é o equivalente local do índice de Moran (a soma de todos os índices locais de uma região é proporcional ao índice de Moran). Ele é obtido pela seguinte expressão:

$$I_i = Z_i \sum_{j=1}^n W_{ij}$$

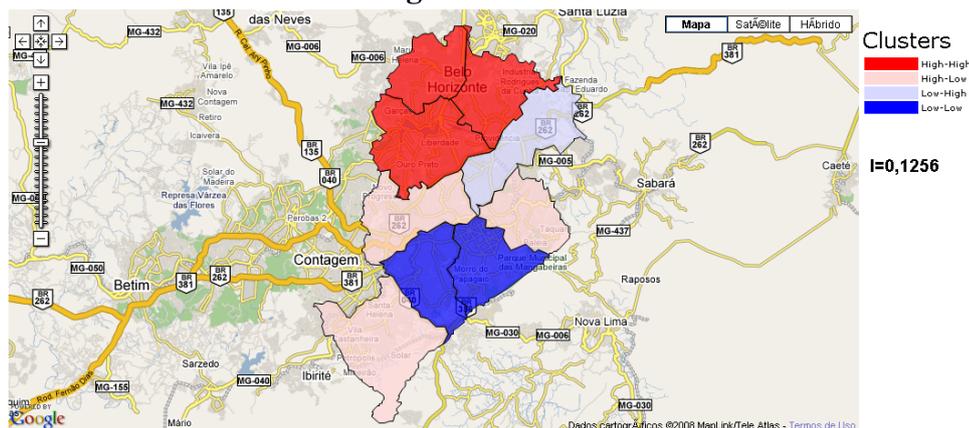
onde  $I_i$  é o índice LISA de correlação espacial local para cada área  $i$  na região considerada.

Para cada área, os índices LISA permitem avaliar sua similaridade em relação às áreas vizinhas e testar sua significância. Desta análise, quatro cenários podem emergir:

- Áreas com valores altos e vizinhos similares: *high-high*. São denominadas *hot spots*.
- Áreas com valores baixos e vizinhos similares: *low-low*. São denominadas *cold spots*.
- Áreas com valores altos e vizinhos com valores baixos: *high-low*. São denominadas *potential spatial outliers*.
- Áreas com valores baixos e vizinhos com valores altos: *low-high*. São também denominadas *potential spatial outliers*.
- Áreas com autocorrelação não significativa.

A análise de autocorrelação de dispositivos móveis pode ser realizada sobre dados de velocidade, fluxo e densidades médias dos mesmos em regiões vizinhas. A Figura 7 exibe um mapa de cenários de índices LISA obtidos sobre a velocidade média de veículos em regionais de Belo Horizonte.

**Figura 7 – Mapa de cenários de indicadores LISA para velocidade média de veículos nas regionais de Belo Horizonte.**

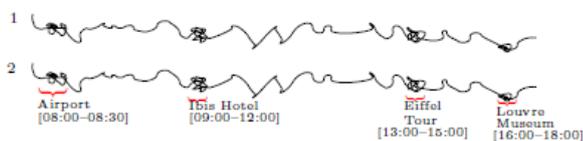


## 6 QUERIES EFICIENTES

### 6.1 Paradas e movimentos

Trajetórias brutas contêm pouco conteúdo semântico. Assim sendo, é essencial que se integrem aos dados tempo-espaciais de trajetórias brutas, informações geográficas que possam acrescentar conteúdo semântico às mesmas. Na Figura 8, a uma trajetória bruta, associam-se pontos geográficos de interesse. A essa composição denomina-se trajetória semântica. Armazenados tais pontos em banco de dados, podem-se realizar pesquisas (*queries*) bastante interessantes. A Figura 9 resume o arcabouço proposto por Alvares et al. (2007) e utilizado no presente trabalho, em que dados brutos de trajetórias e dados geográficos são associados (pré-processados) segundo um algoritmo específico (ver item 7.2 abaixo) e armazenados como trajetórias semânticas, sobre as quais podem-se realizar *queries*.

