



Pós-Graduação em Ciência da Computação

Ygor César Nogueira Sousa

**MAPEAMENTO SEMÂNTICO INCREMENTAL COM  
APRENDIZAGEM ONLINE E NÃO-SUPERVISIONADA**



Universidade Federal de Pernambuco  
posgraduacao@cin.ufpe.br  
[www.cin.ufpe.br/~posgraduacao](http://www.cin.ufpe.br/~posgraduacao)

RECIFE  
2017

Ygor César Nogueira Sousa

**MAPEAMENTO SEMÂNTICO INCREMENTAL COM  
APRENDIZAGEM ONLINE E NÃO-SUPERVISIONADA**

*Trabalho apresentado ao Programa de Pós-graduação em  
Ciência da Computação do Centro de Informática da Univer-  
sidade Federal de Pernambuco como requisito parcial para  
obtenção do grau de Mestre em Ciência da Computação.*

Orientador: *Hansenclever de França Bassani*

RECIFE  
2017

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S725m Sousa, Ygor César Nogueira  
Mapeamento semântico incremental com aprendizagem online e não supervisionada / Ygor César Nogueira Sousa. – 2017.  
80 f.:il., fig., tab.

Orientador: Hansenclever de França Bassani.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2017.  
Inclui referências.

1. Inteligência computacional. 2. Aprendizagem de máquina. I. Bassani, Hansenclever de França (orientador). II. Título.

006.3

CDD (23. ed.)

UFPE- MEI 2017-148

**Ygor César Nogueira Sousa**

**Mapeamento Semântico Incremental com Aprendizagem Online e Não-supervisionada**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 20/03/2017.

**BANCA EXAMINADORA**

---

Prof. Dr. Aluizio Fausto Ribeiro Araújo  
Centro de Informática / UFPE

---

Prof. Dr. Alberto Ferreira de Souza  
Laboratório de Computação de Alto Desempenho/UFES

---

Prof. Dr. Germano Crispim Vasconcelos  
Centro de Informática / UFPE

---

Prof. Dr. Hansenclever de França Bassani  
Centro de Informática / UFPE  
**(Orientador)**

*Dedico este trabalho aos meus pais e irmã.*

# Agradecimentos

Agradeço a Deus, em primeiro lugar, por ter-me concedido força em todos os momentos para superar todas as adversidades.

Aos meus pais (José Cícero e Angela) e irmã (Natália), por tudo que representam na minha vida e por todo o apoio que sempre me deram, foi essencial para que me fosse possível chegar até aqui.

À Raíssa Cardoso, pelo carinho, apoio, suporte e paciência em todos os momentos.

Ao meu orientador, Hansenclever Bassani, por todos os ensinamentos, paciência, dedicação e companheirismo durante essa caminhada.

Aos meus colegas de laboratório, pelos momentos de troca de conhecimentos e descontração compartilhados.

A todos os professores que compartilharam comigo um pouco do seu conhecimento ao longo dessa jornada.

A todos que contribuíram de alguma forma para que eu pudesse alcançar essa conquista.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pelo auxílio financeiro.

# Resumo

A crescente inclusão de robôs móveis na vida cotidiana das pessoas, compartilhando espaço como agentes em diferentes atividades, tem dado impulso à criação de uma série de novas tecnologias compostas. Neste contexto, aparece o Mapeamento Semântico, que visa criar uma abstração ou representação do espaço em que um robô navega, a fim de proporcionar um meio para comum entendimento e comunicação entre estes robôs e seres humanos. Essa abstração é criada sob a forma de um mapa que agrega informações semânticas (isto é, informações que façam sentido para um ser humano em termos de comunicação) sobre o ambiente no qual o robô se encontra. Desta forma, esta dissertação introduz uma abordagem de Mapeamento Semântico incremental, com aprendizagem online e não-supervisionada, baseada em Mapas Auto-organizáveis (SOMs) de topologia variante no tempo. A abordagem se divide no módulo de mapeamento, o qual cria mapas topológicos de ambientes incrementalmente, enriquecidos com objetos reconhecidos como informação semântica determinante, e no módulo de categorização de lugares, dotado de um método de aprendizagem incremental, não-supervisionado, com treinamento online, baseado em SOM. Com o intuito de avaliar a viabilidade da abordagem, a mesma foi testada a partir de experimentos realizados com uma base dados reais, os quais demonstraram de forma promissora sua capacidade na aquisição incremental de mapas topológicos enriquecidos com informações semânticas e na categorização de lugares mapeados a partir destas informações.

**Palavras-chave:** Mapeamento Semântico. Mapas Auto-organizáveis. Aprendizagem de Máquina Não-supervisionada. Aprendizagem Incremental. Robótica Móvel.

# Abstract

The increasing inclusion of mobile robots in people's daily lives, sharing space as agents in a range of different activities, has given rise to the creation of a series of new composite technologies. In this context, it appears the Semantic Mapping, which aims to create an abstraction or representation of space in which a robot navigates in order to provide a means for common understanding and communication between these robots and humans. This abstraction is created in the form of a map that aggregates semantic information (i.e., information that makes sense to a human in terms of communication) about the environment in which the robot is. In this way, this dissertation introduces an incremental semantic mapping approach, with online and unsupervised learning, based on self-organizing maps (SOMs) with time-varying topology. The approach is divided into the mapping module, which incrementally creates topological maps of environments, enriched with recognized objects as determinant semantic information, and in the module of places categorization, endowed with an incremental, unsupervised learning method with online training, based on SOM. In order to evaluate the viability of the approach, it was tested in experiments with real world data, which demonstrated a promising capability for the incremental acquisition of topological maps enriched with semantic information and for the categorization of places based on this information.

**Keywords:** Semantic Mapping. Self-Organizing Maps. Unsupervised Machine Learning. Incremental Learning. Mobile Robotics.

# Lista de Figuras

1.1	(a) Um possível comando em linguagem natural para uma cadeira de rodas autônoma (Duvallet, 2015). (b) Usuário descreve o ambiente para uma cadeira de rodas robótica projetada para ajudar residentes em uma instalação de cuidados de longa duração (Walter <i>et al.</i> , 2014). . . . .	17
1.2	Plataforma de manipulação desenvolvida para coexistir com pessoas em ambientes enquanto realiza manipulação de material sob o comando de humanos por linguagem natural (Walter <i>et al.</i> , 2015). . . . .	17
2.1	Um exemplo de um mapa ampliado com informação semântica. (Pronobis, 2011) .	21
2.2	Visualização da evolução do mapa semântico com lugares hipotetizados. (Duvallet, 2015) . . . . .	22
2.3	Predição de formato de construção utilizando informação parcial. (Duvallet, 2015) .	22
2.4	(a) Um exemplo de mapa métrico 2D de um ambiente interno explorado, (b) o mapa topológico respectivo do mesmo ambiente e (c) a sobreposição dos dois mapas, onde cada nodo do mapa topológico representa uma região no mapa métrico. (Kostavelis e Gasteratos, 2015) . . . . .	23
2.5	Mapas topológicos do ambiente, com informações probabilísticas, associados a um mapa métrico. (Pronobis e Jensfelt, 2012) . . . . .	25
2.6	Um exemplo de grafo topológico semanticamente anotado. Combina informações geométricas com semânticas para diferenciar distintos lugares do mesmo tipo. (Kostavelis e Gasteratos, 2013) . . . . .	25
2.7	Um mapa métrico anotado com pistas (rótulos semânticos) interpretados de textos colocados nas paredes. (Case <i>et al.</i> , 2011) . . . . .	26
2.8	Estrutura do sistema e o fluxo de dados entre seus componentes. (Pronobis e Jensfelt, 2012) . . . . .	26
2.9	Estrutura básica do SOM em grade retangular. Cada nodo na camada de saída é conectado aos seus quatro vizinhos imediatos. (Bassani, 2014) . . . . .	29
3.1	Arquitetura da abordagem proposta. Os componentes com contorno tracejado de cor preta não foram implementados neste trabalho. . . . .	41
3.2	Interação entre SEMMAP e OLARFDSSOM em dois momentos, em marrom o momento de treinamento realizado sempre que o agente transita entre nodos do SEMMAP e em verde o momento de categorização realizado a qualquer instante que se necessite saber a categoria atual para um determinado conjunto de nodos no SEMMAP. . . . .	42

4.1	Gráficos de dispersão do CE obtido em função de cada um dos parâmetros analisados. Os valores dos parâmetros variaram nos intervalos descritos na Tabela 4.2. . . . .	54
4.2	Gráficos de dispersão do CE obtido em função de cada um dos parâmetros analisados. Os três gráficos de cima (primeira linha) foram obtidos com $a_t$ fixado em um valor ótimo e os três de baixo (segunda linha) foram obtidos fixando $a_t$ e $maxcomp$ com valores ótimos. . . . .	55
4.3	Plataformas robóticas utilizadas na aquisição das sub-bases. ActivMedia PeopleBot na COLD-Saarbrücken, ActivMedia Pioneer-3 na COLD-Freiburg e iRobot ATRV-Mini na COLD-Ljubljana. (Pronobis e Caputo, 2009) . . . . .	56
4.4	Exemplos de imagens de alguns tipos de lugares dos laboratórios de Freiburg (a) e Saarbrücken (b). (Pronobis e Caputo, 2009) . . . . .	57
4.5	Mapa de parte do laboratório de Freiburg (a) e de parte do laboratório de Saarbrücken (b). Caminhos seguidos pelos robôs nos ambientes durante a aquisição dos dados são representados por traços, neste caso, dois caminhos de cada laboratório são desenhados (em cores azul e vermelha). (Pronobis e Caputo, 2009) . . . . .	58
4.6	Gráfico do caminho percorrido na sequência 3 do caminho 2 da sub-base COLD-Freiburg. As marcações em cinza representam as coordenadas fornecidas pelo método de SLAM durante o percurso e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles. . . .	63
4.7	Gráfico do caminho percorrido na sequência 3 do caminho 1 da sub-base COLD-Saarbrücken. As marcações em cinza representam as coordenadas fornecidas pelo método de SLAM durante o percurso e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles. . . .	64
4.8	Gráfico do caminho percorrido na sequência 3 do caminho 3 da sub-base COLD-Saarbrücken. As marcações em cinza representam as coordenadas fornecidas e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles. . . . .	65
4.9	Conexão criada erroneamente entre os nodos 5 e 3 durante o processo de mapeamento da sequência 3 do caminho 3 da sub-base COLD-Saarbrücken. . . . .	66
4.10	Mapa semântico resultante da sequência Freiburg (Caminho 2) cloudy 1 utilizada no experimento. Cada coloração nos nodos representa um agrupamento diferente do OLARFDSSOM e os quadrados tracejados delimitando grupos de nodos representam as categorias dos mesmos no <i>ground truth</i> . . . . .	67

4.11	Mapa semântico resultante de uma sequência do caminho 3 da sub-base COLD-Saarbrücken utilizada no experimento. Cada coloração nos nodos representa um agrupamento diferente do OLARFDSSOM e os quadrados tracejados delimitando grupos de nodos representam as categorias dos mesmos no <i>ground truth</i> . . . . .	72
4.12	Demonstração do CE (a) e da Acurácia (b) obtidos (eixo y) na realização de testes com cada uma das 18 sequências do primeiro conjunto (eixo x) em dois diferentes momentos, logo após a aprendizagem (treinamento) completa daquela sequência e após a aprendizagem de todas as sequências de caminhos selecionadas. . . . .	73
4.13	Demonstração do CE (a) e da Acurácia (b) obtidos (eixo y) na realização de testes com cada uma das 6 sequências do segundo conjunto (eixo x) em dois diferentes momentos, logo após a aprendizagem (treinamento) completa daquela sequência e após a aprendizagem de todas as sequências de caminhos selecionadas. . . . .	73

# Lista de Tabelas

4.1	Configuração de parâmetros utilizada em cada um dos experimentos. . . . .	52
4.2	Intervalo paramétrico para a abordagem. . . . .	55
4.3	Relação do número total de caminhos e sequências presentes na base de dados COLD completa e do que foi utilizado nesta dissertação. Caminho foi abreviado para “Cam” e sequência para “Seq”. . . . .	57
4.4	Relação de categorias, número de nodos e suas representações em números utilizados nos experimentos e avaliações. . . . .	59
4.5	Resultado da visualização de separabilidade dos dados resultantes das categorias de 6 sequências (juntas) do caminho 2 da sub-base COLD-Freiburg. As categorias são representadas em números e o nome categoria foi abreviado para Ctg. . . . .	60
4.6	Resultado da visualização de separabilidade dos dados resultantes das categorias de 6 sequências (juntas) do caminho 2 da sub-base COLD-Saarbrücken. As categorias são representadas em números e o nome categoria foi abreviado para Ctg. . . . .	61
4.7	Resultado da visualização de separabilidade dos dados resultantes das categorias de 3 sequências (juntas) do caminho 1 da sub-base COLD-Freiburg. As categorias são representadas em números e o nome categoria foi abreviado para Ctg. . . . .	61
4.8	Resultado da visualização de separabilidade dos dados resultantes das categorias de 3 sequências (juntas) do caminho 3 da sub-base COLD-Saarbrücken. As categorias são representadas em números e o nome categoria foi abreviado para Ctg. . . . .	62
4.9	Resultados apresentados por (Constante <i>et al.</i> , 2013) em valores de Acurácia. Na terceira coluna são listados os resultados do autor utilizando apenas o descritor SPMK, na quarta o SPACT e na quinta a combinação dos dois. Os resultados da abordagem proposta nesta dissertação são descritos na última coluna. Freiburg foi abreviado para <i>Fr</i> e Saarbrücken para <i>Sa</i> . Cloudy ou night representam o estado de iluminação do caminho no momento que a sequência foi capturada. . . . .	66
4.10	Resultados obtidos neste trabalho para cada par de sequências descritas nas duas primeiras colunas, utilizando o par completo para treinamento e apenas a segunda sequência do par para teste. As últimas colunas dispõem os números de agrupamentos obtidos e os números de categorias desejados. Freiburg foi abreviado para <i>Fr</i> e Saarbrücken para <i>Sa</i> . Cloudy ou night representam o estado de iluminação do caminho no momento que a sequência foi capturada. . . . .	67

4.11	Resultados apresentados por <a href="#">Rubio et al. (2016)</a> em forma de Acurácia. Na segunda coluna são listados os resultados para o caminho 2 da sub-base COLD-Freiburg e na terceira coluna os resultados para o caminho 2 da sub-base COLD-Saarbrücken. Também foram adicionados os resultados desde trabalho para fins de comparação em perspectiva geral, os valores entre parêntesis representam o desvio padrão. . . .	69
4.12	Resultados obtidos com a abordagem proposta nesta dissertação no experimento da Seção 4.7 para os caminhos de número 2 das sub-bases COLD-Freiburg e COLD-Saarbrücken. Os valores entre parêntesis representam o desvio padrão de cada resultado. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados. . . . .	70
4.13	Resultados médios obtidos com diferentes conjuntos de treinamento e teste. Os valores entre parêntesis representam o desvio padrão de cada resultado. A primeira coluna coloca as sub-bases que foram utilizadas no treinamento e teste, onde Ambas significa a utilização conjunta dos dados da Freiburg e Saarbrücken. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados. . . . .	71
4.14	Resultados médios obtidos da utilização da abordagem proposta com diferentes números máximo de agrupamentos. Os valores entre parêntesis representam o desvio padrão de cada resultado. As mesmas sequências são utilizadas para treinamento e teste. A primeira coluna coloca os números máximos de agrupamentos utilizados, sendo estes, o número exato de categorias na base de dados utilizada e um valor bastante alto. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados. . . . .	71

# Lista de Acrônimos

<b>CE</b>	Clustering Error
<b>COLD</b>	COsy Localization Database
<b>DSSOM</b>	Dimension Selective Self-Organizing Map
<b>HOG</b>	Histogram of Oriented Gradients
<b>OLARFDSSOM</b>	Online Local Adaptive Receptive Field Dimension Selective Self-Organizing Map
<b>LARFDSSOM</b>	Local Adaptive Receptive Field Dimension Selective Self-Organizing Map
<b>LHS</b>	Lantini Hypercube Sampling
<b>NB</b>	Naive Bayes
<b>SLAM</b>	Simultaneous Localization and Mapping
<b>SOM</b>	Self-Organizing Map
<b>SPMK</b>	Spatial Pyramid Matching Kernel
<b>SPACT</b>	Spatial Principal Component Analysis of Census Transform Histograms
<b>TAN</b>	Tree Augmented Naive Bayes

# Sumário

<b>1</b>	<b>Introdução</b>	<b>16</b>
1.1	Objetivos . . . . .	18
1.2	Estrutura do Trabalho . . . . .	19
<b>2</b>	<b>Fundamentação Teórica</b>	<b>20</b>
2.1	Mapeamento Semântico . . . . .	20
2.1.1	Mapas Topológicos . . . . .	24
2.1.2	Percepção Baseada em Pistas . . . . .	25
2.1.3	Coerência Temporal . . . . .	27
2.1.4	Contextualização do Modelo Proposto . . . . .	27
2.2	Mapas Auto-organizáveis . . . . .	27
2.2.1	Mapa de Kohonen . . . . .	29
2.2.2	Dimension Selective Self-Organizing Map - DSSOM . . . . .	30
2.2.2.1	Escolha do Nodo Vencedor . . . . .	31
2.2.2.2	Atualização do Vetor de Pesos . . . . .	32
2.2.2.3	Auto-Organização . . . . .	32
2.2.2.4	Agrupamento . . . . .	33
2.2.3	Local Adaptive Receptive Field Dimension Selective Self-Organizing Map - LARFDSSOM . . . . .	33
2.2.3.1	Inserção de Nodos e Competição . . . . .	35
2.2.3.2	Atualização do Nodo Vencedor e Seus Vizinhos . . . . .	36
2.2.3.3	Remoção de Nodos . . . . .	37
2.2.3.4	Atualização da Vizinhança . . . . .	37
2.2.3.5	Convergência . . . . .	38
2.2.3.6	Agrupamento . . . . .	38
2.3	Conclusão . . . . .	39
<b>3</b>	<b>Abordagem Desenvolvida</b>	<b>40</b>
3.1	Descrição da Abordagem . . . . .	40
3.2	SEMMA . . . . .	43
3.2.1	Competição . . . . .	43
3.2.2	Atualização de Nodo Vencedor . . . . .	44
3.2.3	Transição entre Nodos . . . . .	46
3.3	OLARFDSSOM . . . . .	46
3.3.1	Competição . . . . .	47

3.3.2	Inserção e Remoção de Nodos . . . . .	48
3.3.3	Atualização do Nodo Vencedor e Seus Vizinhos . . . . .	49
3.3.4	Atualização da Vizinhança . . . . .	49
3.3.5	Agrupamento no OLARFDSSOM . . . . .	49
3.4	Conclusão . . . . .	50
<b>4</b>	<b>Experimentos e Análise de Resultados</b>	<b>51</b>
4.1	Medidas de Qualidade . . . . .	52
4.2	Ajustes Paramétricos . . . . .	53
4.3	Descrição da Base de Dados . . . . .	54
4.4	Avaliação de Separabilidade das Categorias dos Nodos do SEMMAP . . . . .	58
4.5	Avaliação da Topologia dos Mapas Criados . . . . .	61
4.6	Experimento Comparativo 1 . . . . .	63
4.7	Experimento Comparativo 2 . . . . .	68
4.8	Experimento de Avaliação Incremental . . . . .	69
4.9	Estudo de Caso . . . . .	72
4.10	Conclusão . . . . .	74
<b>5</b>	<b>Considerações Finais</b>	<b>75</b>
5.1	Limitações . . . . .	75
5.2	Trabalhos Futuros . . . . .	76
	<b>Referências</b>	<b>78</b>

# 1

## Introdução

Até recentemente, robôs que operavam fora de laboratórios eram limitados a ambientes controlados, preparados explicitamente para prevenir interação com humanos. Contudo, existe uma demanda crescente por robôs que atuem não como máquinas em situações isoladas, mas como co-habitantes que auxiliam humanos em uma variedade de atividades do cotidiano. Isto evidencia a necessidade de que para robôs atuarem efetivamente com humanos como colegas de equipe, eles precisam eficientemente e flexivelmente interpretar e lidar com pedidos em linguagem natural. Reconhecendo essa necessidade, o foco em pesquisas tem aumentado em permitir a robôs interpretar os ambientes em que estão de forma que lhes permita se comunicar e atender comandos em linguagem natural (Dzifcak *et al.*, 2009; Tellex *et al.*, 2011; Walter *et al.*, 2014).

Essa capacidade poderia, por exemplo, permitir a um agente dos bombeiros direcionar um micro veículo aéreo falando “Voe pelas escadas, percorra o corredor e verifique a segunda sala à direita depois da cozinha”. Com isso, um desafio fundamental é associar corretamente elementos linguísticos de um comando ao entendimento do mundo externo de um robô. Logo, esse problema pode ser mitigado permitindo a robôs formular representações de conhecimento que modelam as propriedades semânticas de alto nível de seus ambientes (Walter *et al.*, 2014).

Neste cenário, mapas semânticos vem para sanar essa necessidade, provendo robôs com modelos de seus ambientes centrados em humanos (Pronobis e Jensfelt, 2012). Essas abordagens geralmente tomam como entrada sensores de baixo nível (por exemplo, *laser range scanners* e câmeras) e estendem modelos tradicionais de mapeamento (métricos e topológicos) de ambientes para incluir propriedades semânticas de alto nível, ou seja, informações que façam sentido para um ser humano em termos de comunicação por linguagem natural. Modelos existentes populam o mapa semântico com propriedades semânticas da cena (como tamanho, categoria, forma e objetos presentes em um lugar) que podem ser inferidas, por exemplo, de classificadores baseados em imagens, dados de *laser scanners* ou em outros atributos semânticos (principalmente na inferência da categoria de um lugar) (Walter *et al.*, 2014).

Vários trabalhos na literatura fazem uso de mapas semânticos sob diferentes perspectivas e os integram em *frameworks* envolvendo humanos e robôs. Utilizados em aplicações como



**Figura 1.1:** (a) Um possível comando em linguagem natural para uma cadeira de rodas autônoma (Duvallet, 2015). (b) Usuário descreve o ambiente para uma cadeira de rodas robótica projetada para ajudar residentes em uma instalação de cuidados de longa duração (Walter *et al.*, 2014).

robôs domésticos, robôs companheiros, empilhadeiras autônomas, carros autônomos, dentre outras. Em Duvallet (2015), mapas semânticos são incorporados a uma solução aplicada a cadeira de rodas que proporciona comandos por linguagem natural (a Figura 1.1(a) ilustra um exemplo). Ainda aplicando ao uso em cadeira de rodas, Walter *et al.* (2014) incorpora descrições em linguagem natural providas por um agente externo (i.e. uma pessoa) sobre lugares como informação incorporada aos mapas semânticos criados (exemplificado na Figura 1.1(b)). Um outro trabalho proposto por Walter *et al.* (2015), faz o uso de mapas semânticos em conjunto com interação por linguagem natural em uma empilhadeira utilizada para a manipulação de material (Figura 1.2).



**Figura 1.2:** Plataforma de manipulação desenvolvida para coexistir com pessoas em ambientes enquanto realiza manipulação de material sob o comando de humanos por linguagem natural (Walter *et al.*, 2015).

Desta maneira, para que robôs possam ter uma longa vida de utilização é necessário que aos mesmos seja incorporado aprendizado em longo termo, organizado e incremental (conhecimento construído gradualmente) (Karaoğuz e Bozma, 2016). Portanto, a utilização de abordagens com aprendizagem supervisionada poderiam prover boa performance, porém iriam requerer frequente intervenção humana para atualização de sua arquitetura ou retreinamento, dificultando

o uso em longo termo. Alternativamente, abordagens com aprendizagem não-supervisionada permitiriam reduzir a necessidade de tais intervenções, principalmente se associadas a uma estratégia de aprendizagem incremental poderiam, além disso, continuar aprendendo conforme mais informação seja fornecida.

Na perspectiva espacial da criação de mapas semânticos, a maioria dos trabalhos presentes na literatura realiza a operação de forma incremental, criando mapas progressivamente enquanto operam e incorporando a eles informações semânticas reconhecidas no ambiente. Entretanto, quando se considera as categorias de lugar dos ambientes mapeados, grande parte dos modelos atualmente propostos inferem automaticamente sobre essa propriedade utilizando de algum método com aprendizagem não incremental e supervisionada (Pronobis e Jensfelt, 2012; Kostavelis e Gasteratos, 2013, 2017), alguns modificam métodos de aprendizagem supervisionada para aprender categorias adicionais incrementalmente (Sunderhauf *et al.*, 2016) e outros não realizam categorização automática (Bastianelli *et al.*, 2013; Gemignani *et al.*, 2016). Isso limita suas abordagens para atuação em atividades de longa duração.

Perante o apresentado, o problema de pesquisa investigado neste trabalho pode ser definido como:

**Problema de Pesquisa:** *É possível e viável a criação de uma abordagem que permita a robôs formular automaticamente representações de conhecimento que modelam as propriedades semânticas de alto nível de seus ambientes, incorporando aprendizagem em longo termo, não-supervisionada e incremental?*

Diante do exposto, na seção seguinte são delineados os objetivos a serem atingidos neste trabalho, os quais foram especificados com o intuito de responder o problema de pesquisa descrito.

## 1.1 Objetivos

De forma geral, o objetivo deste trabalho é propor uma abordagem de mapeamento semântico incremental e não-supervisionada baseada em Mapas Auto-Organizáveis (SOM), com características online e de topologia variante no tempo. Contendo como objetivos específicos:

- Desenvolvimento de abordagem incremental baseada em SOM para criação de mapas topológicos enriquecidos com informação semântica;
- Utilização de algoritmo baseado em Mapas Auto-organizáveis de topologia variante no tempo com aprendizagem online e não-supervisionada para categorização de lugares baseando-se em informação semântica.

## 1.2 Estrutura do Trabalho

Em continuidade ao apresentado, os demais capítulos da presente dissertação estão organizados da seguinte maneira: o Capítulo 2 dispõe os principais conceitos e fundamentos teóricos que contribuíram para a construção da abordagem; o Capítulo 3 descreve a abordagem desenvolvida neste trabalho, com fluxo de funcionamento, módulos identificados e técnicas utilizadas; o Capítulo 4 apresenta o planejamento e execução dos experimentos realizados, assim como análise dos resultados obtidos; por fim, no Capítulo 5 são feitas as considerações finais do trabalho, indicando as presentes limitações e possíveis trabalhos futuros.