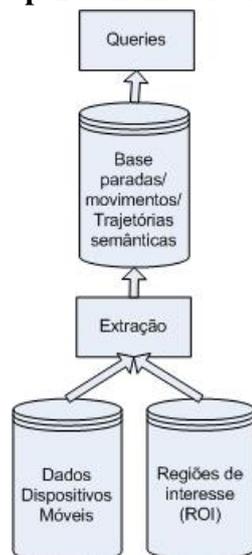
**Figura 8 – 1 representa um conjunto de pontos de uma trajetória. 2 representa a mesma trajetória a mesma trajetória com conteúdo semântico acrescentado.**



Fonte: (ALVARES et al., 2007)

O modelo de pré-processamento de trajetórias proposto por Alvares et al. (2007) associa dados geográficos com dados espaço-temporais de trajetórias, por meio de identificação de paradas (*stops*) e movimentos (*moves*). Em uma trajetória bruta, podem existir paradas candidatas que são objetos geográficos de interesse (que podem ser de qualquer geometria) por onde ela passa. Dispositivos móveis, nesse modelo, têm paradas por determinado tempo em determinadas paradas candidatas quando permanecem nas mesmas por um tempo mínimo pré-determinado. Movimentos são sub-trajetórias que ocorrem entre duas paradas consecutivas. Na Figura 10, temos três paradas candidatas (regiões RC1, RC2 e RC3). Se o dispositivo móvel

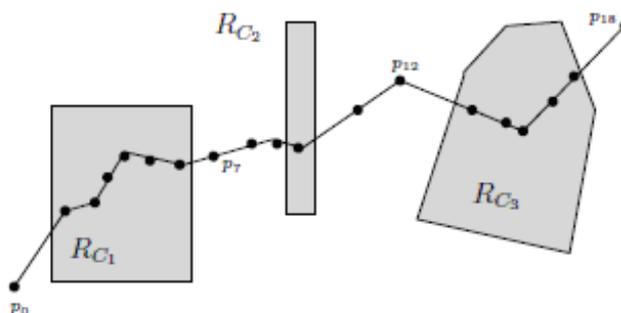
**Figura 9 – Arcabouço para análise de trajetórias semânticas.**



Fonte: (ALVARES et al., 2007).

permanecer dentro do objeto geográfico RC1 por mais que um tempo mínimo especificado, então estará caracterizada a primeira parada para os pontos 1 a 6 da trajetória em RC1. A trajetória entra em RC2, mas se supusermos que permanece lá por um período igual ou menor que um tempo inferior ao especificado, então RC2 não será uma parada. Da mesma forma, se em RC3 a trajetória permanecer por mais tempo que um mínimo especificado, então os pontos de 13 a 17 da mesma em RC3 constituirão outra parada. Por conseguinte, os movimentos serão três: do ponto inicial até RC1, da saída de RC1 até RC3, (pontos 7, 8 e 9, 10, 11 e 12) e da saída de RC3 até o final.

**Figura 10 – Exemplo com três paradas candidatas.**



Fonte: (ALVARES et al., 2007).

## 6.2 Algoritmo para extração de paradas e movimentos

Alvares e outros 2007, apresentam um algoritmo para extração de paradas e movimentos que verifica se cada ponto de uma trajetória intercepta a geometria de uma parada candidata. Se intercepta, verifica se a duração da interseção é maior que um tempo mínimo. Se for, é considerada uma parada, senão é descartada. Movimentos são então registrados entre duas

paradas sucessivas.