# 2

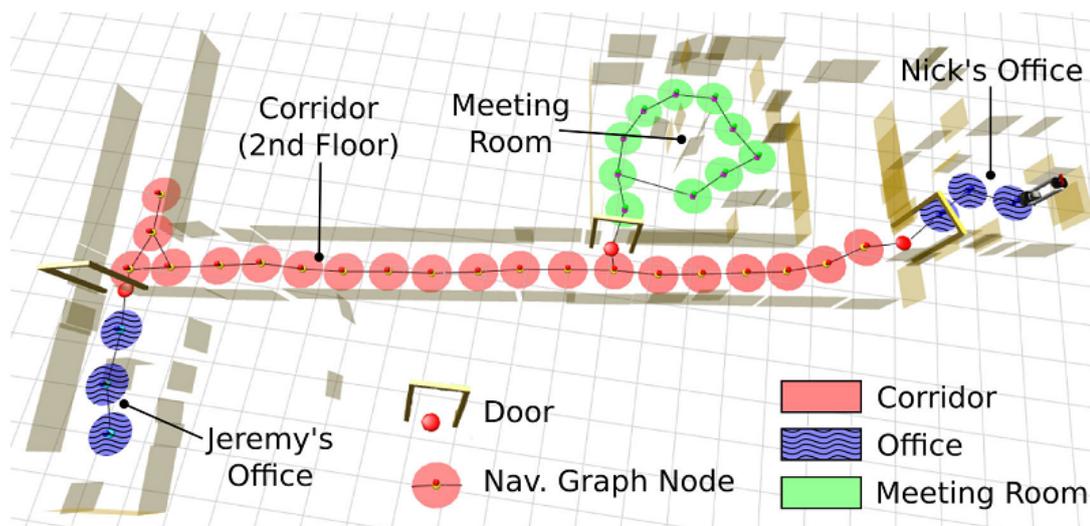
## Fundamentação Teórica

Este capítulo introduz os conceitos e fundamentos que serviram de base para a construção da abordagem proposta neste trabalho. Primeiramente, conforme mencionado no capítulo anterior, esta dissertação tem por objetivo propor uma abordagem que realize aquisição de conhecimento semântico em forma de mapa sobre ambientes visitados por um agente. Esta é uma proposição popularizada recentemente, que agrega muitas outras áreas de conhecimento, denominada na literatura de Mapeamento Semântico (Pronobis e Jensfelt, 2012; Bastianelli *et al.*, 2013; Kostavelis e Gasteratos, 2015; Kollar *et al.*, 2013). A Seção 2.1 descreve algumas das principais características deste problema, introduzindo alguns trabalhos presentes na área.

Os ambientes mapeados, em geral são divididos em regiões (ou seja, lugares) e estas recebem categorias (ou classes) inferidos por métodos inteligentes através de informações semânticas identificadas nessas regiões no momento do mapeamento, a categoria em si agrega mais uma informação semântica à região. Logo, como método inteligente, destinado a essa atividade de categorização, esta dissertação buscou implementar um método de aprendizagem não-supervisionada baseado no LARFDSSOM (*Local Adaptive Receptive Field Dimension Selective Self-Organizing Map*) (Bassani e Araujo, 2015), um Mapa Auto-organizável de topologia variante no tempo. Os Mapas Auto-organizáveis, assim como o LARFDSSOM, têm sua origem e características detalhadas na Seção 2.2.

### 2.1 Mapeamento Semântico

A preocupação do mapeamento semântico é permitir a robôs criar modelos dos seus ambientes que sejam centrados em humanos, para que assim, quando interagirem com pessoas eles possam raciocinar sobre propriedades de alto nível (Duvallet, 2015). Ter representações em mapa tão ricas permite a robôs realizar ações que requerem raciocínio sobre componentes com significado semântico no mapa. Bastianelli *et al.* (2013) define Mapeamento Semântico como o termo usado para denotar o processo que permite ao robô enriquecer o mapa utilizado para navegação com informação semântica sobre o ambiente. A Figura 2.1 apresenta uma ilustração de um mapa ampliado com informação semântica.



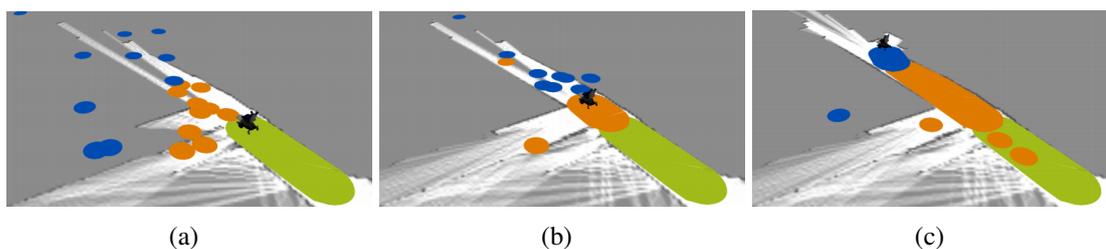
**Figura 2.1:** Um exemplo de um mapa ampliado com informação semântica. (Pronobis, 2011)

A maior parte das abordagens utilizadas na construção de mapas semânticos encontradas no estado da arte realizam interpretação autônoma de informações, ou seja, não requerem entradas de um usuário humano em sua geração (Bastianelli *et al.*, 2013). Outro tipo de abordagens encontradas são aquelas apoiadas por interação com um operador humano, possibilitando aos robôs construir representações do ambiente mais centradas no humano e recolher informações difíceis de serem deduzidas automaticamente. Nesse tipo de abordagem é bastante comum a utilização de interfaces multimodais (incluindo linguagem natural) para construir um mapa do ambiente (Bastianelli *et al.*, 2013; Kollar *et al.*, 2013; Walter *et al.*, 2014; Deeken *et al.*, 2015), possibilitando ao usuário descrever em linguagem natural o lugar em que o robô está ou apontar para um objeto e o descrever (Duvalllet, 2015).

Uma possibilidade que utiliza uma abordagem apoiada na interação humano-máquina são os Grafos Semânticos, um *framework* apresentado em Hemachandra *et al.* (2014) como um mapa que acopla informações métricas, topológicas e semânticas do ambiente, contendo locais categorizados semanticamente, descritos em linguagem natural por um operador. Essa abordagem estende um mapa métrico de SLAM (do inglês, *Simultaneous Localization and Mapping*) padrão com uma representação topológica do ambiente e propriedades semânticas de cada região. As regiões são categorizadas utilizando um método probabilístico supervisionado através de propriedades semânticas inferidas utilizando classificação de cena baseada em imagem e laser, assim como por descrição em linguagem natural (Hemachandra *et al.*, 2014; Duvalllet, 2015).

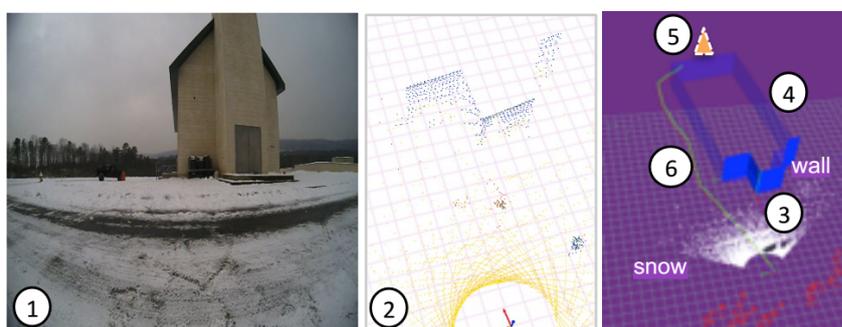
O que pode ser encarado como uma forma de evolução da abordagem de Grafos Semânticos proposta por Hemachandra *et al.* (2014), mencionada anteriormente, Duvalllet (2015) adiciona a habilidade de criar hipóteses sobre novos lugares no ambiente utilizando informação contida em comandos por linguagem natural, buscando tratar a linguagem em si como mais um sensor a ser utilizado na construção do mapa, abordagem a qual, apresentou bons resultados

em ambientes internos (*indoor*). A Figura 2.2 ilustra a evolução do mapa semântico do robô ao seguir o comando “*go to the kitchen that is down the hallway*”. Regiões amostrais são desenhadas em círculos pequenos e regiões visitadas são desenhadas com a área preenchida. *Hallway* é representado pela cor laranja, *kitchen* pela cor azul e onde o robô inicialmente se encontrada pela cor verde. O robô primeiramente estima as possíveis localizações da cozinha e se move em sua direção (Figura 2.2(a)), em seguida observa o corredor (*hallway*) e então refina sua estimativa usando a relação “*down*”provida pelo usuário (Figura 2.2(b)). Logo depois, o robô chega a cozinha (*kitchen*) e declara que finalizou a direção dada pelo comando (Figura 2.2(c)).



**Figura 2.2:** Visualização da evolução do mapa semântico com lugares hipotetizados. (Duvallet, 2015)

Neste mesmo trabalho, Duvallet (2015) apresenta uma extensão da sua abordagem para a operação em ambientes externos (*outdoor*), oriunda de um trabalho publicado anteriormente por Oh *et al.* (2015). Nessa extensão, o trabalho introduz a previsão do formato de construções inicialmente não completamente perceptíveis. Ele combina seus dados anteriormente obtidos com padrões geométricos em sua memória declarativa. Por exemplo, na Figura 2.3, utilizando uma visão da frente da construção o robô extrai grupos de nuvens de pontos das paredes e prediz o restante da construção baseando-se em exemplos da sua memória e na localização do seu objetivo descrito em linguagem natural.

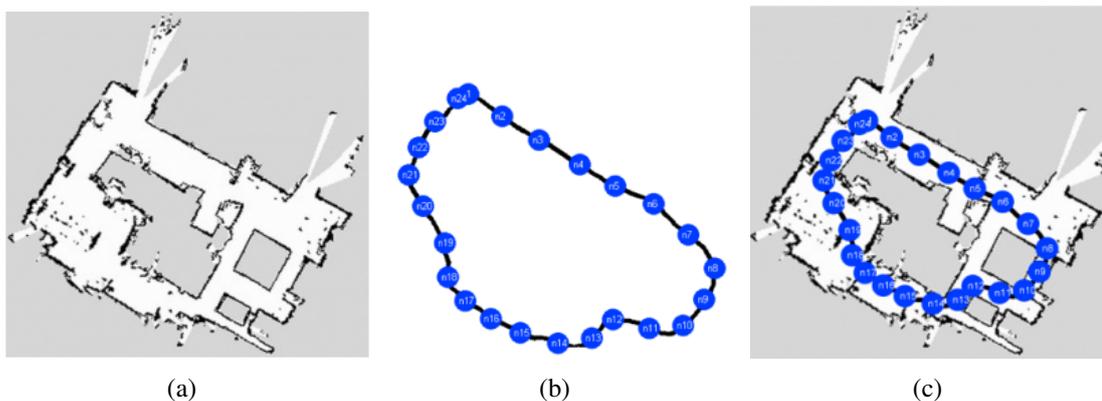


**Figura 2.3:** Predição de formato de construção utilizando informação parcial. (Duvallet, 2015)

Em Kostavelis e Gasteratos (2015) é feito um levantamento da literatura científica e são destacadas características principais frequentemente presentes em metodologias de mapeamentos semântico. Características consideradas essenciais para que um método de mapeamento

semântico seja considerado completo. A percepção baseada em pistas é uma dessas características. Onde em alguns casos, apenas pistas semânticas únicas (*single cues*), como objetos, são utilizadas para inferir sobre a categoria de um lugar, enquanto outras metodologias exploram em conjunto múltiplas pistas (*multiple cues*) semanticamente lógicas sobre uma área, como objetos, texturas e formas (Kostavelis e Gasteratos, 2015).

A existência de um mapa topológico é outro atributo importante que um método de mapeamento semântico típico possui, o qual é uma abstração do ambiente explorado geralmente em termos de um grafo. A existência de um mapa métrico 3D ou 2D do ambiente explorado, seja ele interior (*indoor*) ou exterior (*outdoor*) é considerado um componente complementar que geralmente provê suporte aos atributos implementados pelos métodos de lógica semântica, entretanto algumas abordagens na literatura trabalham anotando diretamente os mapas métricos, como em Case *et al.* (2011) e Gemignani *et al.* (2016). A Figura 2.4 demonstra um exemplo de construção de mapa topológico a partir de informações fornecidas por um mapa métrico 2D.



**Figura 2.4:** (a) Um exemplo de mapa métrico 2D de um ambiente interno explorado, (b) o mapa topológico respectivo do mesmo ambiente e (c) a sobreposição dos dois mapas, onde cada nodo do mapa topológico representa uma região no mapa métrico. (Kostavelis e Gasteratos, 2015)

Em geral, mapas semânticos são de larga escala, ou seja, criados incrementalmente e relacionados a um sistema de coordenadas global, simultaneamente anotados com características de alto nível como tipos de objetos, interpretações de formas, rótulos linguísticos e categorias de lugares. Entretanto, na literatura existem trabalhos que não fazem a aquisição de mapas semânticos dos ambientes mas inferem a localização semântica (ou seja, o tipo de lugar em que está inserido) de um agente por meio do raciocínio de rótulos (i.e., categorias de lugares como banheiro ou escritório) uma instância de quadro por vez, conhecidos como métodos de cena única. Os trabalhos descritos em Constante *et al.* (2013) e Rubio *et al.* (2016) se encaixam neste contexto e são utilizados mais a frente no capítulo de experimentação por utilizarem a mesma base de dados usada nesta dissertação.

Uma outra característica importante destacada por Kostavelis e Gasteratos (2015) está exatamente na utilização de informações adquiridas em vários instantes de tempo durante a

aquisição do mapa, denominada de coerência temporal. Métodos baseados apenas em cena única perdem essa possibilidade.

Desta forma, as subseções posteriores detalham um pouco mais as características apresentadas e expõem alguns trabalhos acadêmicos encontrados em mapeamento semântico incorporando as características em questão (Subseções 2.1.1, 2.1.2 e 2.1.3). Além disso, a abordagem de mapeamento semântico proposta nesta dissertação é contextualizada na Subseção 2.1.4.

### 2.1.1 Mapas Topológicos

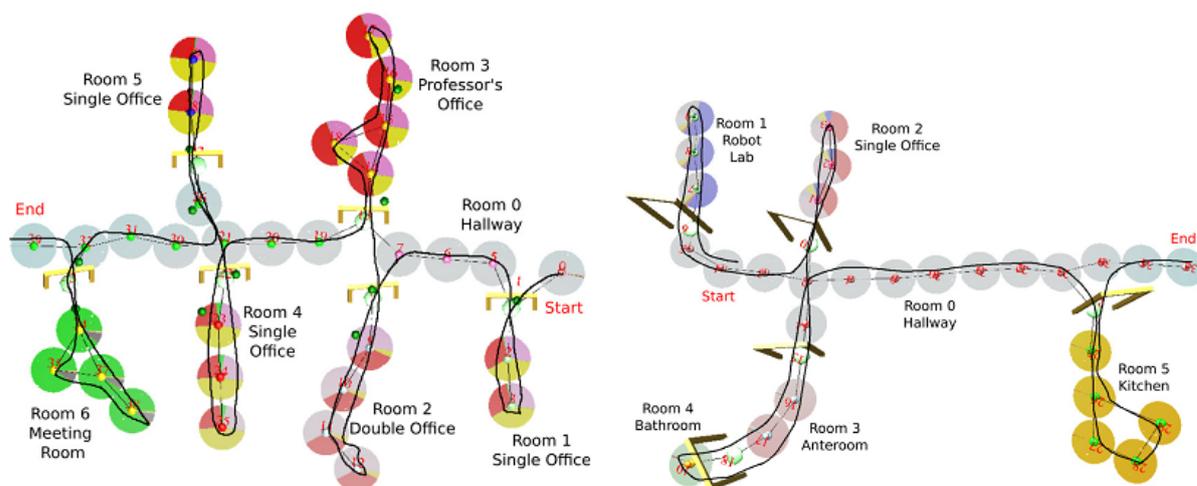
Uma forma de revelar informações relacionais escondidas nos mapas métricos, que são organizados de forma geométrica para favorecer sua informação espacial, é organizá-lo em termos de um mapa topológico. Em geral, na forma de grafo, onde os vértices correspondem aos locais e as arestas aos caminhos entre os locais. Desta forma, em um mapa topológico a estruturação de um ambiente contém tanto informação geométrica para a disposição dos lugares visitados quanto a informação conceitual sobre a categoria que eles pertencem. Portanto, esses grafos constituem uma das características fundamentais no mapeamento semântico, devido ao fato deles possibilitarem uma abstração dos mapas métricos, assim como das pistas conceituais ([Kostavelis e Gasteratos, 2015](#)).

Os mapas topológicos podem conter apenas características espaciais ou semasiológicas, onde sempre que o robô viaja uma certa distância ou o mecanismo de inferência converge para uma classe de lugar específica, um nó é adicionado. Ou podem conter várias restrições de acordo com seus atributos semânticos, como (dentre outras possibilidades) a incerteza sobre a classe do lugar atualmente visitado ou a viabilidade de transição entre as áreas mapeadas. Um exemplo é o trabalho de [Krishnan e Krishna \(2010\)](#), nele um mapa semântico em forma de grafo é construído em um nível superior a um mapa métrico. Cada vértice do grafo corresponde a uma categoria de lugar reconhecida, como uma sala ou corredor. Enquanto as arestas representam um ponto de transição, como uma porta ([Kostavelis e Gasteratos, 2015](#)).

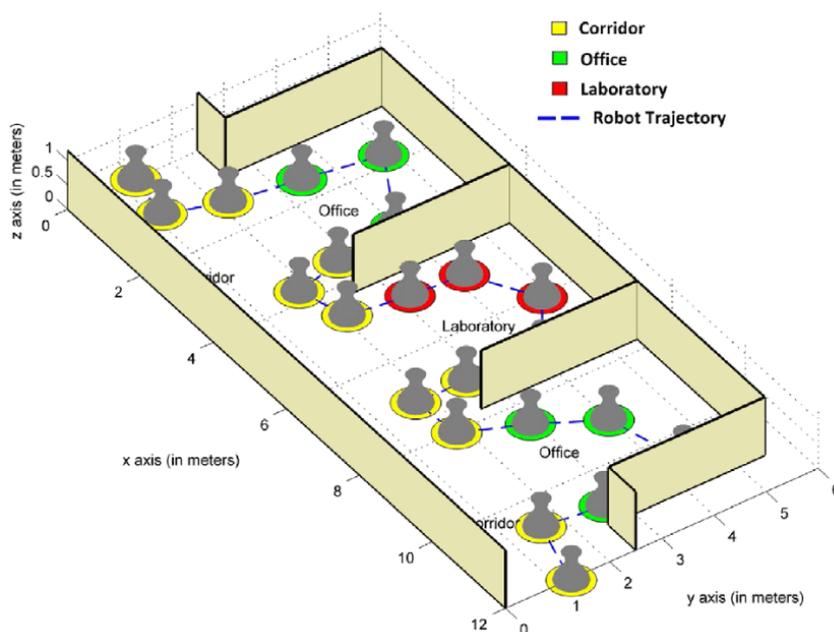
Em alguns outros trabalhos, informações probabilísticas são incorporadas aos mapas topológicos ([Pronobis e Jensfelt, 2012](#); [Pronobis et al., 2010](#); [Aydemir et al., 2011](#)), resultando em mapas semânticos mais intuitivos, como pode ser visto na Figura 2.5 ([Kostavelis e Gasteratos, 2015](#)). Nela são apresentados dois mapas topológicos do ambiente associados a mapas métricos, indicando os resultados das segmentações dos ambientes e categorizações. Os círculos indicam a localização de lugares no ambiente e a linha preta, a trajetória do robô. Os gráficos em forma de pizza indicam as distribuições de probabilidade sobre as possíveis categorias para cada lugar (cada fração do gráfico de pizza corresponde a uma categoria) ([Pronobis e Jensfelt, 2012](#)).

Além do já citado, [Kostavelis e Gasteratos \(2013\)](#) introduz um mapa topológico semanticamente anotado que contém restrições geométricas e semânticas usadas para diferenciar entre locais distintos de uma mesma categoria. Como um exemplo, na Figura 2.6 pode-se identificar a existência de dois lugares de categoria “office” em diferentes regiões (ou seja, *office 1* e *office 2*).

Um modelo de SVM (*Support Vector Machine*) é utilizado para inferir as categorias de lugar nesse trabalho.



**Figura 2.5:** Mapas topológicos do ambiente, com informações probabilísticas, associados a um mapa métrico. (Pronobis e Jensfelt, 2012)

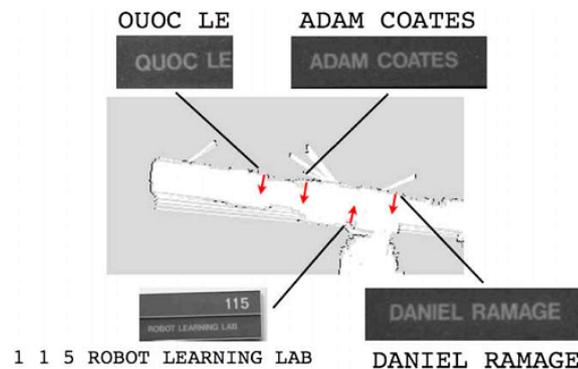


**Figura 2.6:** Um exemplo de grafo topológico semanticamente anotado. Combina informações geométricas com semânticas para diferenciar distintos lugares do mesmo tipo. (Kostavelis e Gasteratos, 2013)

### 2.1.2 Percepção Baseada em Pistas

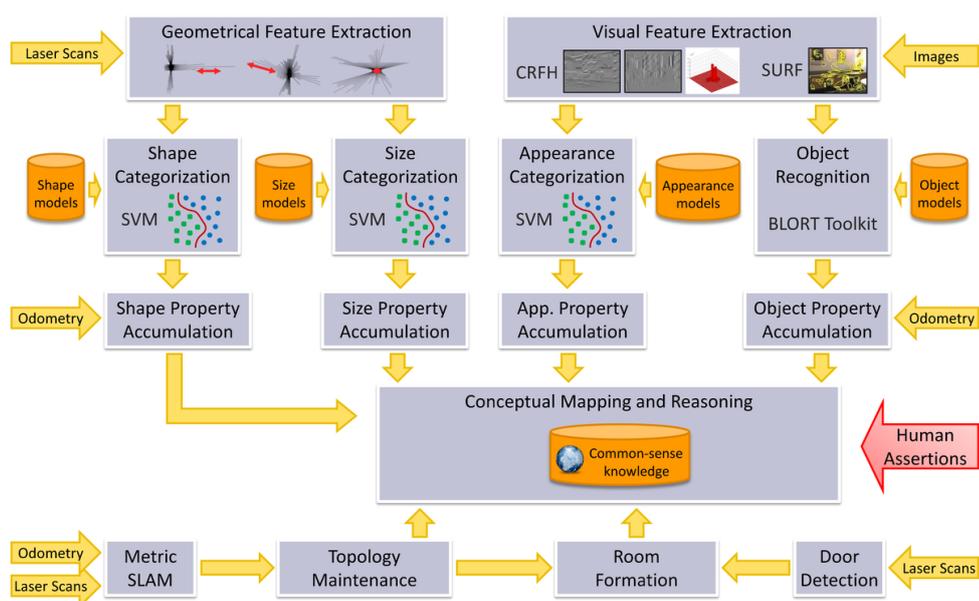
No contexto de percepção baseada em pistas únicas (*single cues*), Bastianelli *et al.* (2013) e Gemignani *et al.* (2016) descrevem modelos que incrementalmente enriquecem mapas semânticos de múltiplas camadas com objetos reconhecidos apontados por usuários utilizando

apontadores *laser*. O método proposto por [Case et al. \(2011\)](#) realiza a aquisição de mapas semânticos enriquecidos com dados de texto reconhecidos automaticamente por imagem, como número de salas e nomes de pessoas. Os textos reconhecidos são colocados na orientação em que foram reconhecidos dentro do mapa métrico criado (Figura 2.7).



**Figura 2.7:** Um mapa métrico anotado com pistas (rótulos semânticos) interpretados de textos colocados nas paredes. ([Case et al., 2011](#))

Em percepção com múltiplas pistas (*multiple cues*), no método apresentado por [Pronobis e Jensfelt \(2012\)](#) foram utilizados múltiplos sensores para reconhecer diferentes características semânticas do ambiente. O reconhecimento de objetos e aparência de lugares são feitos através de técnicas de visão computacional, enquanto os tamanhos e formas dos quartos são extraídos utilizando *laser range scanners*. As informações adquiridas são unificadas em modelos de SVM e um modelo probabilístico de *Chain Graph* ([Lauritzen e Richardson, 2002](#)) infere sobre as categorias dos lugares. A Figura 2.8 exibe a estrutura do sistema proposto por [Pronobis e Jensfelt \(2012\)](#) assim como o fluxo de dados entre seus componentes.



**Figura 2.8:** Estrutura do sistema e o fluxo de dados entre seus componentes. ([Pronobis e Jensfelt, 2012](#))

### 2.1.3 Coerência Temporal

Um outro atributo comum entre os modelos de mapeamento semântico é a coerência temporal, que pode ser levada em conta tanto na construção dos mapas topológicos quanto na formulação conceitual dos modelos utilizados para categorização de lugares. Nesse contexto, mais uma vez o trabalho de [Pronobis e Jensfelt \(2012\)](#) pode ser citado, ele defende que sistemas de mapeamento semântico desenvolvidos devem possuir integração entre espaço e tempo devido ao fato de que informação adquirida em um único ponto raramente provê evidência suficiente para uma classificação confiável de lugares e objetos ([Kostavelis e Gasteratos, 2015](#)).

### 2.1.4 Contextualização do Modelo Proposto

Apesar dos muitos avanços, a limitação que muitos dos modelos atualmente propostos na literatura tem em realizar a categorização automática dos lugares mapeados semanticamente utilizando de métodos com aprendizagem supervisionada e não incremental ([Kostavelis e Gasteratos, 2013](#); [Pronobis e Jensfelt, 2012](#); [Hemachandra et al., 2014](#); [Kostavelis e Gasteratos, 2017](#)), dificulta o uso dos mesmos em aplicações de longa duração, fazendo-se necessário intervenção humana para atualização da arquitetura dos métodos ou para retreinamento. [Sunderhauf et al. \(2016\)](#) propôs um modelo que aprende novas classes incrementalmente, mas que ainda assim requer supervisão no seu processo de aprendizagem. Uma forma de mitigar esse problema seria o desenvolvimento de uma abordagem de mapeamento semântico que pudesse aprender categorias de lugares sem supervisão e incrementalmente com o passar do tempo. Até o presente momento, o autor desta dissertação não encontrou nenhum trabalho na literatura com características similares.

Neste contexto se apresenta o modelo desenvolvido nesta dissertação. O mesmo contém todas as características destacadas no começo desta seção, construindo mapas topológicos progressivamente a partir de informações fornecidas por um método de SLAM métrico, enriquecidos com informações semânticas, explorando apenas objetos reconhecidos como pista única. Além disso, agrega um método para categorização de lugares com aprendizagem não-supervisionada, incremental e treinamento online, baseado em um mapa auto-organizável de topologia variante no tempo, o LARFDSSOM ([Bassani e Araujo, 2015](#)). Isso possibilita a aprendizagem de categorias com o passar do tempo e o uso da abordagem em aplicações de longa duração. A próxima seção descreve o funcionamento dos mapas auto-organizáveis e do LARFDSSOM.

## 2.2 Mapas Auto-organizáveis

O Mapa Auto-organizável (SOM, do inglês *Self-Organizing Map*) é uma rede neural proposta por [Kohonen \(1982\)](#), originalmente criada como uma ferramenta para visualização de dados de alta-dimensionalidade. Ela mapeia uma distribuição de alta dimensão em uma grade regular de menor dimensão, sendo capaz de compactar informações preservando os

relacionamentos topológicos e métricas dos dados originais, mostrando grande capacidade em criar abstração e exibição simplificada de informações (Bassani, 2014). Estas características fizeram com que os SOMs fossem frequentemente utilizados para estabelecer agrupamentos de forma não-supervisionada (Haykin, 1998).

A estrutura básica de um SOM típico é formada por uma camada de entrada e uma de saída. A camada de entrada propaga os valores recebidos a todos os nodos da camada de saída, na qual é criada uma distribuição topológica que representa de forma sumarizada a informação recebida nas entradas ao longo do tempo. O processo de aprendizagem no SOM além de não-supervisionado é competitivo e cooperativo. Competitivo porque os nodos na camada de saída competem entre si e aquele mais semelhante a entrada atual é considerado o vencedor, com isso, sendo atualizado de forma a melhor representar esta entrada. Cooperativo porque além do nodo vencedor ser atualizado, seus vizinhos topológicos também são.

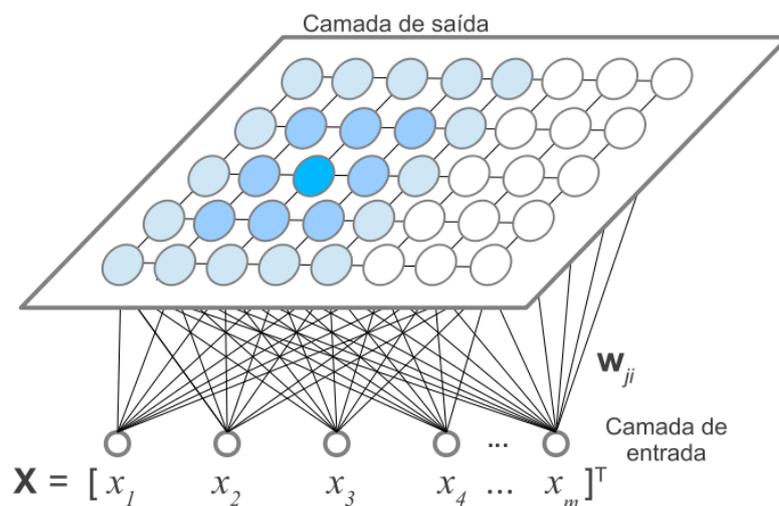
As competições em SOMs tradicionais são realizadas através do cálculo da distância euclidiana simples, dos nodos presentes na camada de saída para os padrões de entrada recebidos, considerando-se todas as dimensões dos dados nesses padrões de forma igualitária. No entanto, em alguns tipos de aplicações, principalmente em dados de alta dimensionalidade, muitas dessas dimensões são possivelmente irrelevantes para certo grupos e relevantes para outros. Podendo ocorrer diversos problemas no processo de agrupamento pela consideração igual de todas as dimensões, como a possível ocultação de agrupamentos pela consideração de dados relevantes como ruído. Uma forma de lidar com este problema está na utilização de uma distância euclidiana ponderada nas competições, possibilitando ao método aplicar relevâncias (pesos) diferentes para cada uma das dimensões de entrada. O DSSOM (*Dimension Selective Self-Organizing Map*), proposto por Bassani e Araujo (2012), é um mapa auto-organizável seletivo a dimensões de entrada que utiliza desse tipo de procedimento, descrito em detalhes na Subseção 2.2.2.

A camada de saída no SOM clássico de Kohonen (1982) (detalhado na Subseção 2.2.1) possui topologia fixa, ou seja, o número de nodos e conexões entre eles são estabelecidos com antecedência. Fazendo-se necessário conhecimento prévio sobre o conjunto de dados a ser analisado para se obter bons resultados, mesmo assim o mapa produzido em muitos casos pode não se adequar bem aos dados, prejudicando a qualidade dos agrupamentos formados. Por isso e pela necessidade em diferentes aplicações de que o SOM se adapte incrementalmente conforme dados são disponibilizados, muitos modelos foram desenvolvidos (Fritzke, 1994; Kunze e Steffens, 1995; Marsland *et al.*, 2002) com o objetivo de criar mapas auto-organizáveis com topologia que pudesse variar dinamicamente com o tempo, modificando o número de nodos e as relações de vizinhança entre eles. Na Subseção 2.2.3 é descrito um desses modelos, o LARFDSSOM, proposto por Bassani e Araujo (2015) como uma evolução com topologia variante no tempo ao DSSOM, mencionado anteriormente.

### 2.2.1 Mapa de Kohonen

O Mapa Auto-organizável (SOM), também conhecido como mapas de Kohonen, foi proposto por [Kohonen \(1982\)](#) como um dos primeiros trabalhos sobre redes neurais artificiais de aprendizagem não-supervisionada nas décadas de 1970 e 1980, servindo de base para um número expressivo de derivações até os dias de hoje. O funcionamento básico do SOM define um mapeamento partindo de um conjunto de dados multidimensional de entrada para uma grade regular, usualmente bidimensional de nodos. O método é capaz de transformar relacionamentos complexos e não lineares entre dados de alta dimensionalidade em uma visualização simples de baixa dimensionalidade.

Conforme mencionado anteriormente, o SOM é composto de duas camadas. A camada de entrada é um vetor de nodos  $\mathbf{x}$  de tamanho  $m$ , onde  $m$  é o número de dimensões do espaço de entrada. A camada de saída é uma grade de nodos organizados segundo uma distribuição de vizinhança geralmente retangular (Figura 2.9), onde cada nodo na camada de saída é conectado aos seus vizinhos imediatos. Cada nodo  $j$  na camada de saída está associado a um vetor de pesos sinápticos  $\mathbf{w}_j$  de tamanho  $m$ , ou seja, um peso é utilizado para cada nodo na cama de entrada.



**Figura 2.9:** Estrutura básica do SOM em grade retangular. Cada nodo na camada de saída é conectado aos seus quatro vizinhos imediatos. ([Bassani, 2014](#))

O processo de aprendizagem é composto pelas etapas de competição, adaptação e cooperação. Primeiramente, sempre que um estímulo  $\mathbf{x}$  é apresentado a camada de entrada do mapa, ele é propagado para todos os nodos da camada de saída e os mesmos competem para determinar aquele que mais se aproxima do estímulo. O nodo vencedor é aquele cujo vetor de pesos sinápticos,  $\mathbf{w}_j$ , apresenta a menor distância para o estímulo  $\mathbf{x}$ , geralmente determinada pela distância euclidiana:

$$v(\mathbf{x}) = \arg \min_j [D(\mathbf{x}, \mathbf{w}_j)] \quad (2.1)$$

onde  $D(\mathbf{x}, \mathbf{w}_j)$  é distância euclidiana simples descrita abaixo, assumindo  $m$  como o número de

dimensões presente na camada de entrada:

$$D(\mathbf{x}, \mathbf{w}_j) = \sqrt{\sum_{i=1}^m (x_i - w_{ji})^2} \quad (2.2)$$

A etapa de adaptação envolve o nodo vencedor, movendo seu vetor de pesos,  $\mathbf{w}_j$ , em direção ao estímulo  $\mathbf{x}$ . Realizada conforme a equação que segue:

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n)h_{jv(x)}(n)(\mathbf{x} - \mathbf{w}_j(n)) \quad (2.3)$$

onde  $h_{jv(x)}$  é a função de decaimento da vizinhança e  $\eta(n)$  indica a taxa de aprendizagem, calculada através de uma função que decai com o decorrer das  $n$  iterações, conforme a Eq 2.4.

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_1}\right), n = 0, 1, 2, \dots \quad (2.4)$$

onde  $\tau_1$  é a constante de tempo que ajusta a velocidade de decaimento.

A cooperação é feita através da adaptação dos nodos vizinhos do nodo vencedor  $j$ , movendo-os também em direção ao estímulo de entrada com um menor nível de adaptação reduzido pela função  $h_{jv(x)}$  da Eq. 2.5. Esta função atinge seu máximo para o nodo vencedor, onde  $j = i$  e decai através de uma gaussiana com a distância lateral entre os vizinhos.

$$h_{jv(x)} = \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_v\|^2}{2\sigma^2(n)}\right), n = 0, 1, 2, \dots \quad (2.5)$$

onde  $\sigma(n)$  corresponde ao raio da função de vizinhança,  $\mathbf{r}_v \in \mathbb{R}_2$  e  $\mathbf{r}_j \in \mathbb{R}_2$  são as posições vetoriais do vencedor  $v$  e do seu vizinho  $j$  na grade. O valor de  $\sigma(n)$  decresce com o decorrer das iterações conforme a equação que segue:

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_2}\right), n = 0, 1, 2, \dots \quad (2.6)$$

onde  $\tau_2$  é a constante de tempo que ajusta a velocidade de decaimento.

### 2.2.2 Dimension Selective Self-Organizing Map - DSSOM

O DSSOM é um Mapa Auto-organizável proposto por [Bassani e Araujo \(2012\)](#) de topologia fixa e seletivo a dimensões de entrada, capaz de ajustar a relevância de cada uma das dimensões no cálculo da unidade vencedora para cada nodo na grade presente na unidade de saída. Permitindo que a relevância de algumas dimensões seja praticamente anulada para um dado nodo, fazendo com que estas dimensões não interfiram no agrupamento estabelecido por este. O ajuste destas relevâncias é feito em forma de pesos de maneira adaptativa durante o processo de treinamento. O DSSOM possibilita que mais de um nodo vença para um dado

estímulo de entrada, permitindo que nodos que reconheçam características distintas (ou seja, aplicam diferentes relevâncias para as dimensões de entrada) das consideradas pelo primeiro vencedor também possam agrupar esse estímulo.

Portanto, a cada nodo  $j$  presente na grade (camada de saída) é associado um vetor de ponderação  $\boldsymbol{\omega}_j = \{\omega_{ji}, i = 1 \dots m\}$ , onde  $m$  é o número de dimensões de entrada. Esse vetor indica a relevância de cada atributo de entrada para cada nodo e tem valores que variam entre  $[0, 1]$ . O processo de treinamento, também chamado de auto-organização, do mapa para um conjunto de padrões de entrada é detalhado no Alg. 1.

**Algoritmo 1:** Auto-organização no DSSOM

```

1 Inicializa os parâmetros  $\eta_0, \sigma_0, \tau_1, \tau_2, \beta, \varepsilon_p, t_{max}, n_{max}, k_{max}$ 
2 Inicializa os vetores de pesos  $\mathbf{w}_j$  aleatoriamente
3 Inicializa os vetores de médias  $\boldsymbol{\delta}_j \leftarrow 0$ 
4 para  $t \leftarrow 0$  até  $t_{max}$  faça
5   para  $n \leftarrow 0$  até  $n_{max}$  faça
6     Inicializa o vetor global de relevâncias  $\boldsymbol{\rho} \leftarrow 1$ 
7     Apresenta ao mapa um padrão  $\mathbf{x}$  escolhido aleatoriamente
8      $k = 1$ 
9     while  $\rho_{max} > \varepsilon_p$  e  $k \leq k_{max}$  do
10      se  $k = 1$  então
11        | Encontra o primeiro vencedor  $s_1$  (Eq. 2.7)
12      senão
13        | Encontra outro vencedor  $s_k$  (Eq. 2.14)
14      fim
15      Atualiza o vetor de médias  $\boldsymbol{\delta}_j$  do vencedor e dos seus vizinhos (Eq. 2.11)
16      Atualiza o vetor de relevâncias  $\boldsymbol{\omega}_j$  do vencedor e dos seus vizinhos
17        (Eq. 2.12)
18      Atualiza o vetor de pesos  $\mathbf{w}_j$  do vencedor e dos seus vizinhos (Eq. 2.3)
19      Atualiza o vetor global de relevâncias  $\boldsymbol{\rho}$  (Eq. 2.13)
20       $k = k + 1$ 
21    end
22  fim
23 fim

```

### 2.2.2.1 Escolha do Nodo Vencedor

O primeiro vencedor de uma competição é o nodo que apresenta a maior ativação para o padrão de entrada. Calculado conforme segue:

$$s_1(\mathbf{x}) = \arg \max_j [ac(D_\omega(\mathbf{x}, \mathbf{w}_j), Sum(\boldsymbol{\omega}_j))] \quad (2.7)$$

Definindo  $D_\omega(\mathbf{x}, \mathbf{w}_j)$  como a distância entre o centro do protótipo e o padrão de entrada,

ponderada pelo vetor de relevâncias, determinada por:

$$D_{\omega}(\mathbf{x}, \mathbf{w}_j) = \sum_{i=1}^m \omega_{ij}^2 (x_i - w_{ji})^2 \quad (2.8)$$

A ativação de um nodo é calculada por uma função da distância ponderada para o padrão de entrada,  $D_{\omega}(\mathbf{x}, \mathbf{w}_j)$ , e da soma dos componentes do seu vetor de relevâncias,  $Sum(\boldsymbol{\omega}_j)$ .

$$ac(D_{\omega}(\mathbf{x}, \mathbf{w}_j), Sum(\boldsymbol{\omega}_j)) = \frac{Sum(\boldsymbol{\omega}_j)}{D_{\omega}(\mathbf{x}, \mathbf{w}_j) + Sum(\boldsymbol{\omega}_j) + \varepsilon} \quad (2.9)$$

onde  $\varepsilon$  é um valor pequeno para evitar divisão por zero e  $Sum(\boldsymbol{\omega}_j)$  é dado por:

$$Sum(\boldsymbol{\omega}_j) = \sum_{i=1}^m \omega_{ji} \quad (2.10)$$

#### 2.2.2.2 Atualização do Vetor de Pesos

A atualização do nodo vencedor é feita de forma semelhante ao mapa de Kohonen (Seção 2.2.1), atualizando-se os pesos sinápticos do vencedor e seus vizinhos. Entretanto, com a adição da atualização também do vetor de ponderação  $\boldsymbol{\omega}_j$ . Para isto, cada nodo também conta com um vetor,  $\boldsymbol{\delta}_j$ , que estima a variação média entre os estímulos para os quais este nodo foi vencedor. Esta variação média é calculada através de uma média móvel controlada pelo parâmetro  $\beta$ , o qual indica a velocidade com que a média móvel se modifica. Este vetor é inicializado com zeros e atualizado conforme a equação que segue:

$$\boldsymbol{\delta}_j(n+1) = (1 - \beta)h_{jv(x)}\boldsymbol{\delta}_j(n) + \beta h_{jv(x)}|\mathbf{x} - \mathbf{w}_j(n)| \quad (2.11)$$

onde  $|\cdot|$  denota o valor absoluto resultante aplicado as componentes do vetor,  $\beta \in ]0, 1[$  e a função de vizinhança  $h_{jv(x)}$  é utilizada para que a atualização das distâncias dos vizinhos do nodo vencedor ocorra de forma proporcional a distância para o vencedor.

Após a atualização do vetor de variações médias, cada componente  $i$  do vetor de ponderação é atualizada conforme a Eq. 2.12.

$$\omega_{ji} = \begin{cases} 1 - \left( \frac{\delta_{ji}}{\delta_{jimax}} \right) & \text{se } \delta_{jimax} > 0 \\ 1 & \text{se } \delta_{jimax} = 0 \end{cases} \quad (2.12)$$

onde  $\delta_{jimax}$  é o componente de maior valor do vetor  $\boldsymbol{\delta}_j$ .

#### 2.2.2.3 Auto-Organização

O DSSOM permite que mais de um vetor seja vencedor de uma competição para um dado estímulo de entrada. Para isso foi introduzido um vetor de relevâncias global  $\boldsymbol{\rho} = \{\rho_i, i =$

$1\dots m\}$ . Este vetor penaliza as dimensões já consideradas por vencedores anteriores. Todos os componentes desse vetor são inicializados com um e essa inicialização é feita sempre que um novo padrão de entrada é apresentado. Quando um vencedor é encontrado, o vetor de relevâncias global é atualizado em função do vetor de relevâncias do vencedor anterior, conforme segue:

$$\rho_i = \rho_i(1 - \omega_{ki}) \text{ para } i = 0, 1, 2\dots m \quad (2.13)$$

onde  $k$  é o índice do nodo vencedor e  $\omega_{ki}$  é o  $i$ -ésimo componente do vetor de relevâncias do nodo vencedor anterior.

Fazendo com que as dimensões consideradas pelo nodo vencedor sejam menos relevantes nas próximas competições. Posteriormente, enquanto o número máximo de vencedores,  $k_{max}$ , não for atingido ou o valor do maior componente do vetor de relevâncias global,  $\rho_{max}$ , for maior do que um limiar  $\varepsilon_p$ , outro vencedor  $s_k$  é determinado. O novo vencedor é calculado de acordo com a Eq. 2.14, utilizando o vetor de relevâncias global ao invés do associado a cada nodo.

$$s_k(\mathbf{x}) = \arg \max_j [ac(D_\rho(\mathbf{x}, \mathbf{w}_j), Sum(\boldsymbol{\rho}_j))] \quad (2.14)$$

onde  $D_\rho(\mathbf{x}, \mathbf{w}_j)$  é calculado conforme segue:

$$D_\rho(\mathbf{x}, \mathbf{w}_j) = \sum_{i=1}^m \rho_i^2 (x_i - w_{ji})^2 \quad (2.15)$$

onde  $\rho_i$  é o  $i$ -ésimo componente do vetor de relevâncias global, atualizado de acordo com a Eq. 2.13, após cada novo nodo vencedor ser determinado.

#### 2.2.2.4 Agrupamento

Após a auto-organização, o mapa pode ser utilizado para agrupar padrões de entrada. Cada nodo no mapa é associado a um índice que representa um grupo e os padrões de entrada são associados a um ou mais índices de agrupamento. O número máximo de grupos aos quais um padrão pode pertencer é delimitado pelo parâmetro  $k_{max}$ . Se a ativação produzida pelo padrão de entrada para o primeiro vencedor for menor do que um limiar  $th$ , o padrão é considerado ruído. O Alg. 2 detalha o processo de agrupamento.

### 2.2.3 Local Adaptive Receptive Field Dimension Selective Self-Organizing Map - LARFDSSOM

O LARFDSSOM é um Mapa Auto-organizável de topologia variante no tempo proposto por [Bassani e Araujo \(2015\)](#), que assim como o DSSOM, aplica relevâncias distintas as dimensões de entrada e possibilita que mais de um nodo vença para um dado estímulo de entrada. No LARFDSSOM, a vizinhança de nodos na grade é formada por nodos que levam em consideração

**Algoritmo 2:** Agrupamento no DSSOM

```

1 para cada padrão de entrada (x) faça
2   Inicializa o vetor global de relevâncias  $\rho \leftarrow 1$ 
3   Apresenta x para o DSSOM
4    $k = 1$ 
5   while  $\rho_{max} > \varepsilon_p$  e  $k \leq k_{max}$  do
6     se  $k = 1$  então
7       Encontra o primeiro vencedor  $s_1$  (Eq.2.7)
8       se  $ac(s_1) < th$  então
9         Associa x ao grupo de ruído
10        pare
11       fim
12     senão
13       Encontra outro vencedor  $s_k$  (Eq.2.14)
14     fim
15     Associa x ao grupo do nodo vencedor
16     Atualiza o vetor global de relevâncias  $\rho$  (Eq.2.13)
17      $k = k + 1$ 
18   end
19 fim

```

um conjunto similar de atributos. Os passos de competição e cooperação são repetidos por um número determinado de épocas. Durante esse processo, os nodos que não vencerem para uma proporção mínima de padrões são removidos do mapa e sempre que nenhum nodo presente atingir um valor mínimo de ativação para um dado estímulo de entrada, um novo nodo é adicionado para satisfazer aquele estímulo.

Da mesma forma que no DSSOM, cada nodo  $j$  presente no mapa (na camada de saída) esta associado a um vetor de relevâncias  $\omega_j = \{\omega_{ji}, i = 1..m\}$ , um vetor de variações médias das distâncias observadas  $\delta_j = \{\delta_{ji}, i = 1..m\}$  e um vetor de pesos sinápticos, renomeados para centros,  $\mathbf{c}_j = \{c_{ji}, i = 1..m\}$ , onde  $m$  é o número de dimensões de entrada. Quando um novo nodo é adicionado ao mapa, seu centro  $\mathbf{c}_j$  é inicializado na posição do último padrão de entrada, os vetores  $\omega_j$  e  $\delta_j$  tem todos seus componentes inicializados com valores um e zero, respectivamente. Esses valores são atualizados após o passo de competição.

O processo de treinamento do mapa para um determinado conjunto de padrões é realizado em duas fases denominadas de auto-organização (Alg. 3) e convergência (Alg. 4). Por fim, com o mapa já treinado os padrões são agrupados na terceira e última fase, detalhada no Alg. 5.

**Algoritmo 3:** Auto-Organização do LARFDSSOM

```

1 Inicializa os parâmetros  $a_t, l_p, e_b, e_n, \beta, maxcomp, s, t_{max}, N_{max}$ ;
2 Inicializa o mapa com um nodo com  $\mathbf{c}_j$  igual ao primeiro padrão de entrada,  $\delta_j \leftarrow 0$ ,
    $\omega_j \leftarrow 1$  e  $wins_j \leftarrow 0$ ;
3 Inicializa a variável  $nwins \leftarrow 1$ ;
4 para  $t \leftarrow 0$  até  $t_{max}$  faça
5   Apresenta randomicamente os padrões de entrada  $\mathbf{x}$  ao mapa
6   Computa a ativação de todos os nodos (Eq. 2.17)
7   Encontra o nodo vencedor  $s$  com maior valor de ativação ( $a_s$ ) (Eq. 2.16)
8   se  $a_s < a_t$  e  $N < N_{max}$  então
9     Cria um novo nodo  $j$  e atribui:  $\mathbf{c}_j \leftarrow \mathbf{x}$ ,  $\delta_j \leftarrow 0$ ,  $\omega_j \leftarrow 1$  e
        $wins_j \leftarrow l_p \times nwins$ 
10    Conecta  $j$  aos outros nodos através da Eq. 2.22
11  senão
12    Atualiza o vetor de distâncias  $\delta_s$  do nodo vencedor e de seus vizinhos
       (Eq. 2.20)
13    Atualiza o vetor de relevâncias  $\omega_s$  do nodo vencedor e de seus vizinhos
       (Eq. 2.21)
14    Atualiza o vetor de centro  $\mathbf{c}_s$  do nodo vencedor e de seus vizinhos (Eq. 2.19)
15    Atribui  $wins_s \leftarrow wins_s + 1$ 
16  fim
17  se  $nwins = maxcomp$  então
18    Remove os nodos com  $wins_j < l_p \times maxcomp$ 
19    Atualiza as conexões restantes através da Eq. 2.22
20    Reseta o número de vitórias dos nodos restantes:  $wins_j \leftarrow 0$ 
21     $nwins \leftarrow 0$ 
22  fim
23   $nwins \leftarrow nwins + 1$ 
24 fim
25 Executa a fase de convergência (Alg. 4)

```

## 2.2.3.1 Inserção de Nodos e Competição

O vencedor de uma competição  $s(\mathbf{x})$ , é o nodo que apresenta maior ativação,  $ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \omega_j)$ , para um padrão de entrada.

$$s(\mathbf{x}) = \arg \max_j [ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \omega_j)] \quad (2.16)$$

A ativação  $ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \omega_j)$  de cada nodo é calculada por uma função de base radial da distância ponderada,  $D_\omega(\mathbf{x}, \mathbf{c}_j)$ , com o campo receptivo ajustado em função da norma do vetor de relevâncias. Com isso, a ativação cresce com a norma do vetor de relevâncias, fazendo com que nodos que levam em consideração mais fortemente um maior número de dimensões tenham

vantagem na competição.

$$ac(D_{\omega}(\mathbf{x}, \mathbf{c}_j), \boldsymbol{\omega}_j) = \frac{1}{1 + \frac{D_{\omega}(\mathbf{x}, \mathbf{c}_j)}{(\|\boldsymbol{\omega}_j\|^2 + \varepsilon)}} \quad (2.17)$$

onde  $\|\boldsymbol{\omega}_j\|$  é a norma do vetor de relevâncias,  $\varepsilon$  é um valor bem pequeno para evitar divisão por zero,  $D_{\omega}(\mathbf{x}, \mathbf{c}_j)$  é a distância entre o centro do protótipo e o padrão de entrada, ponderada pelo vetor de relevâncias, dada por:

$$D_{\omega}(\mathbf{x}, \mathbf{c}_j) = \sqrt{\sum_{i=1}^m \omega_{ij}(x_i - c_{ji})^2} \quad (2.18)$$

O nodo que for vencedor de uma competição terá seus pesos atualizados, assim como seus vizinhos, pelo procedimento descrito na próxima seção. Entretanto, sempre que todos os nodos no mapa apresentarem ativação inferior a um limiar de ativação,  $a_t$ , um novo nodo é adicionado na posição do padrão de entrada e nenhum nodo é atualizado. Também é possível determinar um limite máximo,  $N_{max}$ , para o número de nodos no mapa. Se este limite for alcançado, novos nodos não serão mais inseridos, mas os nodos existentes continuam sendo atualizados.