### 6.3 Realização de queries mais eficientes (ALVARES et al., 2007)

A forma de armazenamento por paradas e movimentos facilita a realização de queries. Poderíamos fazer, por exemplo, a seguinte pesquisa: quantas trajetórias vão de um hotel a pelo menos um ponto turístico? A seguinte *query* fornece a resposta:

```
SELECT distinct count (a.Tid)
FROM stop a, stop b
WHERE a.SFTname = 'Hotel' AND
      b.SFTname = 'Touristic_Place' AND a.Tid = b.Tid
AND a.Sid < b.Sid
```

Esta QUERY é mais simples que a convencional, que necessita de dois *joins* espaciais:

```
SELECT distinct count (t.Tid)
FROM trajectory t, trajectory u,
      hotel h, touriticPlace p
WHERE intersects (t.geometry, h.geometry) AND
      intersects (t.geometry, h.geometry) AND
      t.Tid=u.Tid AND u.timest>t.timest
```

Outro exemplo que ilustra a simplificação obtida: quantas trajetórias visitam a Igreja de Notre Dame, o Pompidou e a Ponte Arcole. Abaixo, respectivamente a *as queries* convencional e a obtida com o uso de paradas e movimentos (ALVARES et al., 2007).

```
SELECT distinct count (t.Tid)
FROM trajectory t, trajectory u, trajectory v,
      touriticPlace p, touriticPlace q, bridge b
WHERE intersects (t.geometry, p.geometry) AND
      p.Name='Notre_Dame' AND
      intersects (u.geometry, q.geometry) AND
      q.Name='Pompidou_Center' AND
      intersects (v.geometry, b.geometry) AND
      b.Name='Tid' AND t.Tid=v.Tid AND
      t.Tid=u.Tid AND u.Tid=v.Tid AND
      t.timest < v.timest AND v.timest < v.timest
```

```
SELECT distinct count (a.Tid)
FROM move m, touriticPlace p,
      touriticPlace q, bridge b
WHERE m.Slid=p.Id AND p.Name='Notre_Dame' AND
      m.Slid=q.Id AND q.Name='Pompidou_Center' AND
```

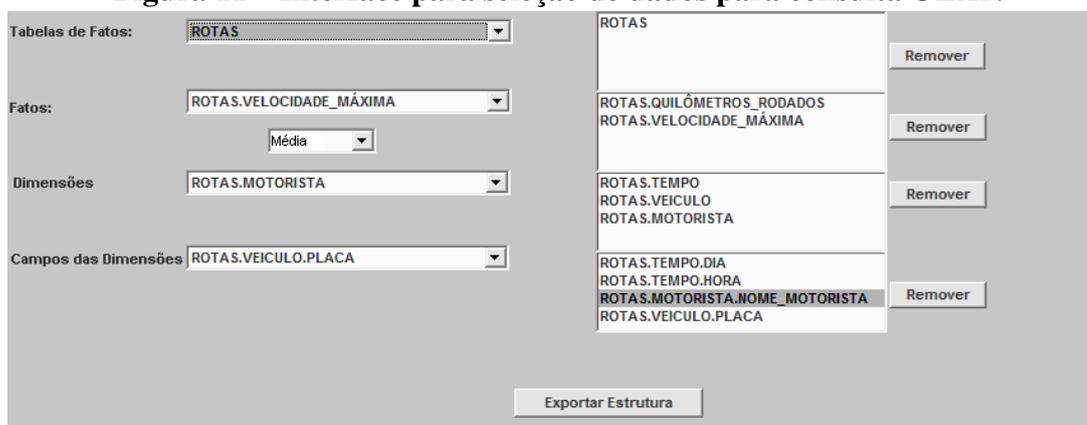
```
intersects (m.geometry , b.geometry) AND  
b.Name= 'Arcole '
```

## 7 CONSULTAS OLAP

Uma vez estruturados em bases de dados multidimensionais, dados de dispositivos móveis podem ser analisados e visualizados por meio de uma ferramenta *Online Analytical Processing* (OLAP). O A Seção 8.2 apresenta os modelos multidimensionais definidos para o arcabouço. A estrutura básica do OLAP é o denominado cubo multidimensional, constituído por diversas dimensões em função das quais fatos são exibidos. Por exemplo, para a tabela de fatos DW\_FAT\_DISP\_REGIAO (Figura 17) pode-se definir um cubo formado pelo fato Velocidade\_Media, o qual pode ser visualizado em um mês específico, por categoria e por região de interesse. Esse cubo pode ser modificado para visualização de Velocidade\_Media em outro nível de agregação, por exemplo, em um dia específico e por município.