### 2.2.3.2 Atualização do Nodo Vencedor e Seus Vizinhos

A atualização do nodo vencedor e de seus vizinhos, é feita levando em consideração duas taxas de aprendizagem:  $e_b \in ]0, 1[$  é a taxa utilizada na atualização do vencedor e  $e_n \in ]0, e_b[$  é a taxa utilizada na atualização dos vizinhos. Desta maneira, os centros são atualizados conforme a equação que segue:

$$\mathbf{c}_j(n+1) = \mathbf{c}_j(n) + e(\mathbf{x} - \mathbf{c}_j(n)) \quad (2.19)$$

onde  $e$  é a taxa de aprendizagem dada por:  $e = e_b$  se  $j$  é o nodo vencedor, ou  $e = e_n$  se  $j$  é um vizinho do nodo vencedor.

O vetor de relevâncias também precisa ser atualizado, entretanto primeiro é estimada a distância média do nodo para os padrões de entrada por ele agrupados. Logo, a cada vitória o vetor de distâncias é atualizado através da média móvel das distâncias observadas, conforme a seguinte função:

$$\boldsymbol{\delta}_j(n+1) = (1 - e\beta)\boldsymbol{\delta}_j(n) + e\beta(|\mathbf{x} - \mathbf{c}_j(n)|) \quad (2.20)$$

onde  $\beta \in ]0, 1[$  controla a velocidade de mudança da média móvel, o operador  $|\cdot|$  denota o valor absoluto da diferença entre os vetores e  $e$  é a taxa de aprendizagem dada por:  $e = e_b$  se  $j$  é o nodo vencedor, ou  $e = e_n$  se  $j$  é um vizinho do nodo vencedor.

Com o vetor de distâncias atualizado, cada componente  $i$  do vetor de relevâncias  $\boldsymbol{\omega}_j$  é

calculada por uma função logística inversa das distâncias  $\delta_{ji}$ , definida por:

$$\omega_{ji} = \begin{cases} \frac{1}{1 + \exp\left(\frac{\delta_{ji} - \delta_{jimean}}{s(\delta_{jimax} - \delta_{jimin})}\right)} & \text{se } \delta_{jimin} \neq \delta_{jimax} \\ 1 & \text{caso contrário} \end{cases} \quad (2.21)$$

onde  $\delta_{jimin}$ ,  $\delta_{jimax}$  e  $\delta_{jimean}$  são respectivamente o valor mínimo, o valor máximo e a média dos  $i$  componentes do vetor de distâncias  $\delta_j$ . As relevâncias estão dentro do intervalo  $[0,1]$  e tendem a zero para dimensões com variabilidade próximas de  $\delta_{jimax}$ . A variável  $s > 0$  controla a suavidade da função logística.

### 2.2.3.3 Remoção de Nodos

Após um número determinado,  $maxcomp$ , de competições, um *reset* ocorre no mapa. Nele são removidos os nodos que apresentarem um número de vitórias abaixo do limiar  $lp \times maxcomp$ , onde  $lp$  é um parâmetro que determina a porcentagem mínima de vitórias necessária para um nodo permanecer no mapa. Esse número de vitórias é contabilizado por uma variável,  $wins_j$ , que é associada a cada nodo  $j$  presente no mapa, ela conta o número de vitórias do nodo desde o último *reset*. Após cada *reset*, o número de vitórias dos nodos que não foram removidos é reiniciado para zero. Adicionalmente, para evitar a remoção prematura de nodos, quando um novo nodo é adicionado seu número de vitórias é iniciado com  $lp \times nwins$ , onde  $nwins$  é o número de competições que ocorreram desde o último *reset*.

### 2.2.3.4 Atualização da Vizinhança

A existência de uma conexão entre nodos significa que eles agrupam padrões em subespaços semelhantes. Logo, a similaridade entre os vetores de relevâncias entre cada par de nodos no mapa determina a existência dessas conexões, sendo expressa pela seguinte equação:

$$\text{nodos } i \text{ e } j \text{ são } \begin{cases} \text{conectados} & \text{se } \|\omega_i - \omega_j\| < c\sqrt{m} \\ \text{desconectados} & \text{caso contrário} \end{cases} \quad (2.22)$$

onde  $c$  é um valor que define um limiar de conexão e  $m$  é o número de dimensões de entrada. Como a relevância varia entre zero e um,  $\sqrt{m}$  é o valor máximo para a norma das distâncias entre dois vetores de relevância. Com  $c = 1/2$  tem-se um limiar de conexão entre dois nodos onde apenas pares de nodos com diferenças de relevâncias que estejam abaixo da metade do máximo possível serão conectados.

A cada *reset* as conexões entre todos os nodos são atualizadas, assim como, quando um novo nodo é adicionado ao mapa, as conexões entre o mesmo e todos os outros são atualizadas.

## 2.2.3.5 Convergência

Ao fim da fase de auto-organização, o mapa entra em fase de convergência. Nesta fase o mapa continua sua auto-organização, mas sem a adição de novos nodos. Os nodos existentes no mapa continuam sendo treinados com o conjunto de padrões de entrada e os que não atingirem uma porcentagem mínima de vitórias,  $lp \times maxcomp$ , são removidos. Esse processo continua até que não sejam removidos mais nodos. O Alg. 4 detalha o funcionamento desta fase, nele  $N_{max}$  é ajustado a cada iteração com o número atual de nodos no mapa.

**Algoritmo 4:** Fase de Convergência do LARFDSSOM

```

1 enquanto Verdadeiro faça
2    $N_{max} \leftarrow N$ 
3   Remove os nodos com  $wins_j < lp \times maxcomp$ 
4   se ( $N = N_{max}$  ou  $N = 1$ ) então
5     retorna
6   Atualiza as conexões de todos os nodos através da Eq. 2.22
7   Reseta o número de vitórias de todos os nodos:  $wins_j \leftarrow 0$ 
8   para  $t \leftarrow 0$  até  $t_{max}$  faça
9     Apresenta um padrão  $x$  randomicamente escolhido ao mapa
10    Computa a ativação de todos os nodos (Eq. 2.17)
11    Encontra o nodo vencedor  $s$  com a maior ativação (Eq. 2.16)
12    Atualiza o vetor de médias  $\delta_j$  do vencedor e dos seus vizinhos (Eq. 2.20)
13    Atualiza o vetor de relevâncias  $\omega_j$  do vencedor e dos seus vizinhos
        (Eq. 2.21)
14    Atualiza o vetor de centros  $c_j$  do vencedor e dos seus vizinhos (Eq. 2.19)
15     $wins_s \leftarrow wins_s + 1$ 
16 fim
17 fim

```

## 2.2.3.6 Agrupamento

Após o processo de treinamento dado pelas fases de auto-organização e convergência, o mapa está pronto para ser utilizado para agrupar o conjunto de padrões de entrada. Neste modelo, um padrão de entrada pode pertencer a vários grupos, atendendo ao requisito de agrupamento em subespaços. Desta forma, a cada nodo presente no mapa é associado um índice representando um grupo, e um padrão de entrada pode ser associado a um ou mais desses grupos, desde que eles apresentem ativação acima do limiar,  $a_t$ , para este padrão. Entretanto, também pode ser realizado o agrupamento tradicional, denominado de agrupamento projetivo, no qual cada padrão de entrada é associado a apenas o agrupamento do nodo de maior ativação.

Em ambos os tipos de agrupamento, caso nenhum dos nodos produza ativação igual ou superior ao limiar para um dado padrão de entrada, este padrão é atribuído ao conjunto de *outliers*. O Alg. 5 descreve o processo de agrupamento.

**Algoritmo 5:** Agrupamento no LARFDSSOM

```

1 para cada padrão de entrada ( $\mathbf{x}$ ) no conjunto de dados faça
2   Apresenta  $\mathbf{x}$  para o LARFDSSOM
3   Calcula a ativação de todos os nodos (Eq. 2.17)
4   Encontra o nodo vencedor  $s$  com a maior ativação ( $a_s$ ) (Eq. 2.16)
5   se  $a_s \geq a_t$  então
6     repita
7       Atribui  $\mathbf{x}$  para o grupo com índice do nodo vencedor  $s$ 
8       se agrupamento projetivo então
9         pare
10        Encontra o próximo nodo no mapa com a maior ativação  $a_s$ 
11        desconsiderando os vencedores anteriores
12      até  $a_s < a_t$ 
13    senão
14      Atribui  $\mathbf{x}$  para o conjunto de outliers
15  fim

```

## 2.3 Conclusão

Este capítulo introduziu algumas características e trabalhos atuais na literatura sobre Mapeamento Semântico na Seção 2.1 e detalhou o funcionamento de alguns tipos de Mapas Auto-organizáveis na Seção 2.2. Conceitos que serviram de base para a abordagem proposta na presente dissertação (exposta em detalhes no próximo capítulo), uma abordagem que faz a utilização de redes neurais não-supervisionadas de aprendizagem competitiva baseadas em Mapas Auto-organizáveis (SOM, do inglês *Self Organizing Maps*) (Kohonen, 1982) para a atividade de mapeamento semântico, assim como a categorização dos tipos de lugares mapeados associada.

Desta forma, o Mapa Auto-organizável que serviu de base foi uma variante do modelo SOM clássico com topologia variante no tempo, o LARFDSSOM (Bassani e Araujo, 2015). Embora o mesmo apresente características de aprendizagem incremental, o treinamento realizado é *offline*, por isso, na presente dissertação é proposta uma modificação do mesmo com treinamento *online* para a atividade de categorização dos lugares mapeados, conforme é descrito no próximo capítulo.

# 3

## Abordagem Desenvolvida

Este capítulo tem por objetivo apresentar a abordagem desenvolvida neste trabalho, primeiramente expondo sua arquitetura e descrição na Seção 3.1, em seguida detalhando o funcionamento dos módulos desenvolvidos, SEMMAP (Seção 3.2) e OLARFDSSOM (Seção 3.3). Por fim, a conclusão do capítulo é feita na Seção 3.4.

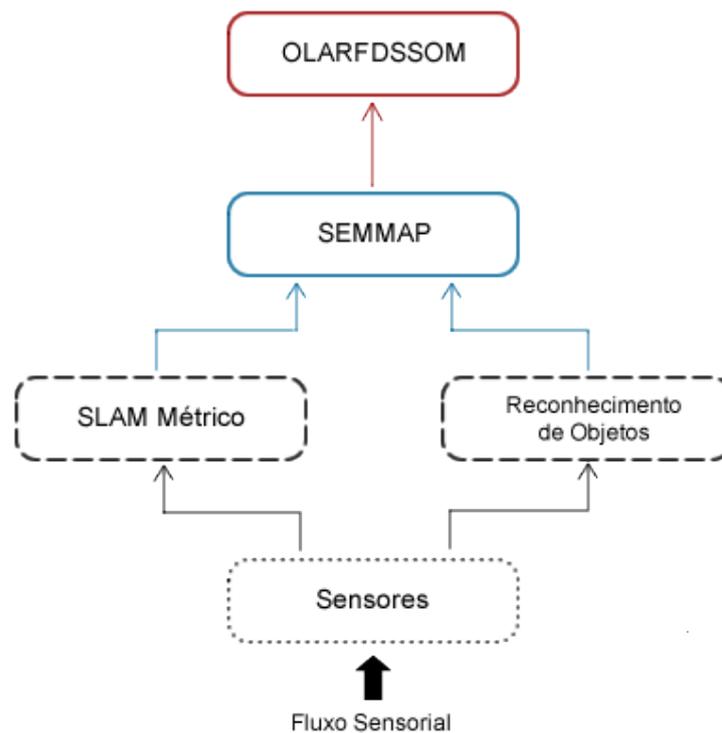
### 3.1 Descrição da Abordagem

A abordagem proposta nesta dissertação realiza a aquisição de mapas semânticos sem nenhum tipo de supervisão e de forma incremental (ou seja, progressivamente enquanto opera), não sendo necessário conhecimento prévio do ambiente no qual estiver inserida. O tipo de ambiente em que a abordagem desenvolvida pode atuar, interno (*indoor*) ou externo (*outdoor*), não foi restrito em sua concepção, tanto que um método de detecção de portas não foi incluso. Entretanto, no atual estágio da pesquisa, a abordagem foi experimentada apenas com uma base de dados de ambientes *indoor* (conforme é exposto posteriormente no Capítulo 4). Em suma, a abordagem coleta dados sensoriais enquanto navega em um determinado ambiente e constrói progressivamente um mapa topológico enriquecido com informações semânticas, assim como um modelo com aprendizagem incremental que realiza a categorização dos diferentes tipos de lugares mapeados.

Deste modo, a abordagem é composta de dois módulos principais, os quais realizam a aquisição de informações semânticas em forma de mapa e categorização de lugares mapeados, com aprendizagem incremental e não-supervisionada. São eles:

- SEMMAP: Um método que constrói um mapa topológico do ambiente incrementalmente, enriquecido com informações semânticas, explorando na atual fase da pesquisa apenas objetos reconhecidos no ambiente como pistas semânticas únicas;
- OLARFDSSOM: Uma modificação do LARFDSSOM (Bassani e Araujo, 2015) como categorizador de aprendizagem não-supervisionada, topologia variante no tempo e treinamento online.

Estes módulos são auxiliados por outros dois módulos que não foram desenvolvidos neste trabalho. O de SLAM Métrico, o qual realiza simultaneamente o mapeamento métrico e a localização do agente no ambiente no qual ele está inserido. O outro módulo realiza o reconhecimento de um conjunto pré-definido de objetos no ambiente conforme o agente se desloca. Objetos foram escolhidos como sendo as únicas pistas semânticas determinantes a serem utilizadas no atual estágio da pesquisa, entretanto a abordagem contempla a utilização simultânea de múltiplas propriedades semânticas do ambiente como pistas e pretende-se expandi-la desta forma em trabalhos futuros. A arquitetura exibida na Figura 3.1 ilustra a relação entre os módulos, expressa em conjunto com seu fluxo de funcionamento. Os sensores presentes na arquitetura fornecem dados aos módulos auxiliares em intervalos de tempo  $t$ , caracterizando o fluxo sensorial. Estes sensores podem variar nos mais diversos tipos (como *laser range scanner*, câmera regular, câmera omnidirecional e etc), a depender dos métodos utilizados pelos módulos auxiliares.



**Figura 3.1:** Arquitetura da abordagem proposta. Os componentes com contorno tracejado de cor preta não foram implementados neste trabalho.

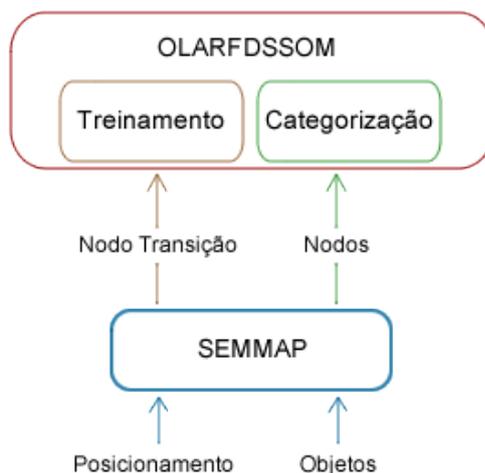
Abaixo, são detalhadas brevemente as atividades desempenhadas por cada um dos módulos da arquitetura no fluxo de funcionamento.

**SLAM Métrico:** Algoritmo de SLAM métrico padrão que recebe como entrada dados sensoriais, realiza mapeamento e localização simultâneas do agente no ambiente e fornece como saída seu posicionamento em coordenadas ao SEMMAP;

**Reconhecimento de Objetos:** Recebe como entrada dados sensoriais (em geral, imagens) e reconhece um conjunto de objetos pré-definidos em valores de certeza variando entre  $[0,1]$ , onde 1 representa certeza máxima no reconhecimento de um objeto em questão. Posteriormente, os valores em forma de vetor são fornecidos ao SEMMAP simultaneamente ao módulo de SLAM;

**SEMMAP:** Método de mapeamento semântico que recebe como entrada posicionamento e objetos reconhecidos pelo agente (em tempo real no fluxo de funcionamento) e constrói um mapa topológico do ambiente incrementalmente, enriquecido com informações semânticas que são constantemente enviadas ao OLARFDSSOM para treinamento, conforme o agente transita entre nodos do mapa topológico (ou seja, se desloca pelo ambiente). As informações enviadas são as presentes no último nodo visitado pelo agente no momento de transição;

**OLARFDSSOM:** Mapa auto-organizável de aprendizagem não-supervisionada que recebe como entrada padrões para treinamento (informações semânticas, neste trabalho caracterizadas por um vetor de objetos) incrementalmente fornecidos pelo SEMMAP e os organiza em sua estrutura de aprendizagem. Além disso, o mesmo pode fornecer como saída a atual categoria (ou grupo) de lugar para qualquer nodo do mapa semântico (SEMMAP) a partir de suas propriedades semânticas incorporadas, sempre que necessário, a depender da aplicação. A Figura 3.2 ilustra as possibilidades.



**Figura 3.2:** Interação entre SEMMAP e OLARFDSSOM em dois momentos, em marrom o momento de treinamento realizado sempre que o agente transita entre nodos do SEMMAP e em verde o momento de categorização realizado a qualquer instante que se necessite saber a categoria atual para um determinado conjunto de nodos no SEMMAP.

A seguir são detalhados o funcionamento dos módulos SEMMAP e OLARFDSSOM. Como os módulos de SLAM Métrico e Reconhecimento de Objetos não foram desenvolvidos neste trabalho, os mesmos não são detalhados.

## 3.2 SEMMAP

O SEMMAP é um método de mapeamento semântico de larga escala que cria mapas topológicos de ambientes enriquecidos com informações semânticas sobre os mesmos. A representação topológica é criada em forma de grafo através de um método baseado em Mapas Auto-organizáveis como uma extensão a um método de SLAM (do inglês, *Simultaneous Localization and Mapping*) métrico padrão. Adicionalmente, propriedades semânticas são inferidas externamente e incorporadas ao mapa topológico.

O mapa é criado incrementalmente, conforme um agente se movimenta no ambiente. Nós presentes no mapa competem para agrupar as informações de posição fornecidas pelo método de SLAM métrico e o vencedor de uma competição é o mais ativo de acordo com uma função de base radial que considera apenas a localização espacial, ou seja, o nó vencedor será aquele que a cobertura de posicionamento mais se aproximar da nova posição fornecida. Sempre que nenhum nó atinge uma ativação mínima, um novo nó é criado no mapa para satisfazer a nova posição de entrada. A vizinhança dos nós é formada pela transição do agente entre nós no decorrer da sua movimentação.

Apesar das propriedades semânticas não serem consideradas durante a competição, os nós vencedores incorporam essas propriedades presentes nas regiões cobertas por eles. No atual estágio da pesquisa apenas objetos reconhecidos pelo agente foram incorporados como propriedade semântica. As propriedades semânticas do último nó visitado são enviadas ao OLARFDSSOM para treinamento no instante em que o agente realiza uma transição para outro nó. Os nós também podem ser associados a categorias de lugar, a qualquer momento, pelo OLARFDSSOM, através das propriedades semânticas armazenadas neles.

Os processos descritos acima são apresentados com mais detalhes a seguir, assim como são fornecidas as equações utilizadas e o algoritmo da operação de mapeamento (Alg. 6).

### 3.2.1 Competição

Os padrões de entrada apresentados ao mapa em instantes de tempo são fornecidos em conjunto por métodos externos associados ao agente, de forma que:  $\mathbf{p} = \{p_i, i = 1 \dots m\}$ , é um vetor que representa a posição atual do robô no ambiente fornecida por um método de SLAM métrico, onde  $m$  é o número de coordenadas utilizadas pelo método (geralmente em um plano horizontal  $x$  e  $y$ );  $\mathbf{r} = \{r_i, i = 1 \dots n\}$ , é o vetor de objetos reconhecidos pelo método de reconhecimento de objetos do agente com valores de certeza entre  $[0,1]$ , onde  $n$  é o número de objetos que podem ser reconhecidos.

O mapa é representado por um grafo  $G = (\mathbf{V}, \mathbf{E})$ , onde  $\mathbf{V} = \{v_j, j = 1 \dots k\}$  é o vetor de nós (também denominados de vértices) que representam regiões da topologia do mapa e  $\mathbf{E} = \{e_i, i = 1 \dots l\}$  representa um vetor de relações de transitividade observadas entre nós. Cada nó  $j$  no mapa está associado com três vetores:  $\mathbf{c}_j = \{c_{ji}, i = 1 \dots m\}$  representa a posição

(centro) do nodo  $j$  na topologia;  $\mathbf{o}_j = \{o_{ji}, i = 1 \dots n\}$  é o vetor de objetos reconhecidos na região coberta pelo nodo, com valores de certeza entre  $[0,1]$ ;  $\phi_j = \{\phi_{ji}, i = 1 \dots n\}$  é o vetor que acumula os valores dos objetos reconhecidos em diferentes perspectivas pelo agente, com valores entre  $[0, s_t]$ , onde  $s_t$  é um parâmetro que determina o limite superior de acumulação.

Quando um novo nodo é criado, seu centro  $\mathbf{c}_j$  é inicializado na última posição  $\mathbf{p}$  apresentada ao mapa pelo método de SLAM. O vetor de somas  $\phi_j$  é inicializado com os valores do último vetor apresentado de objetos reconhecidos  $\mathbf{r}$  e todos os componentes do vetor de objetos  $\mathbf{o}_j$  são inicializados de acordo com a Eq. 3.6, descrita na próxima subseção.

O vencedor da competição,  $s(\mathbf{p})$ , é o nodo que apresenta a maior ativação para a atual posição do agente apresentada:

$$s(\mathbf{p}) = \arg \max_j [ac(D(\mathbf{p}, \mathbf{c}_j))] \quad (3.1)$$

onde  $\mathbf{p}$  é o vetor de tamanho  $m$  que representa a atual posição do agente.

A ativação de um nodo,  $ac(D(\mathbf{p}, \mathbf{c}_j))$  é calculada por uma função da distância euclidiana dos vetores:

$$ac(D(\mathbf{p}, \mathbf{c}_j)) = \frac{1}{1 + D(\mathbf{p}, \mathbf{c}_j)} \quad (3.2)$$

onde  $D(\mathbf{p}, \mathbf{c}_j)$  é calculado como uma distância euclidiana tradicional determinada por:

$$D(\mathbf{p}, \mathbf{c}_j) = \sqrt{\sum_{i=1}^m (p_i - c_{ji})^2} \quad (3.3)$$

Se em uma competição, um limiar de ativação  $a_t$  não for atingindo pelo vencedor, um novo nodo é inserido ao mapa na posição do agente provida. Caso contrário, o nodo vencedor é atualizado conforme descrito na subseção a seguir e nas linhas 14 a 16 do Alg. 6.

### 3.2.2 Atualização de Nodo Vencedor

A atualização da posição central do nodo vencedor é feita levando em consideração a taxa de aprendizagem  $e \in ]0, 1[$  e é calculada conforme a equação que segue:

$$\mathbf{c}_j(n+1) = \mathbf{c}_j(n) + e(\mathbf{p} - \mathbf{c}_j(n)) \quad (3.4)$$

Os valores do vetor de objetos  $\mathbf{o}_j$  são estimados após a atualização do vetor de somas de objetos  $\phi_j$ . Esse vetor acumula os valores de reconhecimento de objetos pelo agente na região coberta pelo nodo, visando mitigar o problema de ocultação de objetos pelo reconhecimento em diferentes perspectivas da região (seja o reconhecimento feito através imagens, sensores de distância ou qualquer outra fonte de dados). Desta forma, os  $n$  objetos reconhecidos, com valores de certeza entre  $[0,1]$ , são consolidados no vetor  $\mathbf{r}$  e cada componente  $i$  do vetor de somas  $\phi_j$  é

**Algoritmo 6:** Processo de Mapeamento Semântico do SEMMAP

```

1 Inicializa os parâmetros  $a_t$ ,  $e$ ,  $s_t$ ;
2 Inicializa o mapa com um nodo  $j$  com  $\mathbf{c}_j$  igual a primeira posição  $\mathbf{p}$  apresentada,  $\phi_j$ 
   igual ao primeiro vetor apresentado de objetos reconhecidos  $\mathbf{r}$  e  $\mathbf{o}_j$  calculado de
   acordo com a Eq. 3.6;
3 Atribui  $u \leftarrow j$ , como o último nodo vencedor;
4 para cada padrão de entrada  $(\mathbf{p}, \mathbf{r})$  faça
5     Apresenta ao mapa a posição  $\mathbf{p}$  do agente e objetos  $\mathbf{r}$  reconhecidos pelo mesmo
6     Computa a ativação de todos os nodos (Eq. 3.2)
7     Encontra o nodo vencedor  $v$  com maior valor de ativação ( $a_v$ ) (Eq. 3.1)
8     se  $a_v < a_t$  então
9         Cria um novo nodo  $j$  e atribui:  $\mathbf{c}_j \leftarrow \mathbf{p}$ ,  $\phi_j \leftarrow \mathbf{r}$ ,  $\mathbf{o}_j$  calculado de acordo com
           a Eq. 3.6
10        Conecta  $j$  a  $u$ 
11        Envia o vetor  $\mathbf{o}_u$  do nodo  $u$  para treinamento no OLARFDSSOM
12        Substitui  $u \leftarrow j$ , como o último nodo vencedor
13    senão
14        Atualiza o vetor de somas  $\phi_v$  do nodo vencedor (Eq. 3.5)
15        Atualiza o vetor de objetos  $\mathbf{o}_v$  do nodo vencedor (Eq. 3.6)
16        Atualiza o vetor de centro  $\mathbf{c}_v$  do nodo vencedor (Eq. 3.4)
17        se  $v \neq u$  então
18            Conecta  $v$  a  $u$ 
19            Envia o vetor  $\mathbf{o}_u$  do nodo  $u$  para treinamento no OLARFDSSOM
20        Substitui  $u \leftarrow v$ , como o último nodo vencedor
21    fim
22 fim

```

atualizada numa soma limitada de forma superior pela variável  $s_t$ , calculada conforme segue:

$$\phi_{ji} = \begin{cases} s_t & \text{se } \phi_{ji} > s_t \\ \phi_{ji} = \phi_{ji} + r_i & \text{caso contrário} \end{cases} \quad (3.5)$$

Com o vetor de somas  $\phi_j$  atualizado, cada componente  $i$  do vetor de objetos  $\mathbf{o}_j$  é calculada com uma função logarítmica das somas  $\phi_{ji}$  determinada por:

$$o_{ji} = \log_{1+s_t}(1 + \phi_{ji}) \quad (3.6)$$

onde  $s_t$  é a variável que ao mesmo tempo que limita o valor de soma, controla a base do logaritmo, de forma que os valores resultantes de  $o_{ji}$  fiquem sempre entre  $[0,1]$ .