A Figura 11 exibe a interface da ferramenta OLAP para geração de consultas em bases multidimensionais. A geração das consultas é feita pela seleção de fatos, dimensões e campos das dimensões.

**Figura 11 – Interface para seleção de dados para consulta OLAP.**



A ferramenta permite também a realização de *queries* para filtragens, seleções e operações espaciais sobre os fatos e dimensões selecionados (Figura 12), segundo a seguinte sintaxe que inclui, inclusive, operações espaciais:

Query= [ ( ) Cláusula [ ], and, or, ( ) [query].

Cláusula = [ ( ) [Campo Dimensão | Fato|Função] [operador relacional | operador espacial] [valor numérico| Função| Campo Dimensão | Fato]

[ Operador relacional ] = [ > | < | > = | < = | < > ]

[ Operador espacial ] = [ contém | disjunto | igual a | intersecciona | sobrepõe | toca | está dentro de | distância | cruza ]

Função= [média | máximo | mínimo | soma | contagem | área | centróide | intersecção | união ] ( [ Parâmetros ] )

Parâmetros = Parâmetro [,Parâmetro]

Parâmetro = [Fato | Campo Dimensão | valor numérico]

Cada linha da grade contém uma cláusula que compõe a *query* e é preenchida através de seleções nos *combo boxes*.

**Figura 12 – Interface para definição de filtros, seleção de dados por meio de operadores espaciais e convencionais e funções**

	Fato/Dim	Oper. Rel.	Valor	Oper. Lóg.
(	ROTAS.QUILÔMETROS_RODADOS	>=	220	E
	ROTAS.VELOCIDADE_MÁXIMA	<=	110	
)				

A Figura 13 exibe a interface para definição de agregações (agrupamentos) de dimensões e a ordenação do relatório final.

**Figura 13 – Definição de campos de agrupamento e ordenação**

Agrupar por: ROTAS.QUILÔMETROS\_RODADO [Remover]

Ordenar por: ROTAS.MOTORISTA.NOME\_MOT [Remover] Ordem crescente

As Figuras 14, 15 e 16 a seguir exibem exemplos de relatórios, gráficos e mapas exibidos pela ferramenta.

## 8 BASES DE DADOS DO ARCABOUÇO

### 8.1 Base de dados de dispositivos móveis

Contêm dados tempo-espaciais enviados por dispositivos móveis. Incluem dados de posição (latitude e longitude), hora de envio, velocidade, altitude, orientação e eventos, como saídas de regiões específicas, violação de limites de velocidade, abertura de compartimentos, dentre outros.

Figura 14 – Relatório de Consulta Gerado.



CUBO TEMPO MOTORISTA VEICULO  
QUILÔMETROS\_RODADOS>=220 E VELOCIDADE\_MÁXIMA<=110

29/06/2009 14:39:22

DIA	HORA	NOME_MOTORISTA	PLACA	QUILÔMETROS_RODADOS	VELOCIDADE_MÁXIMA
2008-01-11	15:34:56	José da Silva	wer4567	220	50
2008-01-11	15:34:56	José da Silva	wer4567	230	60
2008-01-09	19:22:00	José da Silva	wer4567	220	50
2008-01-10	15:45:34	José da Silva	wer4567	230	60
2008-01-09	19:22:00	Paulo Matos	ASD4467	220	50
2008-01-10	15:45:34	Silvio Mauricio Tales	QWE4567	230	60

Figura 15 – Gráfico de Barras



## 8.2 Base de dados multidimensional

Extraída a partir da base de dados de dispositivos móveis, é constituída por duas tabelas de fatos: uma que agrega e totaliza fatos para regiões e uma segunda que agrega e totaliza fatos por dispositivos. Os fatos extraídos são velocidade, densidade e fluxo de veículos médio, máximo e mínimo, tempo de permanência dentro de regiões de interesse, tempo médio, máximo e mínimo parado, dentre outros. Ambas as tabelas de fatos possuem dimensões espaciais que vão do menor nível de granularização, *região de interesse*, até o *nível de estado*. Possuem também uma dimensão temporal que permite a agregação de informações do nível *hora* de granularização até o nível *ano*. A Figura 17 exhibe as duas tabelas de fatos, bem como as dimensões a elas relacionadas. A Seção 7 apresenta uma ferramenta OLAP para visualização desses modelos.