Da forma como está colocada, a Eq. 3.5 faz com que o método seja incapaz de perceber a retirada de objetos em ambientes já mapeados. Melhorias nesse sentido estão entre as atividades a serem realizadas em trabalhos futuros.

### 3.2.3 Transição entre Nodos

A conexão entre nodos é determinada no momento em que um nodo  $i$  vencedor de uma competição é diferente do nodo  $j$  vencedor na competição anterior, ou seja, no instante em que é caracterizada transição entre os nodos  $i$  e  $j$ . O mesmo acontece em caso de um novo nodo ser adicionado ao mapa, este nodo e o vencedor da competição anterior terão uma conexão realizada entre eles determinada pela transição (linhas 10 e 17-18 do Alg. 6).

Nesse contexto, enquanto o agente se movimenta pelo ambiente, sempre que ocorrer uma transição de um nodo  $j$  para um nodo  $i$ , os dados atuais do vetor de objetos  $\mathbf{o}_j$  associado ao nodo  $j$  são enviados ao OLARFDSSOM para treinamento, conforme descrito nas linhas 11 e 19 do Alg. 6. Este momento de envio foi escolhido por se entender que ao ser realizada a transição entre nodos, o nodo anteriormente visitado já teria tido a oportunidade de acumular uma quantidade suficiente de informações semânticas sobre a região, de forma a ser útil ao OLARFDSSOM no seu processo de aprendizagem.

## 3.3 OLARFDSSOM

O OLARFDSSOM (*Online Local Adaptive Receptive Field Dimension Selective Self-Organizing Map*) é um mapa auto-organizável com estrutura variante no tempo e treinamento realizado de forma online, originado do LARFDSSOM (Bassani e Araujo, 2015). Assim como no LARFDSSOM original, os nodos podem aplicar relevâncias distintas para as dimensões de entrada e o campo receptivo dos nodos é adaptado durante o processo de auto-organização, o que o permite selecionar automaticamente, atribuindo pesos, as propriedades semânticas que são mais relevantes para os agrupamentos (categorias) criados, evitando também, problemas com a possível alta dimensionalidade dos dados. Entretanto, diferentemente da versão original, a operação do OLARFDSSOM não é realizada em fases. Suas atividades são consolidadas em fase única composta de dois processos: o processo de treinamento, chamado de auto-organização (Alg. 7), o qual é realizado conforme padrões são apresentados ao mapa; e o processo de agrupamento (Alg. 8), realizado com o atual estado de aprendizagem do mapa.

Na auto-organização, os nodos competem para agrupar os padrões de entrada apresentados e o vencedor de uma competição é o mais ativo de acordo com uma função de base radial com o campo receptivo ajustado em função da variância local dos padrões de entrada. Esses nodos são conectados a outros nodos que levam em consideração subconjuntos de atributos similares dos vetores de entrada, formando uma vizinhança. Durante o processo de aprendizagem online os nodos que não vencerem para uma proporção mínima de padrões são periodicamente removidos do mapa e sempre que nenhum nodo atinge um limiar mínimo de ativação em uma competição, um novo nodo é adicionado ao mapa para agrupar o padrão desta competição.

A qualquer momento, paralelamente à auto-organização, o processo de agrupamento pode ser realizado utilizando o mapa em seu estado atual. No entanto, no contexto deste trabalho,

o mapa foi utilizado apenas para agrupamento projetivo, no qual se associa um padrão de entrada apenas ao grupo (ou categoria) do nodo de maior ativação. Entretanto, assim como sua versão original, o mesmo pode ser facilmente modificado para agrupar um padrão em vários grupos com nível de ativação igual ou superior a um limiar mínimo, caracterizando o agrupamento em subespaços.

A seguir os processos brevemente descritos acima, são apresentados com mais detalhes, assim como são fornecidas as equações e o algoritmo de todo o procedimento.

**Algoritmo 7: Auto-Organização do OLARFDSSOM**

```

1 Inicializa os parâmetros  $a_t, l_p, e_b, e_n, \beta, maxcomp, s, c, N_{max}$ ;
2 Inicializa o mapa com um nodo com  $c_j$  igual ao primeiro padrão de entrada,  $\delta_j \leftarrow 0$ ,
    $\omega_j \leftarrow 1$  e  $wins_j \leftarrow 0$ ;
3 Inicializa a variável  $nwins \leftarrow 1$ ;
4 para cada padrão de entrada ( $x$ ) faça
5   Apresenta o padrão de entrada  $x$  ao mapa
6   Computa a ativação de todos os nodos (Eq. 3.8)
7   Encontra o nodo vencedor  $s$  com maior valor de ativação ( $a_s$ ) (Eq. 3.7)
8   se  $a_s < a_t$  e  $N < N_{max}$  então
9     Cria um novo nodo  $j$  e atribui:  $c_j \leftarrow x, \delta_j \leftarrow 0, \omega_j \leftarrow 1$  e
        $wins_j \leftarrow l_p \times nwins$ 
10    Conecta  $j$  aos outros nodos através da Eq. 2.22
11  senão
12    Atualiza o vetor de distâncias  $\delta_s$  do nodo vencedor e de seus vizinhos
       (Eq. 2.20)
13    Atualiza o vetor de relevâncias  $\omega_s$  do nodo vencedor e de seus vizinhos
       (Eq. 2.21)
14    Atualiza o vetor de centro  $c_s$  do nodo vencedor e de seus vizinhos (Eq. 2.19)
15    Atribui  $wins_s \leftarrow wins_s + 1$ 
16  fim
17  se  $nwins \geq maxcomp$  então
18    Remove os nodos com  $wins_j < l_p \times maxcomp$ 
19    Atualiza as conexões restantes através da Eq. 2.22
20     $nwins \leftarrow 0$ 
21  fim
22   $nwins \leftarrow nwins + 1$ 
23 fim

```

### 3.3.1 Competição

Da mesma forma que na versão original, no OLARFDSSOM cada nodo  $j$  no mapa está associado com três vetores de tamanho  $m$  (onde  $m$  é o número de dimensões do vetor de entrada):  $c_j = \{c_{ji}, i = 1 \dots m\}$  representa a posição (centro) do protótipo do grupo  $j$  no espaço de entrada;  $\omega_j = \{\omega_{ji}, i = 1 \dots m\}$  é o vetor de relevâncias no qual cada componente representa um peso (relevância estimada) entre  $[0,1]$  que o nodo  $j$  aplica para a  $i$ -ésima dimensão de entrada;

$\delta_j = \{\delta_{ji}, i = 1 \dots m\}$  é o vetor que armazena uma média móvel da distância observada entre os padrões agrupados no nodo  $j$  e o vetor de centro.

Desta forma, quando um nodo é criado, seu centro  $\mathbf{c}_j$  é inicializado na posição do último padrão  $\mathbf{x}$  apresentado ao mapa. O vetor de relevâncias  $\boldsymbol{\omega}_j$  tem todos seus componentes inicializados com valor um e os componentes do vetor de médias móveis das distâncias  $\delta_j$  são inicializados com valor zero.

O vencedor da competição,  $s(\mathbf{x})$ , é o nodo que apresenta a maior ativação para o padrão de entrada apresentado:

$$s(\mathbf{x}) = \arg \max_j [ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \boldsymbol{\omega}_j)] \quad (3.7)$$

Neste trabalho, a ativação de um nodo,  $ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \boldsymbol{\omega}_j)$ , é calculada por uma função de base radial da distância ponderada  $D_\omega(\mathbf{x}, \mathbf{c}_j)$  (definida no LARFDSSOM pela Eq. 2.18) e da soma dos componentes do seu vetor de relevâncias  $Sum(\boldsymbol{\omega}_j)$ , apresentada no DSSOM (Bassani e Araujo, 2012) e detalhada na Eq. 3.8. Entretanto, a função de ativação da versão original do LARFDSSOM também poderia ser claramente utilizada.

$$ac(D_\omega(\mathbf{x}, \mathbf{c}_j), \boldsymbol{\omega}_j) = \frac{Sum(\boldsymbol{\omega}_j)}{D_\omega(\mathbf{x}, \mathbf{c}_j) + Sum(\boldsymbol{\omega}_j) + \varepsilon} \quad (3.8)$$

onde  $\varepsilon$  é um valor pequeno para evitar divisão por zero e  $Sum(\boldsymbol{\omega}_j)$  é a soma dos componentes do vetor de relevâncias associado ao nodo  $j$ , dada por:

$$Sum(\boldsymbol{\omega}_j) = \sum_{i=1}^m \omega_{ji} \quad (3.9)$$

A função de ativação calcula valores mais elevados para nodos que levam mais dimensões em conta. Se dois nodos possuem a mesma distância ponderada dos seus centros para um dado padrão de entrada, terá maior ativação aquele que levar em consideração mais fortemente o maior número de atributos.

### 3.3.2 Inserção e Remoção de Nodos

Se em uma competição, um certo limiar de ativação  $a_t$  não for atingido pelo vencedor, um novo nodo é inserido ao mapa na posição do padrão de entrada. Caso contrário, o nodo vencedor e seus vizinhos são atualizados, conforme descrito na próxima subseção e nas linhas 5 a 16 do Alg. 7. É possível definir um limite  $N_{max}$  para o número de nodos no mapa, logo, se este limite for alcançado novos nodos não poderão mais ser criados, mas os nodos existentes continuam sendo atualizados.

Cada nodo  $j$  no mapa está associado a uma variável,  $wins_j$ , que conta o número de vitórias desse nodo em competições com ativações não inferiores ao limiar. Um *reset* ocorre após  $maxcomp$  competições, neste momento são removidos do mapa os nodos que apresentam

número de vitórias abaixo do limiar  $lp \times maxcomp$ , onde  $lp$  é um parâmetro que representa a porcentagem mínima de vitórias necessária para que um nodo permaneça no mapa (linhas 17 a 21 do Alg. 7). Além disso, para evitar a remoção prematura de novos nodos, quando um nodo é criado seu número de vitórias é ajustado para  $lp \times nwins$ , onde  $nwins$  é o número de competições que ocorreram desde o último *reset*. Na versão original do LARFDSSOM, a cada *reset* os nodos que não são removidos tem seu número de vitórias  $wins_j$  reiniciado para zero, o que não acontece nesta versão online para evitar que nodos com um número de vitórias superior ao limiar  $lp \times maxcomp$  em um *reset* anterior sejam removidos por não terem sido vitoriosos recentemente.

### 3.3.3 Atualização do Nodo Vencedor e Seus Vizinhos

A atualização do nodo vencedor e de seus vizinhos é realizada sempre que o vencedor atinge um certo limiar de ativação  $a_t$ , pelo mesmo procedimento realizado na versão original do LARFDSSOM, detalhado no Capítulo 2 (Subseção 2.2.3.2). Portanto, o vetor de centros  $\mathbf{c}_j$  é atualizado pela Eq. 2.19, o vetor de médias  $\boldsymbol{\delta}_j$  pela Eq. 2.20 e o vetor de relevâncias  $\boldsymbol{\omega}_j$  pela Eq. 2.21.

### 3.3.4 Atualização da Vizinhança

A similaridade entre os vetores de relevâncias de pares de nodos determina a existência de uma conexão entre eles. Essas conexões são atualizadas entre todos os nodos presentes no mapa a cada *reset* realizado. O mesmo acontece quando um novo nodo é adicionado ao mapa, após a inclusão, as conexões entre ele e os outros nodos já contidos no mapa são atualizadas. O processo que determina a existência dessas conexões é o mesmo utilizado na versão original do LARFDSSOM (Capítulo 2, Subseção 2.2.3.4), calculado pela Eq. 2.22.

### 3.3.5 Agrupamento no OLARFDSSOM

Os vetores de centro e relevâncias armazenados em cada nodo do mapa podem ser utilizados para agrupar padrões de entrada, entretanto, devido as características online e de topologia variante no tempo do modelo, a indexação entre agrupamentos de nodos e padrões de entrada se torna bastante dinâmica. Essa dinamicidade exige que o agrupamento seja feito em momentos, que podem ser encarados como estados de aprendizagem do mapa.

Desta forma, cada nodo no mapa é rotulado com um índice representando um grupo (ou categoria) e a indexação de um padrão de entrada a um grupo de um nodo pode ser realizada a qualquer momento através de uma simples competição dos nodos no atual estado do modelo, sendo o nodo de maior ativação o vencedor, conforme exposto no Alg. 8.

**Algoritmo 8:** Agrupamento no OLARFDSSOM

```
1 para cada padrão de entrada ( $\mathbf{x}$ ) faça  
2   Apresenta o padrão de entrada  $\mathbf{x}$  ao mapa  
3   Computa a ativação de todos os nodos (Eq. 3.8)  
4   Encontra o nodo vencedor  $s$  com maior valor de ativação ( $a_s$ ) (Eq. 3.7)  
5   Atribui  $\mathbf{x}$  ao grupo com índice do nodo vencedor  $s$   
6 fim
```

### 3.4 Conclusão

Este capítulo apresentou a abordagem desenvolvida nesta dissertação, primeiramente descrevendo em visão geral sua arquitetura e fluxo de funcionamento na Seção 3.1. Abordagem em que o módulo de mapeamento, SEMMAP, cria incrementalmente um mapa semântico do ambiente, auxiliado por módulos de SLAM Métrico e de Reconhecimento de Objetos (não desenvolvidos nesta dissertação), os quais realizam, respectivamente, a localização métrica do agente no ambiente e o reconhecimento de um conjunto pré-definido de objetos conforme o agente se desloca. O módulo de categorização, OLARFDSSOM, é treinado incrementalmente com as informações semânticas contidas nos nodos do SEMMAP e pode determinar as categorias de lugar desses nodos sempre que necessário. Os módulos de mapeamento e de categorização foram posteriormente detalhados, SEMMAP na Seção 3.2 e OLARFDSSOM na Seção 3.3, respectivamente.

O capítulo seguinte descreve experimentos realizados para validar o modelo proposto no contexto do problema em questão, através da avaliação da topologia dos mapas criados, das informações semânticas agregadas e das categorias de lugar determinadas para cada nodo. Os experimentos também buscam demonstrar a capacidade de aprendizagem incremental da abordagem de maneira geral.

# 4

## Experimentos e Análise de Resultados

Neste capítulo são descritos experimentos realizados com o objetivo de validar a abordagem proposta, avaliando sua capacidade na criação incremental de mapas semânticos, com lugares categorizados de forma online e não-supervisionada através de informações semânticas mapeadas, assim como demonstrar sua capacidade de aprendizagem incremental.

Inicialmente, na Seção 4.1 são descritas as medidas de qualidade utilizadas para avaliar a abordagem e a Seção 4.2 descreve como foi realizado o ajuste de parâmetros da mesma nos experimentos. Posteriormente, a Seção 4.3 busca descrever a base de dados utilizada nos experimentos e as ferramentas utilizadas em conjunto com a mesma. Em seguida, na Seção 4.4 é feita uma avaliação da separabilidade das categorias de lugares pelas informações semânticas contidas nos nodos mapeados pelo módulo SEMMAP da abordagem com a base de dados descrita. Adicionalmente, na Seção 4.5, uma avaliação da topologia dos mapas criados é feita em torno do posicionamento dos seus nodos e das conexões criadas entre eles.

Logo após, são descritos três experimentos que foram realizados com a abordagem completa com o propósito de avaliar os mapas criados pela categorização dos lugares mapeados. Infelizmente não foram encontradas opções na literatura de mapeamento semântico que utilizassem a mesma base de dados em experimentos, ou que utilizassem uma outra que fosse adequada para avaliar a abordagem proposta. Por isso, em alguns dos experimentos realizados buscou-se fazer comparações apenas em termos gerais, devido às naturezas diferentes das abordagens, com trabalhos da literatura que utilizaram a mesma base de dados, para dar uma perspectiva das capacidades que a abordagem proposta apresenta.

Portanto, o primeiro experimento (Seção 4.6) buscou fazer uma comparação com os resultados de um método de aprendizagem não-supervisionada de categorização de lugares baseado em *transfer learning* descrito em [Constante et al. \(2013\)](#). No segundo experimento (Seção 4.7), a comparação é feita com o trabalho de [Rubio et al. \(2016\)](#), o qual avalia diferentes métodos de aprendizagem supervisionada na mesma base de dados.

O terceiro e último experimento (Seção 4.8), por sua vez, objetivou demonstrar a capacidade de categorização da abordagem proposta em sub-bases da base de dados utilizada, dando foco ao seu potencial incremental. Desta forma, adicionalmente, com o intuito de apresentar mais

indícios da capacidade incremental da abordagem, um estudo de caso (Seção 4.9) foi realizado para demonstrar os resultados de categorização de caminhos ao decorrer do tempo.

A configuração de parâmetros utilizada em cada um dos experimentos é apresentada na Tabela 4.1. O estudo de caso (Seção 4.9) utilizou a mesma configuração do experimento da Seção 4.8, devido à característica mais abrangente do mesmo.

**Tabela 4.1:** Configuração de parâmetros utilizada em cada um dos experimentos.

Parâmetro	1º Experimento	2º Experimento		3º Experimento
<b>OLARFDSSOM</b>		A	B	
Limiar de atividade ( $a_t$ )	0.9668	0.9816	0.9863	0.9879
Tamanho mínimo de um grupo ( $l_p$ )	0.1414	0.0558	0.0977	0.1914
Taxa de relevância ( $\beta$ )	0.0532	0.0579	0.0252	0.0163
Máximo de competições ( $maxcomp$ )	89	42	11	34
Taxa de aprendizagem do vencedor ( $e_b$ )	0.0436	0.0253	0.1035	0.0118
Taxa de aprendizagem dos vizinhos ( $e_n$ )	0.0109	0.0096	0.0087	0.0076
Suavidade da relevância ( $s$ )	0.0453	0.0800	0.0401	0.0781
Limiar de conexão ( $c$ )	0.1108	0.0416	0.0467	0.0301
<b>SEMMAP</b>				
Limiar de atividade ( $a_t$ )	0.5539	0.5539	0.5539	0.5539
Taxa de aprendizagem ( $e$ )	0.0139	0.0139	0.0139	0.0139
Limiar de soma ( $s_t$ )	7	9	5	5

## 4.1 Medidas de Qualidade

Na literatura existem diversas medidas para avaliar a qualidade dos agrupamentos obtidos em comparação com os agrupamentos esperados, usualmente chamados de *ground truth*. Em Müller *et al.* (2009) são descritas algumas dessas medidas especialmente relacionadas a agrupamento em subespaços. Dentre elas foram escolhidas para avaliar a abordagem desse trabalho a Acurácia e o CE (do inglês, *Clustering Error*).

A Acurácia é uma medida comum na literatura e basicamente especificada por  $\frac{T}{N}$ , onde  $N$  é o número total de exemplos e  $T$  o número de exemplos corretamente preditos, ou seja, uma taxa de predições corretas. Devido a essa natureza de simplesmente avaliar a taxa de acerto de todos os exemplos, a Acurácia não funciona bem para bases desbalanceadas.

O CE é calculado como uma taxa de pontos que estão agrupados de maneira diferente do *ground truth*, sendo uma medida que leva em conta não apenas os agrupamentos produzidos, mas também as dimensões relevantes encontradas em cada grupo e penaliza resultados com agrupamentos demais. Essa penalização é feita através da propriedade do método de mapear

no máximo um agrupamento encontrado com um agrupamento do *ground truth*, assim como no máximo um agrupamento do *ground truth* com um agrupamento encontrado pelo modelo. Desta forma penalizando resultados que subdividem agrupamentos em grupos menores (Müller *et al.*, 2009). Na avaliação, pela ausência de informação sobre as dimensões relevantes dos agrupamentos na base de dados utilizada nos experimentos (ou seja, no *ground truth*), optou-se por considerar igualmente relevância máxima para todas as dimensões de entrada em todos os agrupamentos do *ground truth*. Outra decisão foi por utilizar a medida de forma inversa,  $1 - CE$ , sendo o valor máximo 1 correspondente aos resultados com melhores agrupamentos.

Especialmente devido a propriedade do CE de penalizar resultados com muitos agrupamentos, o mesmo foi escolhido como medida principal deste trabalho e foi utilizado para a realização do ajuste dos parâmetros da abordagem em todos os experimentos. A Acurácia por sua vez tem seus resultados finais elevados conforme o número de agrupamentos encontrados aumenta, o que não é interessante para a nossa abordagem. Entretanto, a mesma foi escolhida como medida secundária devido a larga utilização na literatura relacionada a mapeamento semântico e reconhecimento de lugares através de informação semântica. Desta forma, foi utilizada nos experimentos realizados para comparação em termos gerais com outros trabalhos.

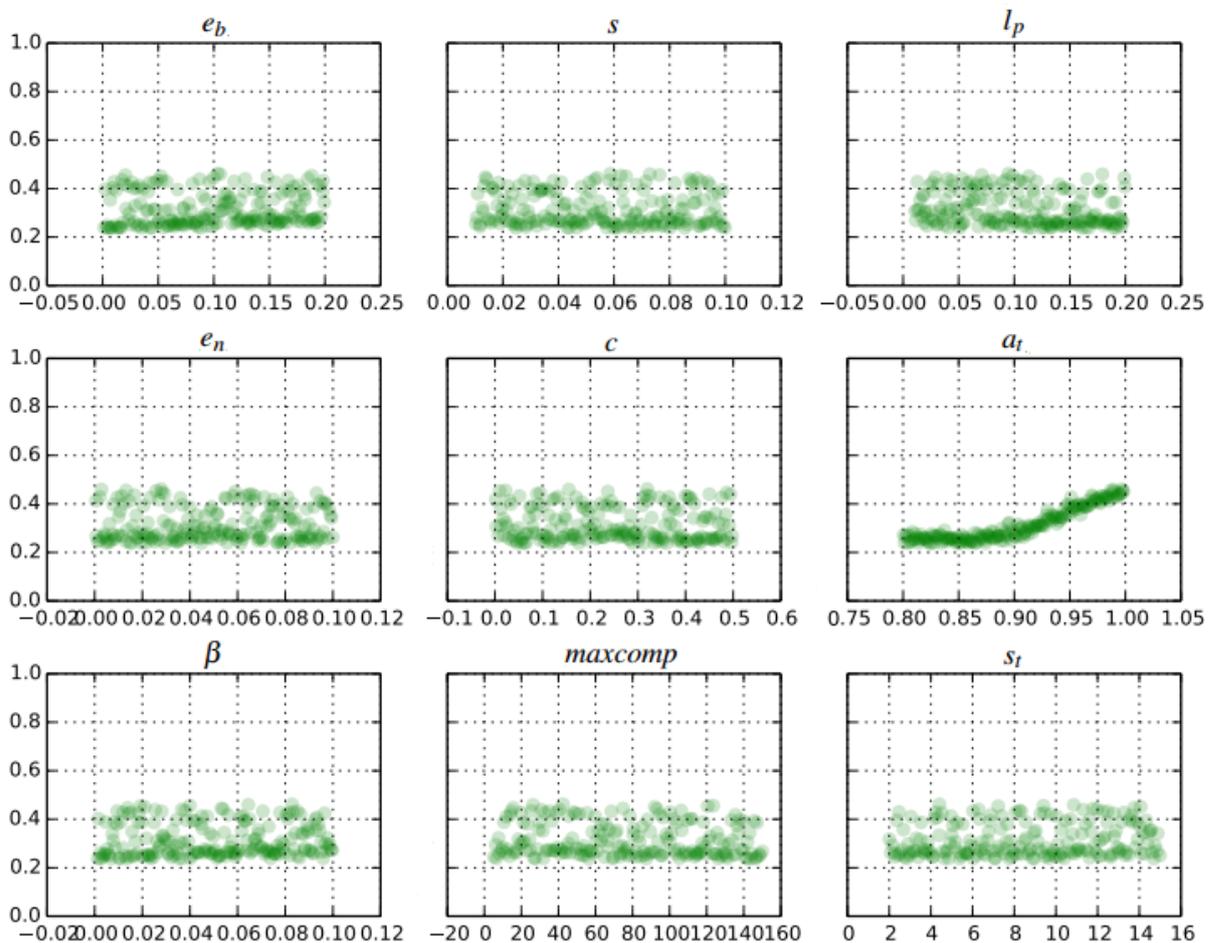
## 4.2 Ajustes Paramétricos

Como a abordagem proposta tem muitos parâmetros, ajustá-los não é uma tarefa fácil. Portanto, para encontrar ajustes ótimos para os experimentos, com exceção dos parâmetros  $a_t$  e  $e$  do módulo de mapeamento semântico (SEMMAP) que foram previamente definidos por afetarem diretamente a construção do mapa topológico, foi adotada uma metodologia baseada em técnicas de análise de sensibilidade (Saltelli *et al.*, 2009) chamada de Amostragem de Hipercubo Latina (LHS, do inglês *Latin Hypercube Sampling*) (Helton *et al.*, 2005). Assim, de acordo com a LHS, a abordagem foi inicialmente executada 200 vezes (em cada experimento) com diferentes parâmetros amostrados dentro de um intervalo previamente estabelecido. A LHS garante uma cobertura completa do espaço de parâmetros, evitando que fiquem regiões grandes sem amostras, como pode ocorrer em uma amostragem puramente aleatória. Para isso, o intervalo de cada parâmetro é dividido em 200 subintervalos de igual probabilidade e um único valor é escolhido aleatoriamente a partir de cada subintervalo. Os intervalos estabelecidos para cada parâmetro estão descritos na Tabela 4.2.