## 8.3 Base de dados de paradas e movimentos

Essa base de dados armazena paradas e movimentos entre regiões de interesse, extraídas da base de dispositivos móveis. A Figura 18 exhibe o modelo de dados dessa base.

Figura 16 – Mapa gerado - velocidade média por regional

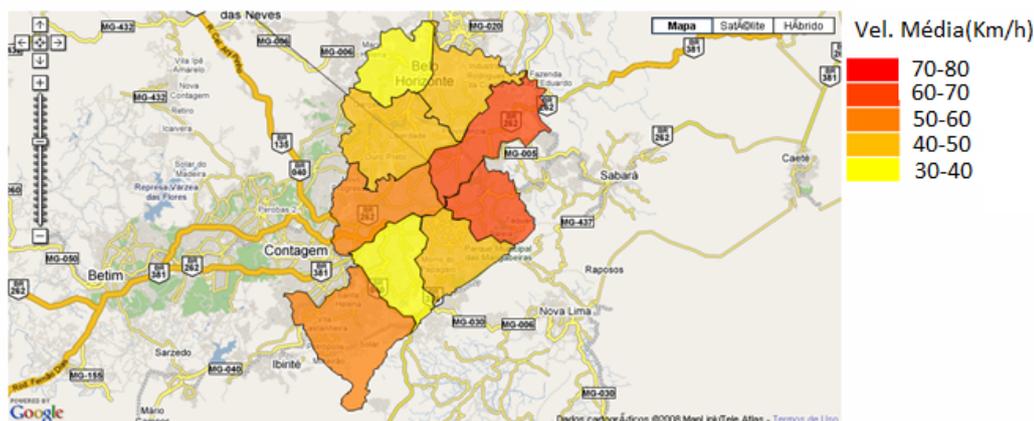
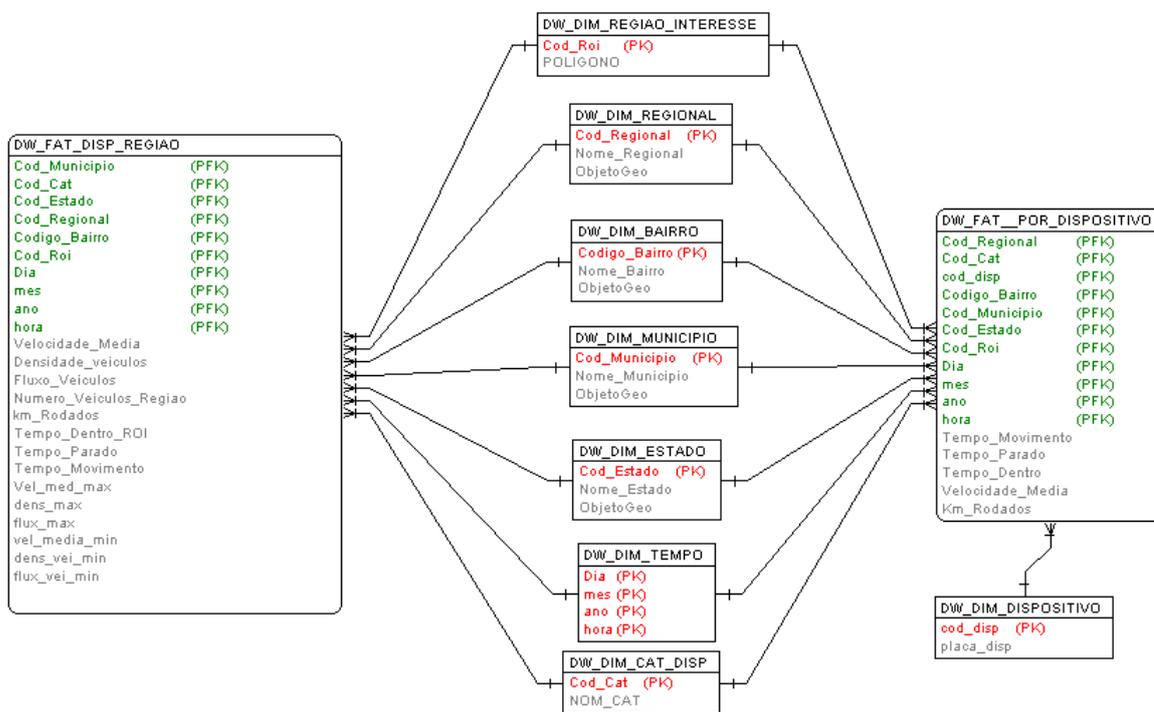