Após as execuções, análises foram realizadas para identificar os parâmetros mais relevantes (ou seja, com maior sensibilidade). Em seguida, a abordagem foi novamente executada 200 vezes (em cada experimento) com os parâmetros mais relevantes amostrados dentro de intervalos menores, favoráveis aos seus melhores resultados e contidos nos intervalos anteriormente estabelecidos (Tabela 4.2). Esse mesmo procedimento foi realizado repetidas vezes com diferentes subintervalos dos intervalos originais, análises foram realizadas e os melhores resultados alcançados em cada experimento foram registrados. Ao fim das execuções, chegou-se

a conclusão de que os parâmetros que se mostraram mais relevantes foram os  $a_t$ ,  $l_p$  e  $maxcomp$ , do OLARFDSSOM.

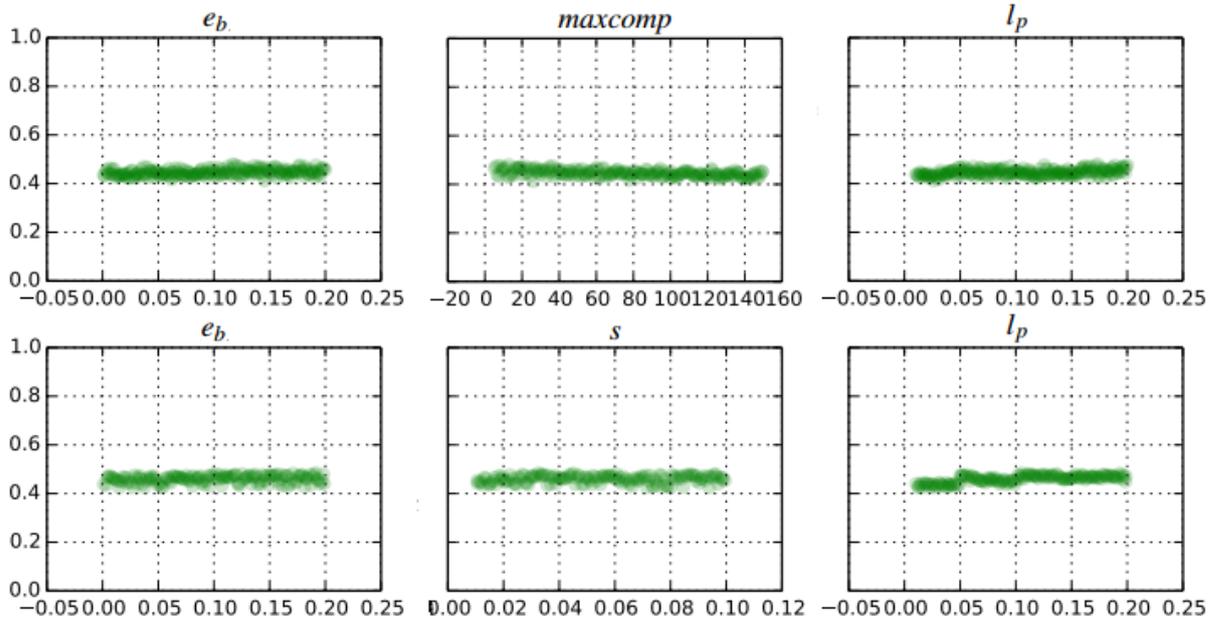
Adicionalmente, com a finalidade de ilustrar o comportamento da análise de sensibilidade dos parâmetros da abordagem baseando-se na métrica de desempenho CE e utilizando o conjunto de dados descrito na Seção 4.3, são apresentados gráficos de dispersão exemplificadores nas Figuras 4.1 e 4.2. Esses gráficos indicam que o parâmetro  $a_t$  é o maior responsável pela variação observada nos resultados (Figura 4.1), seguido por  $maxcomp$  e  $l_p$  (com menor sensibilidade). O efeito de  $maxcomp$  pode ser melhor identificado com a fixação de  $a_t$  em um valor ideal, assim como o  $l_p$  de forma menos evidente. Com a fixação de  $a_t$  e  $maxcomp$ , a sensibilidade de  $l_p$  fica mais evidente (Figura 4.2). As variações nos outros parâmetros não apresentam tanto efeito significativo nos resultados finais quanto os já mencionados.



**Figura 4.1:** Gráficos de dispersão do CE obtido em função de cada um dos parâmetros analisados. Os valores dos parâmetros variaram nos intervalos descritos na Tabela 4.2.

### 4.3 Descrição da Base de Dados

Neste trabalho, os experimentos foram realizados utilizando a base de dados COLD (*COsy Localization Database*), a qual consiste de três partes separadas (*sub-databases*) adquiridas



**Figura 4.2:** Gráficos de dispersão do CE obtido em função de cada um dos parâmetros analisados. Os três gráficos de cima (primeira linha) foram obtidos com  $a_t$  fixado em um valor ótimo e os três de baixo (segunda linha) foram obtidos fixando  $a_t$  e  $maxcomp$  com valores ótimos.

**Tabela 4.2:** Intervalo paramétrico para a abordagem.

Parâmetros	min	max
<b>OLARFDSSOM</b>		
Limiar de atividade ( $a_t$ )	0.8	0.999
Tamanho mínimo de um grupo ( $l_p$ )	0.01	0.2
Taxa de relevância ( $\beta$ )	0.001	0.01
Máximo de competições ( $maxcomp$ )	5	150
Taxa de aprendizagem do vencedor ( $e_b$ )	0.001	0.02
Taxa de aprendizagem dos vizinhos ( $e_n$ )	0.0001	0.01
Suavidade da relevância ( $s$ )	0.01	0.1
Limiar de conexão ( $c$ )	0	0.5
<b>SEMMAP</b>		
Limiar de soma ( $s_t$ )	2	15

de três diferentes laboratórios localizados em três cidades europeias distintas (COLD-Freiburg, COLD-Ljubljana e COLD-Saarbrücken). As mesmas contém sequências de imagens capturadas com câmeras regulares e omnidirecionais, juntamente com dados de *laser range scans*, dados odométricos e informações de localização. Os dados foram adquiridos utilizando plataformas robóticas (Figura 4.3) em diferentes horários, com diferentes condições de iluminação, em áreas de diferentes tipos nos laboratórios descritos, em forma de sequências de dados capturados

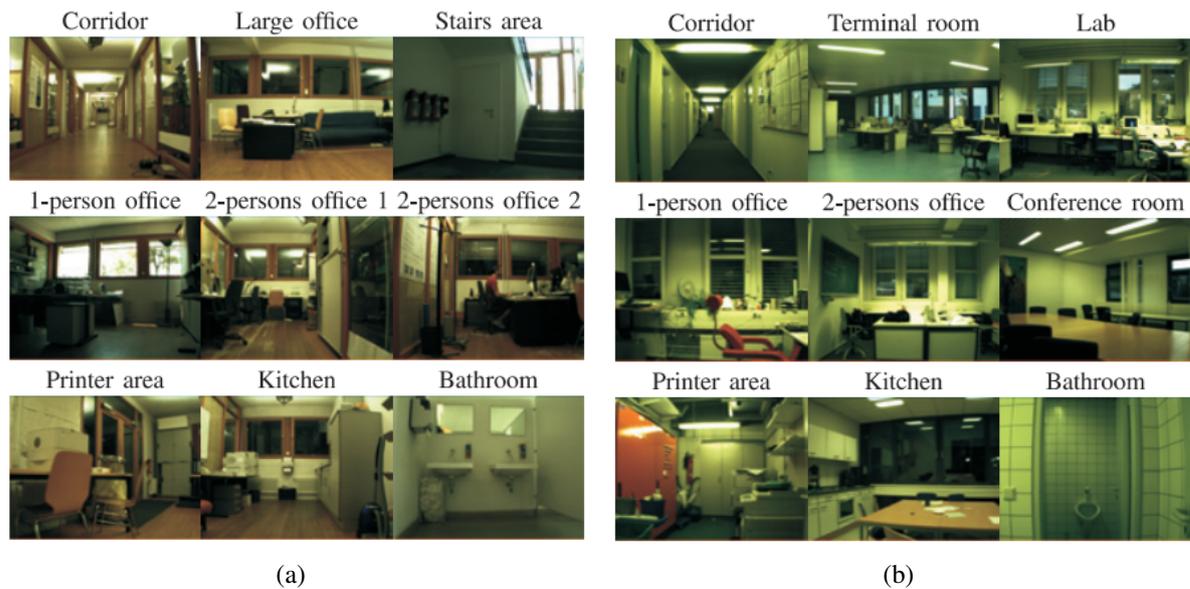
em caminhos percorridos pelas plataformas robóticas nos laboratórios. Apenas as imagens adquiridas de câmeras regulares (apresentando resolução de 640 x 480 pixels, exemplos ilustrados na Figura 4.4) e a localização (posição da plataforma robótica no ambiente) são utilizadas nesta pesquisa. As imagens são utilizadas na atividade de reconhecimento de objetos e as informações de localização são fornecidas à abordagem como as que seriam fornecidas pelo módulo auxiliar de SLAM Métrico. A base de dados é gratuita para a comunidade, descrita por [Pronobis e Caputo \(2009\)](#) e foi escolhida para a realização dos experimentos especialmente por incorporar imagens e informações de localização aos dados fornecidos.



**Figura 4.3:** Plataformas robóticas utilizadas na aquisição das sub-bases. ActivMedia PeopleBot na COLD-Saarbrücken, ActivMedia Pioneer-3 na COLD-Freiburg e iRobot ATRV-Mini na COLD-Ljubljana. ([Pronobis e Caputo, 2009](#))

Ao todo a base de dados contém 76 sequências de dados de 9 caminhos diferentes, sendo destas, 26 sequências de 3 caminhos na sub-base COLD-Freiburg, 18 sequências de 2 caminhos na sub-base COLD-Ljubljana e 32 sequências de 4 caminhos na sub-base COLD-Saarbrücken. Entretanto, devido a problemas nas coordenadas de localização das sequências de dados disponibilizadas, por vezes em falta ou erradas, ao todo nos experimentos apenas 24 sequências de 6 caminhos foram utilizadas (Tabela 4.3), sendo elas 9 sequências de 2 caminhos da COLD-Freiburg e 15 sequências de 4 caminhos da COLD-Saarbrücken (a Figura 4.5 expõe dois caminhos em mapas de partes dos laboratórios de cada sub-base utilizada).

No total (avaliando todos os 6 caminhos utilizados) estão presentes 11 categorias de lugar e os caminhos escolhidos para cada experimento contém um sub-conjunto dessas categorias que pode ou não se diferenciar de um para outro. A Tabela 4.4 descreve as categorias e associa cada uma delas a um índice a ser utilizado como representação nas avaliações e experimentos detalhados nas seções posteriores. A tabela também inclui os números de nodos criados pelo SEMMAP por categoria, quando aplicado à todas as 24 sequências de dados utilizadas. A configuração de parâmetros utilizada foi a do terceiro experimento (descrito anteriormente na



**Figura 4.4:** Exemplos de imagens de alguns tipos de lugares dos laboratórios de Freiburg (a) e Saarbrücken (b). (Pronobis e Caputo, 2009)

Tabela 4.1). Além disso, fica evidente na Tabela 4.4 o desbalanceamento na quantidade de nodos entre as categorias presentes na base de dados, sendo a categoria “corredor” a que apresenta maior quantidade.

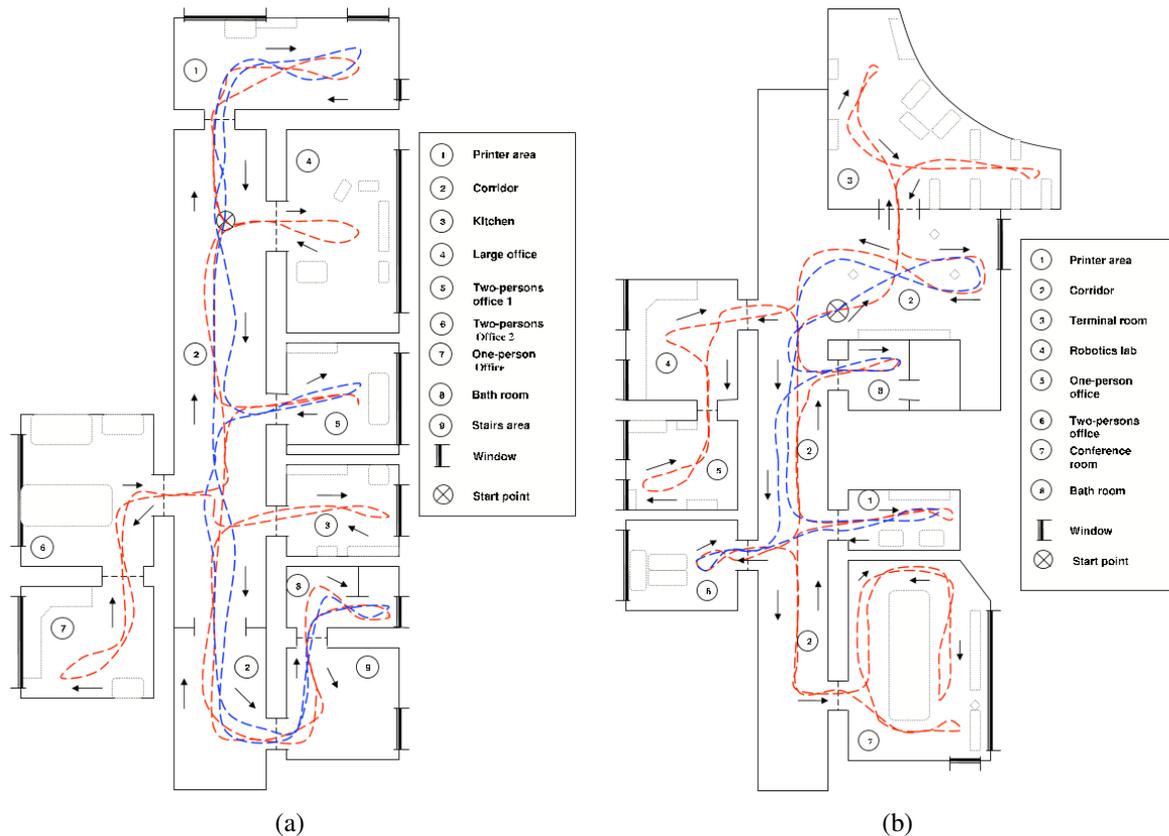
Conforme discutido no capítulo anterior, neste momento da pesquisa, para avaliação da abordagem foi determinado o uso de apenas um tipo de informação semântica: os objetos que estão presentes nos ambientes. Desta forma, a abordagem constrói o mapa semântico dos ambientes incorporando aos nodos os objetos reconhecidos na região coberta por cada nodo. Assim, um conjunto de 18 tipos de objetos foi determinado para reconhecimento. Foram eles: “Persiana, estante de livro, ventilador elétrico, sofá, pia, dispensador de sabão, vaso sanitário, fotocopiadora, monitor, computador de mesa, escrivaninha, mesa, poltrona, corrimão, forno micro-ondas, fogão, máquina de lavar pratos e torradeira”.

**Tabela 4.3:** Relação do número total de caminhos e sequências presentes na base de dados COLD completa e do que foi utilizado nesta dissertação. Caminho foi abreviado para “Cam” e sequência para “Seq”.

Sub-base	Nº Cam. Total	Nº Seq. Total	Nº Cam. Utilizados	Nº Seq. Utilizadas
Freiburg	3	26	2	9
Ljubljana	2	18	0	0
Saarbrücken	4	32	4	15
Total	9	76	6	24

Nesse contexto, para a atividade de reconhecimento dos objetos foi utilizado nas imagens de todas as sequências (ou instâncias) dos caminhos selecionados o modelo Inception-v3 (Szegedy *et al.*, 2015) da biblioteca de software para computação numérica TensorFlow (Abadi *et al.*,

2015). O Inception-V3 foi um modelo treinado utilizando a base de dados ImageNet (Deng *et al.*, 2009) para a competição do ano de 2012 e classifica imagens em 1000 diferentes classes gerais. Nesta dissertação só foram considerados objetos reconhecidos com score igual ou maior que 0.2, esse valor foi escolhido por perceber-se que objetos reconhecidos com scores inferiores não representavam (em geral) corretamente os objetos presentes nas regiões. Os valores de score variam entre  $[0,1]$ , sendo 0 (zero) o valor que representa o não reconhecimento de um objeto e 1 (um) a certeza máxima de que o objeto foi reconhecido.



**Figura 4.5:** Mapa de parte do laboratório de Freiburg (a) e de parte do laboratório de Saarbrücken (b). Caminhos seguidos pelos robôs nos ambientes durante a aquisição dos dados são representados por traços, neste caso, dois caminhos de cada laboratório são desenhados (em cores azul e vermelha). (Pronobis e Caputo, 2009)

## 4.4 Avaliação de Separabilidade das Categorias dos Nodos do SEMMAP

Devido à existência da possibilidade de os mesmos objetos constarem em diferentes categorias de lugares, assim como a de ocorrer erros no reconhecimento dos mesmos por parte do modelo inteligente utilizado devido a diversos fatores, como seu treinamento não específico para os tipos de imagens utilizadas ou a baixa qualidade das imagens apresentadas pela base de

**Tabela 4.4:** Relação de categorias, número de nodos e suas representações em números utilizados nos experimentos e avaliações.

Número	Categoria	Nº de Nodos
0	Corredor	478
1	Área de Impressão	79
2	Laboratório de Robótica	28
3	Área de Escadas	20
4	Banheiro	58
5	Cozinha	51
6	Sala de Terminal	116
7	Sala de Reuniões	90
8	Escritório de Uma Pessoa	60
9	Escritório de Duas Pessoas	65
10	Escritório Grande	22

dados, foi realizada uma análise da separação entre os padrões de cada categoria de lugar com o objetivo de evidenciar o nível de dificuldade dos dados resultantes.

Nesta análise, são criados  $p$  grupos (*clusters*) onde  $p$  é a quantidade de categorias presentes no conjunto de caminhos escolhidos, cada um com  $k$  nodos (criados pelo SEMMAP), sendo  $k$  o número de nodos presentes em cada categoria no *ground truth*. A partir disto, é computada uma métrica de separabilidade através da Eq. 4.1 para cada par de grupos  $(i, j)$ , onde  $i = 1, 2, \dots, p$  e  $j = 1, 2, \dots, p$ .

$$S(\mathbf{m}_i, \mathbf{m}_j) = \frac{D(\mathbf{m}_i, \mathbf{m}_j)}{\sigma_i + \varepsilon} \quad (4.1)$$

onde  $\varepsilon$  é um valor pequeno para evitar divisão por zero,  $\mathbf{m}_i$  e  $\mathbf{m}_j$  são os vetores de médias aritméticas dos vetores de objetos de todos os nodos dos respectivos grupos  $i$  e  $j$ ,  $\sigma_i$  é o desvio padrão das distâncias euclidianas do vetor de médias aritméticas dos objetos do grupo  $i$  para os vetores de objetos de todos os nodos contidos neste mesmo grupo e  $D(\mathbf{m}_i, \mathbf{m}_j)$  é a distância Euclidiana entre os vetores  $\mathbf{m}_i$  e  $\mathbf{m}_j$ , explicitada anteriormente na Eq. 3.3.

Utilizando desta métrica, espera-se que valores inferiores a 1 (um) demonstrem uma sobreposição entre os grupos, quanto mais próximo de zero, maior a sobreposição. Portanto, para dar uma perspectiva do nível de dificuldade dos experimentos que serão tratados nas seções seguintes, foi escolhido apresentar a visualização da separabilidade a partir do cálculo descrito anteriormente (Eq. 4.1) para alguns caminhos das sub-bases COLD-Freiburg e COLD-Saarbrücken. As sequências de dados relativas a esses caminhos são submetidas ao SEMMAP, que cria os mapas semânticos dos ambientes e tem as categorias de cada nodo presente nos mapas definidas pelo *ground truth*. Os nodos resultantes dos mapas, através dos vetores de objetos contidos neles, são então avaliados pela métrica descrita. O ajuste de parâmetros utilizado para o SEMMAP foi o mesmo do terceiro experimento (Seção 4.8), escolhido pela característica mais

abrangente do mesmo em comparação aos demais.

Primeiramente, os caminhos (seis sequências de cada) de número 2 (dois) de cada uma das duas sub-bases foram escolhidos, sendo esses os mais desafiadores de ambas (os mesmos que foram utilizados integralmente nos experimentos da Seção 4.7). As Tabelas 4.5 e 4.6 apresentam os valores calculados e demonstram que algumas categorias de lugares não apresentam muita diferença entre padrões, de acordo com o método utilizado (resultados inferiores a 1 ou próximos). Destaque para a categoria 0 (corredor) que apresenta muita semelhança a muitas outras categorias em ambas as tabelas, assim como as categorias 5 e 10 (cozinha e escritório grande, respectivamente) na Tabela 4.5 e as categorias 6 e 7 (sala de terminal e sala de reuniões, respectivamente) na Tabela 4.6, contrastando com os valores mais altos da categoria 4 (banheiro) para os demais em ambas as tabelas. A Tabela 4.4 indica os números e rótulos de cada categoria.

**Tabela 4.5:** Resultado da visualização de separabilidade dos dados resultantes das categorias de 6 sequências (juntas) do caminho 2 da sub-base COLD-Freiburg. As categorias são representadas em números e o nome categoria foi abreviado para Ctg.

	Ctg 0	Ctg 1	Ctg 10	Ctg 9	Ctg 8	Ctg 3	Ctg 4	Ctg 5	Média
Ctg 0	0.0	2.514	0.905	2.535	1.44	2.72	2.079	0.896	1.636
Ctg 1	2.777	0.0	3.166	3.826	3.323	4.468	3.967	2.613	3.018
Ctg 10	1.044	3.309	0.0	2.874	1.904	3.304	2.509	1.388	2.041
Ctg 9	2.258	3.085	2.217	0.0	1.871	3.37	2.929	2.442	2.272
Ctg 8	1.56	3.259	1.786	2.276	0.0	3.402	2.739	1.86	2.11
Ctg 3	9.105	13.543	9.581	12.667	10.515	0.0	7.885	9.485	9.098
Ctg 4	3.612	6.239	3.776	5.714	4.394	4.092	0.0	3.816	3.955
Ctg 5	0.813	2.146	1.091	2.487	1.557	2.57	1.992	0.0	1.582
Média	2.646	4.262	2.815	4.047	3.126	2.991	3.012	2.813	3.214

De forma complementar, caminhos menos desafiadores foram escolhidos (três sequências de cada), o de número 1 (um) da sub-base COLD-Freiburg e o de número 3 da COLD-Saarbrücken. Ao visualizar as Tabelas 4.7 e 4.8, as quais descrevem os resultados, podemos verificar um comportamento semelhante na similaridade de categorias aos mencionados no parágrafo anterior, com mais uma vez a categoria 0 (corredor) sendo a mais semelhante as demais. Entretanto, com destaque para as sequências da COLD-Saarbrücken, os valores resultantes se apresentam mais elevados (mais distantes de 1) e com média total (encontrado na última linha e última coluna) superior ao das Tabelas 4.5 e 4.6, discutidas anteriormente e relativas as mesmas sub-bases respectivamente. Demonstrando a característica dos caminhos escolhidos de maior separabilidade nos padrões das categorias, logo, de menor dificuldade.

O principal fator identificado como causa da baixa separabilidade entre os padrões de algumas categorias foi a escolha da utilização de apenas objetos reconhecidos através de imagens como pistas semânticas na construção dos nodos pelo SEMMAP. Nos nodos posicionados em áreas de transição entre um tipo de lugar e outro, houve uma certa generalização dos objetos

reconhecidos devido ao fato dos mesmos poderem ser reconhecidos a longa distância, ou seja, a partir de outro ambiente. A categoria mais afetada com isso foi corredor (categoria 0), fator este que contribuiu muito para os baixos valores de separabilidade entre esta categoria e as demais nas avaliações. Além disso, erros no reconhecimento dos objetos por parte do método inteligente de reconhecimento utilizado também foram bastante prejudiciais durante a construção dos nodos em modo geral. Entretanto, apesar dos problemas comentados, a informação armazenada nos nodos possibilita uma separação relativamente boa para a maior parte das categorias presentes nos caminhos analisados, conforme demonstraram os resultados das Tabelas 4.5, 4.6, 4.7 e 4.8.

**Tabela 4.6:** Resultado da visualização de separabilidade dos dados resultantes das categorias de 6 sequências (juntas) do caminho 2 da sub-base COLD-Saarbrücken. As categorias são representadas em números e o nome categoria foi abreviado para Ctg.

	Ctg 0	Ctg 6	Ctg 2	Ctg 8	Ctg 9	Ctg 7	Ctg 1	Ctg 4	Média
Ctg 0	0.0	0.832	2.157	2.985	1.731	0.699	2.48	3.983	1.858
Ctg 6	0.826	0.0	2.04	2.934	1.858	0.875	1.993	4.21	1.842
Ctg 2	4.337	4.134	0.0	3.871	3.776	3.613	6.869	9.193	4.474
Ctg 8	4.153	4.115	2.679	0.0	3.966	3.935	5.515	6.933	3.912
Ctg 9	4.839	5.234	5.249	7.967	0.0	3.66	8.846	12.055	5.981
Ctg 7	0.849	1.071	2.182	3.434	1.59	0.0	3.246	5.064	2.18
Ctg 1	4.01	3.248	5.524	6.408	5.117	4.323	0.0	6.604	4.404
Ctg 4	5.656	6.026	6.492	7.075	6.125	5.922	5.8	0.0	5.387
Média	3.084	3.082	3.29	4.334	3.02	2.879	4.344	6.005	3.755

**Tabela 4.7:** Resultado da visualização de separabilidade dos dados resultantes das categorias de 3 sequências (juntas) do caminho 1 da sub-base COLD-Freiburg. As categorias são representadas em números e o nome categoria foi abreviado para Ctg.

	Ctg 0	Ctg 1	Ctg 9	Ctg 3	Ctg 4	Média
Ctg 0	0.0	2.077	2.668	3.008	2.127	1.976
Ctg 1	3.326	0.0	5.49	6.277	5.148	4.048
Ctg 9	3.158	4.06	0.0	4.806	4.005	3.206
Ctg 3	7.955	10.369	10.735	0.0	5.907	6.993
Ctg 4	4.013	6.066	6.382	4.214	0.0	4.135
Média	3.691	4.514	5.055	3.661	3.438	4.072

## 4.5 Avaliação da Topologia dos Mapas Criados

Antes de experimentar a abordagem completa, a topologia dos mapas criados pelo módulo de mapeamento (SEMMA) foi avaliada em torno de duas características: posição dos nodos e conexões criadas entre eles. Para isso, foram utilizadas 3 (três) sequências de

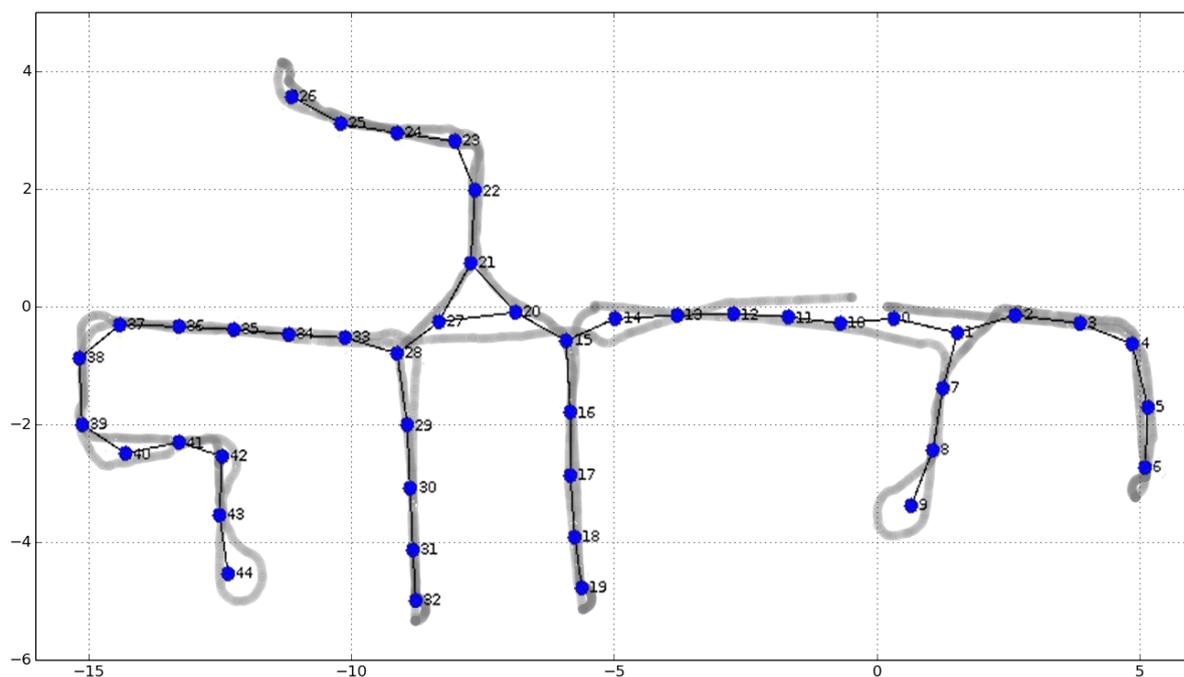
**Tabela 4.8:** Resultado da visualização de separabilidade dos dados resultantes das categorias de 3 sequências (juntas) do caminho 3 da sub-base COLD-Saarbrücken. As categorias são representadas em números e o nome categoria foi abreviado para Ctg.

	Ctg 0	Ctg 4	Ctg 8	Ctg 1	Média
Ctg 0	0.0	3.509	3.518	3.748	2.694
Ctg 4	8.805	0.0	12.784	14.0	8.897
Ctg 8	3.9	5.648	0.0	6.149	3.924
Ctg 1	31.152	46.369	46.099	0.0	30.905
Média	10.964	13.881	15.6	5.974	11.605

cada um dos 6 (seis) caminhos da base de dados, sendo eles, 2 (dois) da COLD-Freiburg e 4 (quatro) da COLD-Saarbrücken. As sequências foram apresentadas a abordagem e os mapas topológicos criados pelo SEMMAP foram avaliados sem levar em consideração as informações semânticas agregadas. O ajuste de parâmetros utilizado no SEMMAP foi mais uma vez o mesmo do experimento descrito na Seção 4.8, os valores podem ser encontrados na Tabela 4.1.

Primeiramente, o posicionamento dos 695 nodos dos 18 mapas criados (provenientes das 18 sequências de dados utilizadas) foram visualmente analisados. Para isso, as coordenadas fornecidas nas sequências de dados foram plotadas em gráficos (um para cada sequência) e os nodos criados pelo SEMMAP foram posicionados em suas coordenadas nestes mesmos gráficos. Os gráficos criados foram analisados visualmente de forma a verificar o posicionamento dos nodos criados em relação as coordenadas fornecidas incrementalmente e os resultados demonstraram que os nodos foram bem posicionados em todos os casos, ficando sempre em posição média em relação ao caminho percorrido. A Figura 4.6 apresenta o gráfico relativo a sequência de dados de número 3 do caminho 2 da sub-base COLD-Freiburg, pode-se perceber que os nodos foram bem posicionados em todas as situações. O mesmo acontece no gráfico exposto na Figura 4.7, relativo a sequência de dados de número 3 do caminho 1 da sub-base COLD-Saarbrücken.

Em seguida, as conexões criadas entre nodos durante o processo de mapeamento foram analisadas. Ao todo, 710 conexões foram formadas, 227 entre os nodos mapeados das sequências de dados da COLD-Freiburg e 483 da COLD-Saarbrücken. Destas, apenas uma conexão ligou nodos erroneamente, resultando numa Acurácia de 0.9986. O erro apresentado se deu numa conexão criada entre os nodos de índice 5 e 3 do mapa relativo a sequência de número 3 do caminho 3 da sub-base COLD-Saarbrücken, a Figura 4.8 traz o gráfico (seguindo o mesmo procedimento dos anteriores) relativo a essa sequência e a Figura 4.9 expõe individualmente essa conexão. Acredita-se que o erro se deu por uma repentina mudança nas coordenadas fornecidas no processo de mapeamento, de forma que a abordagem conectou o nodo visitado atualmente a um outro nodo (que mapeou essa entrada) devido a um erro de posicionamento momentâneo nas coordenadas.



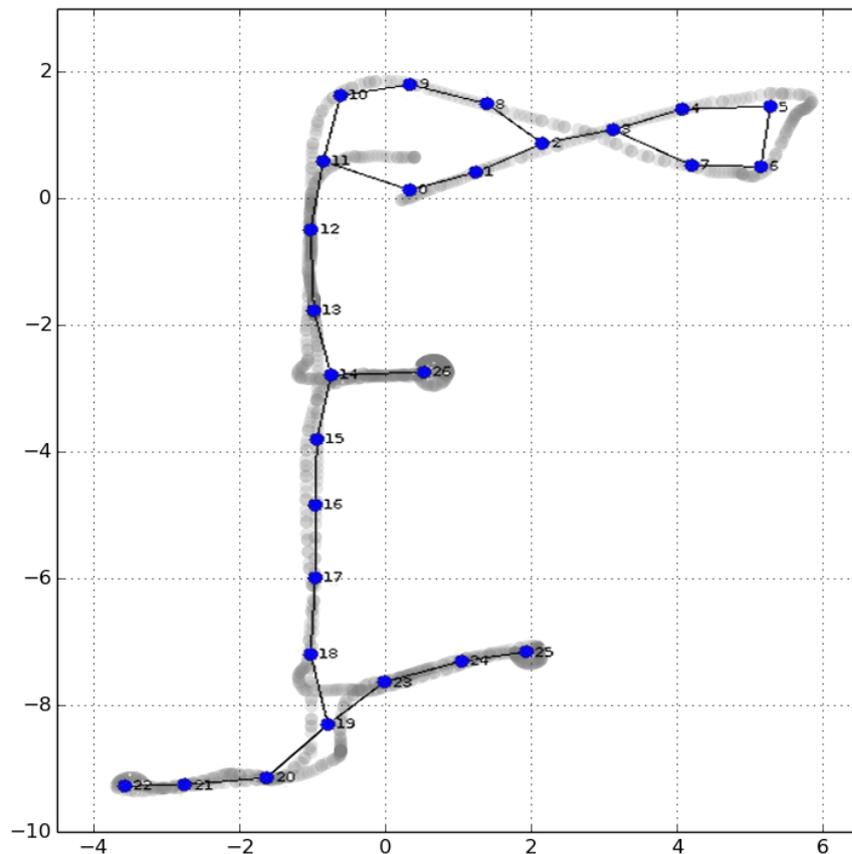
**Figura 4.6:** Gráfico do caminho percorrido na sequência 3 do caminho 2 da sub-base COLD-Freiburg. As marcações em cinza representam as coordenadas fornecidas pelo método de SLAM durante o percurso e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles.

## 4.6 Experimento Comparativo 1

O primeiro dos experimentos realizados com a abordagem completa buscou demonstrar a capacidade da abordagem proposta através da comparação em perspectiva geral com um trabalho da literatura que apesar de bastante diferente em certas características, também busca categorizar lugares através de informações semânticas visuais utilizando uma abordagem de aprendizagem não-supervisionada, o que é um dos objetivos deste trabalho.

Em [Constante \*et al.\* \(2013\)](#) foi introduzido um método de aprendizagem que utiliza de conhecimento de lugares anteriormente rotulados para a categorização de novos ambientes através da transferência de conhecimento de forma não-supervisionada. Essa categorização é feita de forma *offline* e *frame a frame*, portanto apenas informações locais contidas em cada imagem são analisadas por vez e não é criado um mapa semântico dos ambientes.

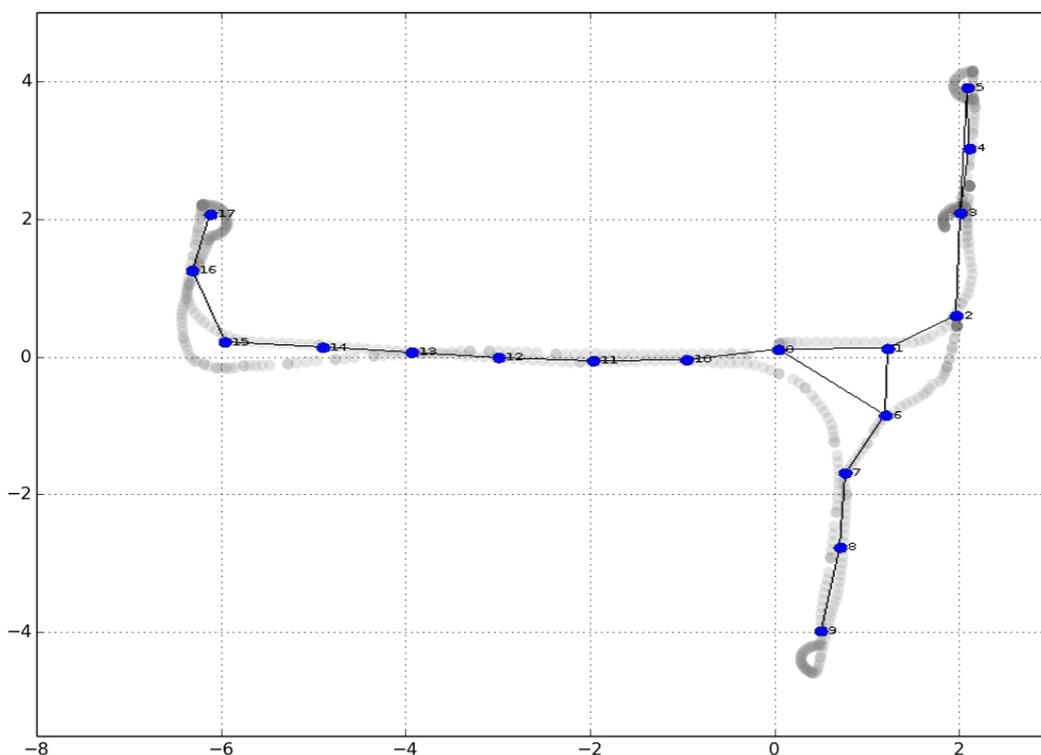
Sendo assim, [Constante \*et al.\* \(2013\)](#) apresenta resultados obtidos da utilização de pares de sequências de caminhos distintos da base de dados descrita na Seção 4.3. Destes pares de sequências, as primeiras foram completamente rotuladas e incorporadas ao método que em seguida utiliza destas informações para transferir parte do conhecimento que é julgado interessante para categorizar uma nova sequência de um caminho. Os caminhos não necessariamente tem as mesmas categorias ou a mesma quantidade de categorias presentes. O método utiliza os descri-



**Figura 4.7:** Gráfico do caminho percorrido na sequência 3 do caminho 1 da sub-base COLD-Saarbrücken. As marcações em cinza representam as coordenadas fornecidas pelo método de SLAM durante o percurso e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles.

tores de imagens SPMK (do inglês, *Spatial Pyramid Matching Kernel*) (Lazebnik *et al.*, 2006) e SPACT (do inglês, *Spatial Principal Component Analysis of Census Transform Histograms*) (Constante *et al.*, 2013) como pistas visuais, primeiramente como pistas únicas e logo após a combinação de ambos.

Desta forma, foram selecionados resultados apresentados pelo autor com 3 pares de sequências contidas na base de dados, também utilizadas neste trabalho, resultados estes fornecidos em forma de Acurácia. Posteriormente, estes mesmos pares foram experimentados com a abordagem deste trabalho. Buscando aproximar os cenários de experimentação, ambas as sequências de cada par foram utilizadas para treinamento e apenas a segunda foi utilizada para teste. O ajuste de parâmetros foi realizado considerando o CE como principal medida de qualidade, todavia, tanto o CE quanto a Acurácia foram calculados. Se os parâmetros fossem ajustados pela Acurácia, resultados ainda melhores poderiam ser obtidos para esta métrica, porém, o número de agrupamentos obtidos pelo OLARFDSSOM se distanciaria dos valores esperados.



**Figura 4.8:** Gráfico do caminho percorrido na sequência 3 do caminho 3 da sub-base COLD-Saarbrücken. As marcações em cinza representam as coordenadas fornecidas e as unidades em azul os nodos criados no mapeamento topológico. Os números associados aos nodos representam a ordem em que foram criados e as arestas entre nodos representam conexão entre eles.

A Tabela 4.9 dispõe os resultados obtidos por [Constante et al. \(2013\)](#), assim como os obtidos pela abordagem deste trabalho em valores de Acurácia, que como pode ser visto se aproximam significativamente dos anteriores. Ficando entre os valores com pistas únicas apresentados por [Constante et al. \(2013\)](#) utilizando o primeiro par de sequências (primeira linha da tabela), acima até mesmo do resultado com a combinação das pistas visuais utilizando o segundo par (segunda linha da tabela), ficando abaixo apenas com o último par (última linha da tabela), mas ainda bastante próximo. Uma comparação utilizando teste de hipótese não foi realizada devido a diferença entre as naturezas dos trabalhos.

Mais detalhes sobre os resultados obtidos com o modelo proposto nesta dissertação podem ser encontrados na Tabela 4.10. Nela encontram-se os valores de CE resultantes para cada par de sequências experimentado, assim como as Acurácias mencionadas no parágrafo anterior. A Tabela 4.10 também dispõe os números de agrupamentos desejados e obtidos com o OLARFDSSOM. Portanto, mesmo a abordagem não apresentando nenhum tipo de supervisão, realizando treinamento online e utilizando apenas objetos reconhecidos como pistas visuais, pode-se afirmar que nesse primeiro experimento foram obtidos resultados de Acurácia compatíveis aos apresentados por [Constante et al. \(2013\)](#), assim como valores dentro do esperado para CE e para o número de agrupamentos apresentado.



**Figura 4.9:** Conexão criada erroneamente entre os nodos 5 e 3 durante o processo de mapeamento da sequência 3 do caminho 3 da sub-base COLD-Saarbrücken.

**Tabela 4.9:** Resultados apresentados por (Constante *et al.*, 2013) em valores de Acurácia. Na terceira coluna são listados os resultados do autor utilizando apenas o descritor SPMK, na quarta o SPACT e na quinta a combinação dos dois. Os resultados da abordagem proposta nesta dissertação são descritos na última coluna. Freiburg foi abreviado para *Fr* e Saarbrücken para *Sa*. Cloudy ou night representam o estado de iluminação do caminho no momento que a sequência foi capturada.

1º Sequência	2º Sequência	SPMK	SPACT	SPMK + SPACT	Este Trabalho
Fr(Caminho 1) cloudy 1	Fr(Caminho 2) cloudy 1	0.7810	0.6204	0.8267	<b>0.6976</b>
Fr(Caminho 1) cloudy 1	Sa(Caminho 1) cloudy 2	0.5019	0.5765	0.6127	<b>0.8148</b>
Fr(Caminho 1) cloudy 1	Sa(Caminho 2) night 1	0.5320	0.5620	0.6089	<b>0.5128</b>

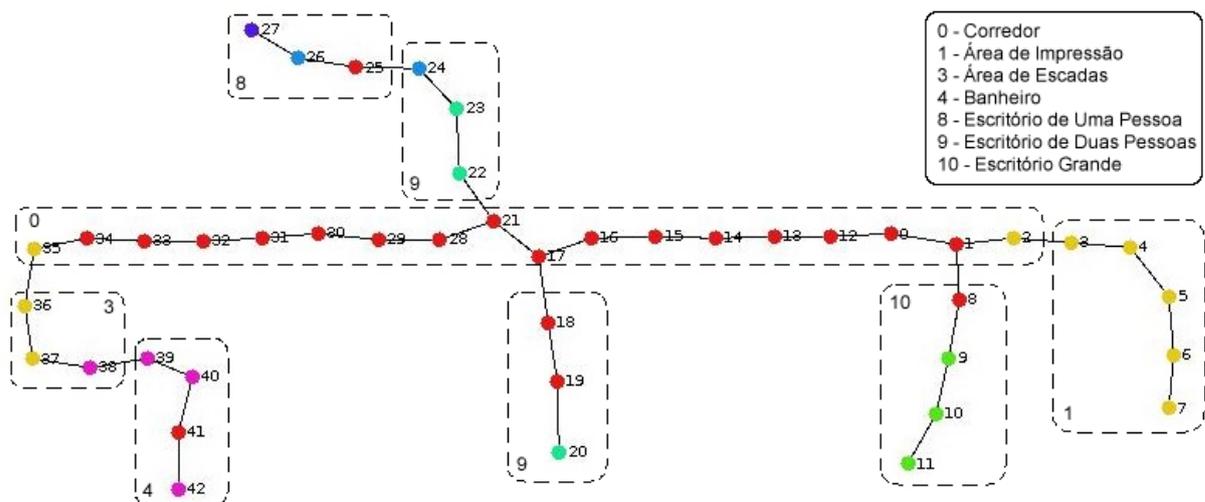
De forma complementar é possível visualizar na Figura 4.10 um dos mapas semânticos criados pela abordagem nesse experimento, cada coloração utilizada nos nodos representa um agrupamento diferente no OLARFDSSOM e os quadrados tracejados delimitando grupos de nodos representam as categorias dos mesmos no *ground truth*. Ao se observar o mapa apresentado na Figura 4.10 pode-se perceber um resultado significativo no agrupamento dos nodos, mas com um problema (recorrente na criação dos mapas) que faz com que nodos em regiões de mesma categoria pertençam a agrupamentos diferentes no OLARFDSSOM, principalmente em nodos que estão perto de áreas de transição (como portas). Acredita-se que isso se dá primeiramente pela escolha de utilizar apenas objetos reconhecidos por um método visual como pista semântica (conforme mencionado anteriormente na Seção 4.3), sem que a distância do robô para estes objetos seja considerada. Fazendo com que seja possível o reconhecimento de objetos presentes em outros ambientes desde que exista visibilidade pelo agente para que o método utilizado o reconheça. Entretanto, espera-se que em trabalhos futuros, utilizando da mesma abordagem mas considerando também outros tipos de propriedades semânticas como pistas (como formato do ambiente, tamanho, textura e etc) seja possível amenizar esse problema.

Em seguida, a utilização de um reconhecedor de objetos de características gerais em

**Tabela 4.10:** Resultados obtidos neste trabalho para cada par de seqüências descritas nas duas primeiras colunas, utilizando o par completo para treinamento e apenas a segunda seqüência do par para teste. As últimas colunas dispõem os números de agrupamentos obtidos e os números de categorias desejados. Freiburg foi abreviado para *Fr* e Saarbrücken para *Sa*. Cloudy ou night representam o estado de iluminação do caminho no momento que a seqüência foi capturada.

1º Sequência	2º Sequência	CE	Acurácia	Nº de Agru- pamentos	Nº de Categorias
Fr (Caminho 1) cloudy 1	Fr(Caminho 2) cloudy 1	0.6537	0.6976	7	7
Fr(Caminho 1) cloudy 1	Sa(Caminho 1) cloudy 2	0.6214	0.8148	4	4
Fr(Caminho 1) cloudy 1	Sa(Caminho 2) night 1	0.4608	0.5128	6	8

conjunto com a baixa qualidade das imagens apresentadas na base de dados acarretou no mau reconhecimento de alguns objetos, confundindo o categorizador OLARFDSSOM (como foi mencionado na avaliação feita na Seção 4.4). Os objetos do conjunto selecionado para reconhecimento também não se mostraram determinantes em algumas categorias, como por exemplo o objeto “Computador de mesa” geralmente presente nas categorias “Escritório de uma pessoa”, “Escritório de duas pessoas” e “Escritório grande”. Em trabalhos futuros, este problema pode ser amenizado com a criação e treinamento de um reconhecedor de objetos específico para a abordagem desenvolvida e pela utilização de novas propriedades semânticas (que considerem diversos tipos de dados sensoriais) além dos objetos como pistas.



**Figura 4.10:** Mapa semântico resultante da seqüência Freiburg (Caminho 2) cloudy 1 utilizada no experimento. Cada coloração nos nodos representa um agrupamento diferente do OLARFDSSOM e os quadrados tracejados delimitando grupos de nodos representam as categorias dos mesmos no *ground truth*.

## 4.7 Experimento Comparativo 2

O segundo experimento segue a mesma linha do anterior, realizando desta vez uma comparação com os resultados de um trabalho da literatura que busca classificar lugares através de informações semânticas visuais utilizando uma variedade de métodos de aprendizagem supervisionada.

Em [Rubio et al. \(2016\)](#) foi exposto um estudo que compara a utilização de classificadores de aprendizagem *offline* e supervisionada baseados em Redes Bayesianas ([Jensen e Nielsen, 2009](#)) e SVM (do inglês, *Support Vector Machines*) ([Cristianini e Shawe-Taylor, 2000](#)) para a tarefa de determinar a classe de lugares utilizando uma pista visual única como informação semântica, o descritor de imagens HOG (do inglês, *Histogram of Oriented Gradients*). Assim como em [Constante et al. \(2013\)](#), no trabalho apresentado por [Rubio et al. \(2016\)](#) a classificação é feita imagem por imagem e utilizando apenas de informações locais contidas na respectiva imagem. Os ambientes também não são mapeados semanticamente.

Desta forma, [Rubio et al. \(2016\)](#) faz a utilização de SVMs com 4 diferentes tipos de funções de *kernel*: linear ( $SVM_{lin}$ ), polinomial ( $SVM_{pol}$ ), radial ( $SVM_{rad}$ ) e sigmoideal ( $SVM_{sig}$ ). Assim como 4 diferentes classificadores Bayesianos, dentre eles duas diferentes versões de *Naive Bayes* (NB), uma discreta ( $NB_d$ ) e uma contínua ( $NB_c$ ). Também foram utilizados um modelo *semi-naive* discreto denominado de TAN (do inglês, *Tree Augmented Naive Bayes*) ([Friedman et al., 1997](#)) e uma rede Bayesiana discreta com aprendizagem estrutural e baseada na métrica  $K2$  ( $BN_{K2}$ ) ([Cooper e Herskovits, 1992](#)).

Com cada um dos classificadores mencionados anteriormente, o autor realiza o processo de validação cruzada 5-fold, com os *folds* estratificados, sobre todas as sequências dos caminhos de número 2 das sub-bases COLD-Freiburg e COLD-Saarbrücken, os quais são os mais desafiadores das respectivas sub-bases e apresenta os resultados em forma de Acurácia, dispostos na Tabela 4.11. Nesse contexto, intencionando aproximar os cenários de experimentação para comparação, foi adaptada uma validação cruzada utilizando cada sequência como sendo um *fold*, onde cada uma delas foi utilizada para teste e o restante para treinamento, sem estratificação. Essa estratégia foi escolhida devido a grande dificuldade em dividir os dados em *folds* estratificados proporcionada pelas características da abordagem desenvolvida no presente trabalho de mapeamento semântico incremental de ambientes com aprendizagem online durante a construção desses mapas.

Portanto, conforme limitações descritas na Seção 4.3, para cada sub-base foram possíveis a utilização de 6 sequências cada, caracterizando uma adaptação da validação cruzada com 6 *folds* (considerados como sequências nesse experimento). Os resultados obtidos com a abordagem estão descritos em forma de Acurácia também na Tabela 4.11, os parâmetros para obtenção dos mesmos foram ajustados com base no CE, todavia tanto o CE quanto a Acurácia foram calculados. A Tabela 4.1 apresentada no início do capítulo, exhibe na coluna “2º Experimento - A” os parâmetros utilizados no experimento com a sub-base COLD-Freiburg e “2º Experimento - B”

**Tabela 4.11:** Resultados apresentados por [Rubio et al. \(2016\)](#) em forma de Acurácia. Na segunda coluna são listados os resultados para o caminho 2 da sub-base COLD-Freiburg e na terceira coluna os resultados para o caminho 2 da sub-base COLD-Saarbrücken. Também foram adicionados os resultados deste trabalho para fins de comparação em perspectiva geral, os valores entre parêntesis representam o desvio padrão.