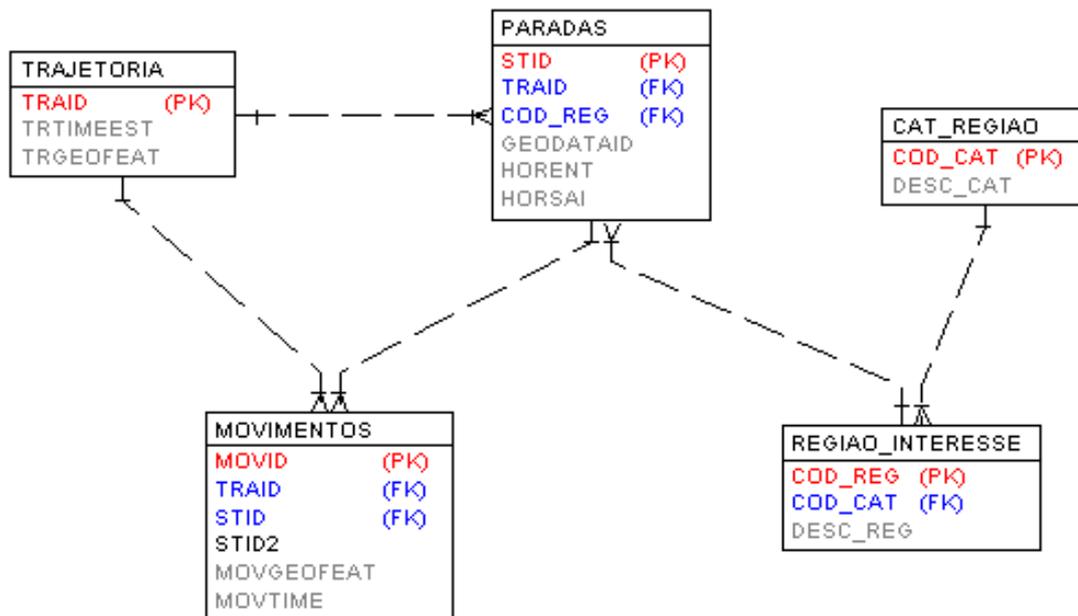
Figura 17 – Modelos multidimensionais estrela.



## 9 EXTRAÇÃO DE DADOS

Fatos e dimensões são extraídos diretamente da base de dados de dispositivos móveis, por meio da realização de operações de agregação e totalização, dos níveis de latitude-longitude e hora-minuto-segundo, para níveis mínimos de agregação *região de interesse e hora*. São realizadas operações de limpeza, como remoção de caracteres espúrios, uniformização de formatos, conversões de unidade e é feita uma verificação de consistência e uniformidade de agregações. Para geração das duas tabelas de fatos, realizam-se operações de “desnormalização”, que consistem em juntar informações de diversas tabelas em uma só, criando redundâncias no modelo relacional, de forma a “desnormalizá-lo”. Esses dados devem ser extraídos com frequência diária.

Figura 18 – Modelo de dados da base de paradas e movimentos



A extração de dados para a tabela de paradas e movimentos é feita diretamente da base de dispositivos móveis, por meio do algoritmo descrito na Seção 6.2 e de uma base de dados espacial de regiões de interesse.

## 10 CONCLUSÕES

A utilização das informações geradas a partir da análise de dados tempo-espaciais pode trazer contribuições e impactos notáveis em diversas áreas, quais sejam:

- *Segurança pública.* Identificação de áreas de maior risco de acidentes, de maior risco de roubos e assaltos, dentre outras.
- *Econômica.* Prospecção de novas possibilidades de mercado e o aperfeiçoamento de estratégias de marketing, o que poderá contribuir com a melhoria da situação econômica das empresas;
- *Controle de tráfego.* O conhecimento extraído de condições do trânsito possibilitará análises que permitam buscar alternativas para melhoria do tráfego em cidades.
- *Atuária.* O conhecimento extraído permitirá inferir o perfil de motorista, possibilitando a criação de modalidades de seguros menos onerosas e que reflitam mais precisamente o perfil do condutor.

- *Transportes.* A identificação do comportamento de condução de motoristas de ônibus e caminhões, bem como de possíveis áreas de gargalo, contribui na definição de melhores rotas tanto de transporte de mercadorias como de transporte coletivo, tornando mais eficiente a logística de empresas de transporte e contribuindo também com a melhoria do tráfego em cidades.
- *Planejamento urbano.* Obtenção de padrões de movimentação de condutores de veículos auxilia na adoção de medidas que permitam a distribuição mais racional de estabelecimentos comerciais e residenciais.

O alcance desse trabalho tende a ser incrementado, uma vez que a resolução 128 do CONTRAN estabelece a obrigatoriedade de instalação de dispositivos rastreadores em todos os veículos fabricados no país a partir de agosto de 2013. Com o crescimento notório da base de veículos rastreados, conseqüentemente, a necessidade por ferramentas para análise de dados tempo-espaciais oriundos dessa base crescerá substancialmente. Esse fato e, ainda, o notório e acelerado aumento da frota de veículos brasileira, abrirão um mercado bastante grande para o uso de ferramentas que auxiliem no processamento e, conseqüentemente, na extração de inteligência dessa informação.

#### **AGRADECIMENTOS**

Esse trabalho contou com o apoio do CNPq e do FIP-PUC Minas.

## REFERÊNCIAS

ABREU, J. F.; MIRANDA-JR., P. O. Mapping geographical voices, with an application of hagerstrand's space-time model. In: MEETING ASSOCIATION OF AMERICAN GEOGRAPHERS - AAG MEETING, 2007. **Annals...** S.Francisco, 2007. v. 1.

AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast algorithms for mining association rules in large databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20, 1994, Santiago, Chile. **Proceedings...** San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994. (VLDB '94), p. 487–499. ISBN 1-55860-153-8.

ALVARES, Luis Otavio et al. A model for enriching trajectories with semantic geographical information. In: ANNUAL ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 15, 2007, Seattle. **Proceedings...** New York, NY, USA: ACM, 2007. (GIS '07), p. 22:1–22:8. ISBN 978-1-59593-914-2.

ANSELIN, Luc. Local indicators of spatial associationa lisa. **Geographical Analysis**, Blackwell Publishing Ltd, v. 27, n. 2, p. 93–115, 1995. ISSN 1538-4632.

BOGORNY, Vania. **Enhancing Spatial Association Rule Mining in Geographic Databases**. 2006. Tese (Doutorado) — UFRGS, Porto Alegre.

GEOPKDD. First report on methods for mining trajectories of moving objects.: Geographic privacy-aware knowledge discovery and delivery. 2006.

KAUFMAN, Leonard; ROUSSEEUW, Peter J. **Finding Groups in Data: an Introduction to Cluster Analysis**. New York: John Wiley & Sons, 1990.

NG, R.T.; HAN, Jiawei. Clarans: a method for clustering objects for spatial data mining. **IEEE Transactions on Knowledge and Data Engineering**, v. 14, n. 5, p. 1003–1016, 2002. ISSN 1041-4347.

WITTEN, Ian H.; FRANK, Eibe. **Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)**. 2. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 0120884070.