Modelo	Freiburg (Caminho 2)	Saarbrücken (Caminho 2)
$SVM_{lin}$	0.465	0.449
$SVM_{pol}$	0.404	0.353
$SVM_{rad}$	0.404	0.353
$SVM_{sig}$	0.404	0.353
$NB_c$	0.575	0.497
$NB_d$	0.546	0.529
TAN	0.695	0.726
$BN_{K2}$	0.823	0.844
Este trabalho	<b>0.5429 (0.06)</b>	<b>0.5321 (0.10)</b>

com a sub-base COLD-Saarbrücken. Logo, como pode ser visto na Tabela 4.11, os resultados obtidos com a abordagem desenvolvida se apresentam entre os três mais altos relativos ao caminho da sub-base COLD-Saarbrücken e entre os quatro mais altos relativos ao caminho da sub-base COLD-Freiburg. Uma comparação utilizando teste de hipótese não foi realizada devido a diferença entre as naturezas do trabalho de [Rubio et al. \(2016\)](#) e da abordagem desenvolvida na presente dissertação.

De forma complementar, a Tabela 4.12 apresenta mais detalhes dos resultados obtidos com a abordagem proposta, nela encontram-se os valores de CE e Acurácia resultantes para cada caminho experimentado. A Tabela 4.12 também dispõe os números finais de agrupamentos criados e os números de categorias desejados, os quais se apresentam dentro do mesmo intervalo de valores. Em vista disso, mesmo a abordagem não apresentando nenhum tipo de supervisão, realizando treinamento online e utilizando apenas de objetos reconhecidos como pistas visuais, pode-se afirmar que nesse experimento foram obtidos resultados compatíveis aos apresentados pelos métodos de aprendizagem offline e supervisionada expostos por [Rubio et al. \(2016\)](#), assim como valores esperados de CE e de números de agrupamentos apresentados.

## 4.8 Experimento de Avaliação Incremental

O terceiro e último experimento, diferentemente dos dois anteriores, objetivou demonstrar a capacidade de categorização da abordagem perante sub-grupos de caminhos da base de dados utilizada, possuindo como foco o seu potencial incremental. Para isso, foram realizados dois sub-experimentos utilizando 3 (três) sequências de cada caminho possível das sub-bases, destes 2 (dois) da COLD-Freiburg e 4 (quatro) da COLD-Saarbrücken. Em ambos os sub-experimentos

**Tabela 4.12:** Resultados obtidos com a abordagem proposta nesta dissertação no experimento da Seção 4.7 para os caminhos de número 2 das sub-bases COLD-Freiburg e COLD-Saarbrücken. Os valores entre parêntesis representam o desvio padrão de cada resultado. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados.

<b>Caminho</b>	<b>CE</b>	<b>Acurácia</b>	<b>Nº de Agru- pamentos</b>	<b>Nº de Categorias</b>
Freiburg (Caminho 2)	0.5316 (0.04)	0.5429 (0.06)	7.16 (1.34)	8
Saarbrücken (Caminho 2)	0.4496 (0.07)	0.5321 (0.10)	8.33 (1.24)	8

a abordagem foi executada 30 (trinta) vezes com sequências aleatoriamente ordenadas, média e desvio padrão foram computados.

No primeiro dos sub-experimentos a abordagem foi posta a prova em duas oportunidades. Na primeira delas foi treinada com as sequências de ambas as sub-bases e testada com cada uma das sub-bases de forma isolada (Ambas/Freiburg e Ambas/Saarbrücken)<sup>1</sup>. Em seguida, foi treinada e testada com as sequências de cada sub-base de forma isolada (Freiburg/Freiburg e Saarbrücken/Saarbrücken). Os resultados podem ser encontrados na Tabela 4.13 em valores de CE e Acurácia, bem como os números de agrupamentos obtidos e categorias desejadas. Como pode ser visto, levando em consideração as oportunidades em que a abordagem foi testada com a mesma sub-base que foi treinada (Freiburg/Freiburg e Saarbrücken/Saarbrücken), os resultados de quando a abordagem foi treinada com ambas as sub-bases (Ambas/Freiburg e Ambas/Saarbrücken) se mostram em geral levemente superiores. Portanto, demonstrando resultados iguais ou superiores quando a abordagem é treinada com mais dados, evidenciando sua capacidade de aprendizagem incremental. Isto pode ser confirmado comparando os resultados de CE das condições Ambas/Freiburg e Freiburg/Freiburg, assim como de Ambas/Saarbrücken e Saarbrücken/Saarbrücken, por meio da variação de Welch do teste estatístico T (com normalidade dos dados confirmada por meio dos testes de Kolmogorov-Smirnov e Shapiro-Wilk) unilateral com nível de significância de 5%, não obtendo indícios para rejeitar a hipótese nula, em ambos os casos, de que o resultado com treinamento utilizando ambas as sub-bases é igual ou superior ao treinado com apenas uma sub-base.

O segundo sub-experimento coloca a abordagem em dois cenários, o primeiro limita o número máximo de agrupamentos a serem criados pelo OLARFDSSOM ao número de categorias esperado na base de dados de treinamento e o segundo utiliza um número muito superior ao número de elementos na base de dados, de forma que este limite não afete os resultados. A abordagem foi treinada e testada com todas as sequências selecionadas das sub-bases COLD-Freiburg e COLD-Saarbrücken. Com base nisso, os números máximos de agrupamentos selecionados para os cenários foram 11 (onze) e 10000 (dez mil), respectivamente.

Os resultados obtidos são encontrados na Tabela 4.14 e como pode-se perceber, os valores

<sup>1</sup>Notação: sub-bases de treinamento/sub-bases de teste

**Tabela 4.13:** Resultados médios obtidos com diferentes conjuntos de treinamento e teste. Os valores entre parêntesis representam o desvio padrão de cada resultado. A primeira coluna coloca as sub-bases que foram utilizadas no treinamento e teste, onde Ambas significa a utilização conjunta dos dados da Freiburg e Saarbrücken. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados.

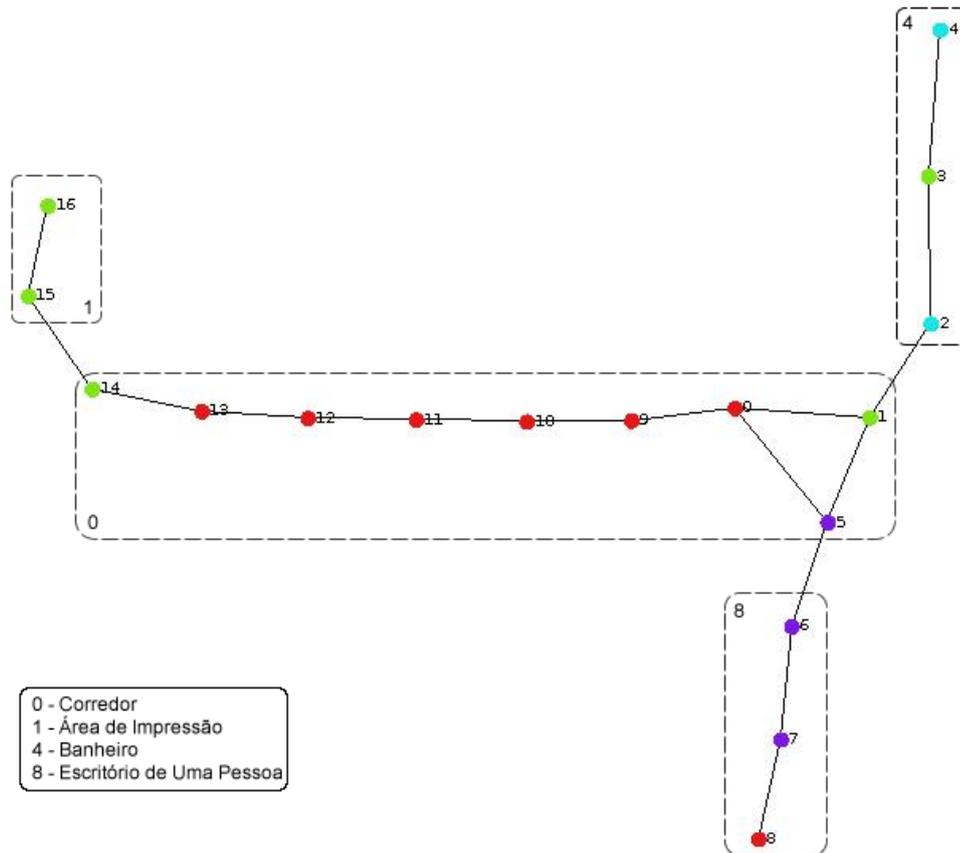
Treinamento/Teste	CE	Acurácia	Nº de Agrupamentos	Nº de Categorias
Ambas/Freiburg	0.6009 (0.02)	0.6785 (0.02)	9.40 (1.54)	11
Freiburg/Freiburg	0.5822 (0.02)	0.6505 (0.02)	8.03 (1.90)	8
Ambas/Saarbrücken	0.4541 (0.02)	0.5401 (0.03)	9.13 (1.74)	11
Saarbrücken/Saarbrücken	0.4352 (0.03)	0.5410 (0.02)	10.46 (2.23)	9

de CE ficaram bastante próximos, assim como os de Acurácia, evidenciando o potencial da abordagem em criar categorias sem qualquer supervisão com resultados bastante equiparáveis aos com número de categorias limitado, consequentemente evidenciando também mais uma característica da sua capacidade incremental. Isso pode ser mais uma vez confirmado através da comparação dos resultados de CE com a variação de Welch do teste estatístico T (com normalidade dos dados confirmada por meio dos testes de Kolmogorov-Smirnov e Shapiro-Wilk) bilateral com nível de significância de 5%, não obtendo indícios para rejeitar a hipótese nula de que os resultados obtidos (em valores de médias) são iguais.

**Tabela 4.14:** Resultados médios obtidos da utilização da abordagem proposta com diferentes números máximo de agrupamentos. Os valores entre parêntesis representam o desvio padrão de cada resultado. As mesmas sequências são utilizadas para treinamento e teste. A primeira coluna coloca os números máximos de agrupamentos utilizados, sendo estes, o número exato de categorias na base de dados utilizada e um valor bastante alto. As últimas colunas dispõem os números de agrupamentos resultantes e os números de categorias desejados.

Nº Máximo de Agrupamentos	CE	Acurácia	Nº de Agrupamentos	Nº de Categorias
11	0.4812 (0.02)	0.5523 (0.02)	8.40 (1.42)	11
10000	0.4782 (0.01)	0.5530 (0.01)	9.66 (1.77)	11

Adicionalmente, de forma ilustrativa é possível visualizar na Figura 4.11 um dos mapas semânticos criados pela abordagem durante sua execução para uma sequência de um caminho da sub-base COLD-Saarbrücken, cada coloração utilizada nos nodos representa um agrupamento diferente no OLARFDSSOM e os quadrados tracejados delimitando grupos de nodos representam as categorias dos mesmos no *ground truth*. Conforme o esperado, esse mapa apresenta o mesmo comportamento da Figura 4.10, especificado na Seção 4.6.



**Figura 4.11:** Mapa semântico resultante de uma sequência do caminho 3 da sub-base COLD-Saarbrücken utilizada no experimento. Cada coloração nos nós representa um agrupamento diferente do OLARFDSSOM e os quadrados tracejados delimitando grupos de nós representam as categorias dos mesmos no *ground truth*.

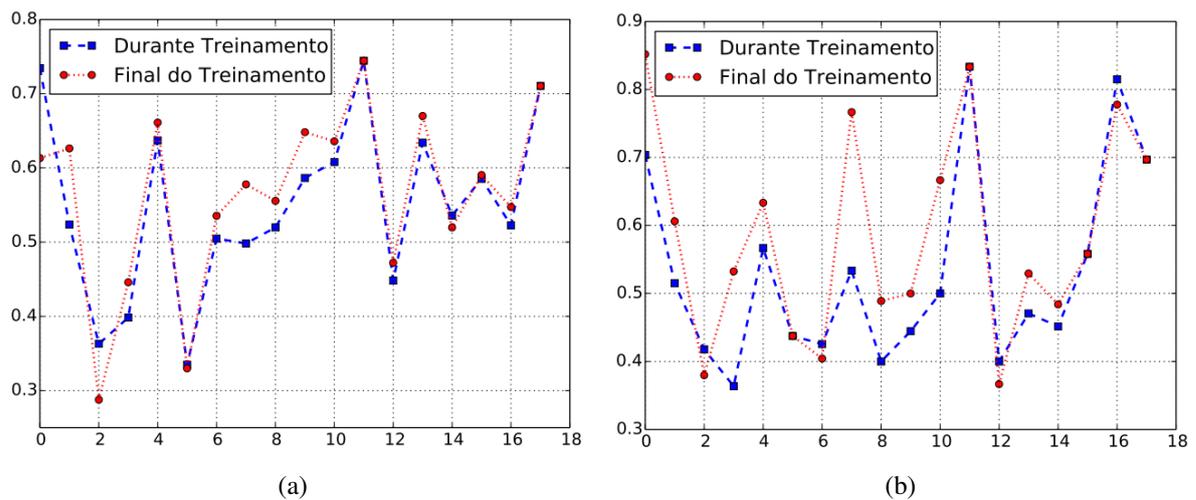
## 4.9 Estudo de Caso

De forma complementar, intencionando expor novas evidências da capacidade incremental da abordagem proposta, foi realizado um estudo de caso onde um conjunto de sequências é treinado pela abordagem e cada uma dessas sequências é testada em dois momentos: logo após o treinamento das mesmas e novamente ao fim de todo o treinamento. Desta maneira, o primeiro conjunto de sequências utilizado na avaliação foi composto de 3 sequências de cada um dos 6 (seis) caminhos possíveis (ordenadas aleatoriamente) das sub-bases, 2 (dois) da COLD-Freiburg e 4 (quatro) da COLD-Saarbrücken, totalizando 18 (dezoito) sequências. Além do já mencionado, mais um conjunto composto pelos mesmos caminhos foi utilizado, entretanto contendo apenas uma sequência de cada (totalizando 6 sequências).

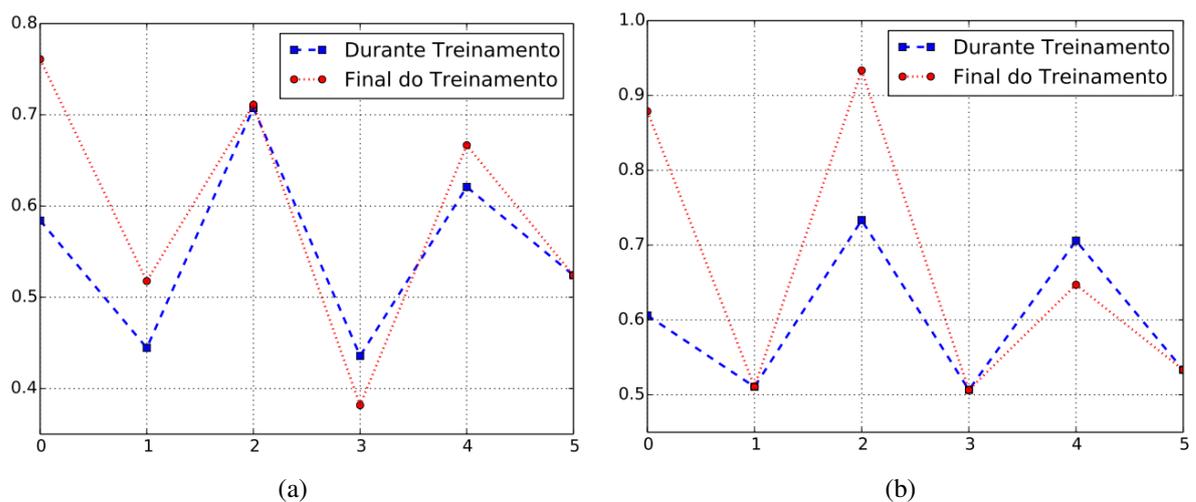
Os resultados obtidos com o primeiro conjunto de sequências são apresentados em forma de gráficos na Figura 4.12(a) em valores de CE e na Figura 4.12(b) em valores de Acurácia. Conforme mostram os resultados, em ambas as medidas de qualidade utilizadas os valores resultantes ao final do treinamento ficaram na grande maioria das vezes com valores iguais ou superiores aos obtidos em testes durante o treinamento (83.4% no caso do CE e 77.8% da

Acurácia) e mesmo nos poucos casos em que foram inferiores o método apresentou valores próximos, demonstrando a capacidade da abordagem em aprender incrementalmente.

As Figuras 4.13(a) e 4.13(b) apresentam os resultados obtidos utilizando o segundo conjunto de sequências mencionado anteriormente, em valores de CE e Acurácia, respectivamente. De forma similar à demonstração anterior, com ambas as medidas de qualidade os valores ao final do treinamento foram na maioria dos casos iguais ou superiores aos durante o treinamento (83.4% para ambas as medidas) e nos poucos casos em que foram inferiores o método apresentou valores próximos aos apresentados durante o treinamento.



**Figura 4.12:** Demonstração do CE (a) e da Acurácia (b) obtidos (eixo y) na realização de testes com cada uma das 18 sequências do primeiro conjunto (eixo x) em dois diferentes momentos, logo após a aprendizagem (treinamento) completa daquela sequência e após a aprendizagem de todas as sequências de caminhos selecionadas.



**Figura 4.13:** Demonstração do CE (a) e da Acurácia (b) obtidos (eixo y) na realização de testes com cada uma das 6 sequências do segundo conjunto (eixo x) em dois diferentes momentos, logo após a aprendizagem (treinamento) completa daquela sequência e após a aprendizagem de todas as sequências de caminhos selecionadas.

## 4.10 Conclusão

Neste capítulo foram descritos os experimentos realizados com a finalidade de avaliar a capacidade da abordagem proposta na criação incremental de mapas semânticos, com lugares categorizados de forma não-supervisionada, através de um método com treinamento online e aprendizagem incremental. Primeiramente, na Seção 4.5 a topologia dos mapas criados foi avaliada, demonstrando coerência com o *ground truth*. Em seguida, nos experimentos das Seções 4.6 e 4.7 foram realizadas comparações em perspectiva geral com os resultados obtidos por [Constante et al. \(2013\)](#) e [Rubio et al. \(2016\)](#), conseguindo a abordagem (dentro das limitações comparativas) demonstrar sua capacidade no contexto de classificação/categorização de lugares com resultados bastante compatíveis com os apresentados pelos trabalhos mencionados. As Seções 4.8 e 4.9, posteriormente, descreveram cenários de experimentação e estudo que tiveram como principal objetivo apresentar indícios do potencial da aprendizagem incremental, não-supervisionada e treinamento online da abordagem no contexto em questão.

No próximo capítulo são feitas as considerações finais desta dissertação, onde estão expostas as conclusões e limitações da abordagem proposta, bem como possíveis trabalhos futuros.

# 5

## Considerações Finais

Nesta dissertação foi apresentada uma abordagem de mapeamento semântico de larga escala, com aprendizagem online e não-supervisionada, baseada em Mapas Auto-organizáveis de topologia variante no tempo. A abordagem se divide no módulo de mapeamento, o SEMMAP, o qual cria mapas topológicos de ambientes incrementalmente, enriquecidos com objetos como informação semântica determinante, e no módulo de categorização de lugares, dotado de um método de aprendizagem incremental, não-supervisionado e treinamento online, o OLARFDSSOM.

O Capítulo 2 introduziu os conceitos que serviram de base para a construção da abordagem proposta. Primeiramente foram apresentadas as principais características e alguns trabalhos na literatura sobre Mapeamento Semântico. Posteriormente foram abordados os Mapas Auto-organizáveis que serviram de base para a abordagem. No Capítulo 3 a abordagem em si foi apresentada em detalhes, dispondo sua arquitetura, fluxo de funcionamento, módulos principais e auxiliares.

A abordagem foi posta à prova no Capítulo 4 em experimentos que demonstraram sua viabilidade e promissora capacidade na aquisição de mapas topológicos enriquecidos com informações semânticas, assim como na categorização dos lugares mapeados a partir de suas informações semânticas. A capacidade de aprendizagem incremental da abordagem, com treinamento online e não-supervisionado, entrou em evidência nos experimentos, demonstrando uma promissora capacidade de aprendizagem em longo termo, validando-a em relação ao problema central de pesquisa definido no Capítulo 1.

A seguir são descritas as principais limitações identificadas na solução proposta, assim como possíveis trabalhos futuros e aplicações.

### 5.1 Limitações

Mesmo com a capacidade demonstrada pela abordagem nos experimentos, limitações já eram conhecidas e algumas outras se evidenciaram na execução dos mesmos. Uma vez que a abordagem proposta não incorpora mecanismos para segmentação de regiões, como um método para detecção de portas, a mesma encontrou problemas na determinação de categorias de nodos

que se encontravam em regiões de transição de tipos de lugares dentro dos ambientes mapeados. Porém, ainda assim, a maior parte dos nodos de forma geral foi categorizada adequadamente.

Um outro fator que contribuiu para a mesma limitação, foi a utilização, como pista semântica única determinante, de apenas objetos reconhecidos através de imagens por um método inteligente de reconhecimento, o qual não foi treinado com imagens específicas para o problema em questão e por esse motivo apresentou baixa confiabilidade. Além da taxa de erro que esses métodos agregam, objetos reconhecidos apenas por informações visuais podem ser reconhecidos a distância, a partir de outros ambientes, mesmo com incerteza. Espera-se que com a inclusão de diferentes tipos de informações semânticas, como dados de sensores que detectem a distância para cada objeto reconhecido, seja possível obter resultados ainda melhores neste quesito.

Um outro problema já conhecido está na estratégia de acumulação de objetos utilizada no SEMMAP. Essa estratégia é inadequada para utilização em ambientes dinâmicos, pois o modelo não é capaz de esquecer objetos vistos anteriormente. Em ambientes dinâmicos, os objetos considerados podem mudar de posição com certa frequência. O estudo de uma estratégia diferente de acumulação é um possível trabalho futuro.

Adicionalmente, a inclusão de rótulos linguísticos associados às categorias identificadas pelo OLARFDSSOM é uma característica presente nos trabalhos da literatura de mapeamento semântico que falta à abordagem proposta. Como o método de categorização é incremental, categorias podem sumir e surgir com o tempo. Devido a isso, a inclusão de tal característica ficou como um estudo a ser realizado como trabalho futuro.

## 5.2 Trabalhos Futuros

A partir da abordagem proposta nesta dissertação, é possível o surgimento de muitos outros trabalhos para, por exemplo, tratar das limitações da mesma, evoluir suas técnicas, usá-la como base para outros modelos ou realizar a utilização da abordagem em aplicações práticas. Algumas dessas possibilidades a serem exploradas de forma imediata são citadas a seguir:

- Incorporação de rótulos linguísticos às categorias criadas incrementalmente pelo OLARFDSSOM, por exemplo, através de consultas ao usuário em linguagem natural;
- Utilização de múltiplas informações semânticas, provenientes de diversas fontes de dados, como formato do lugar, tamanho e textura, na construção dos mapas semânticos;
- Possibilidade de inclusão de novas informações semânticas durante o processo de aprendizagem incremental do OLARFDSSOM;
- Associação do mapa semântico a uma abordagem de comunicação por linguagem natural aplicada em diversos tipos de atividades reais com plataformas robóticas

móveis ou outros tipos de plataformas como cadeiras de rodas;

- Desenvolvimento de módulos auxiliares que façam o reconhecimento de informações semânticas determinantes que possam ser utilizadas pela abordagem na criação dos mapas semânticos.

# Referências

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.  
URL <http://tensorflow.org/>
- Aydemir, A., Gobelbecker, M., Pronobis, A., Sjoo, K., Jensfelt, P., Sept 2011. Plan-based object search and exploration using semantic spatial knowledge in the real world. Em: European Conference on Mobile Robotics. pp. 13–18.
- Bassani, H., 2014. Modelos neurais de aquisição de linguagem natural para agentes incorporados. Tese de Doutorado, Universidade Federal de Pernambuco.
- Bassani, H. F., Araujo, A. F. R., June 2012. Dimension selective self-organizing maps for clustering high dimensional. Em: International Joint Conference on Neural Networks. IEEE, pp. 1–8.
- Bassani, H. F., Araujo, A. F. R., 2015. Dimension selective self-organizing maps with time-varying structure for subspace and projected clustering. IEEE Transactions on Neural Networks and Learning Systems 26 (3), 458–471.
- Bastianelli, E., Bloisi, D. D., Capobianco, R., Cossu, F., Germignani, G., Iocchi, L., Nardi, D., Nov 2013. On-line semantic mapping. Em: International Conference on Advanced Robotics. IEEE Computer Society, pp. 1–6.
- Case, C., Suresh, B., Coates, A., Ng, A. Y., May 2011. Autonomous sign reading for semantic mapping. Em: International Conference on Robotics and Automation. IEEE, pp. 3297–3303.
- Constante, G., Ciarfuglia, T. A., Valigi, P., Ricci, E., Nov 2013. A transfer learning approach for multi-cue semantic place recognition. Em: International Conference on Intelligent Robots and Systems. pp. 2122–2129.
- Cooper, G. F., Herskovits, E., 1992. A bayesian method for the induction of probabilistic networks from data. Machine Learning 9 (4), 309–347.
- Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.
- Deeken, H., Wiemann, T., Lingemann, K., Hertzberg, J., Sept 2015. Semap - a semantic environment mapping framework. Em: European Conference on Mobile Robots (ECMR). pp. 1–6.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. Em: CVPR09.  
URL <http://image-net.org/>
- Duvallet, F., 2015. Natural language direction following for robots in unstructured unknown environments. Tese de Doutorado, Carnegie Mellon University.

- Dzifcak, J., Scheutz, M., Baral, C., May 2009. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. Em: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 4163–4168.
- Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. *Machine Learning* 29 (2), 131–163.
- Fritzke, B., 1994. Growing cell structures—a self-organizing network for unsupervised and supervised learning. *Neural Networks* 7 (9), 1441 – 1460.
- Gemignani, G., Capobianco, R., Bastianelli, E., Bloisi, D. D., Iocchi, L., Nardi, D., 2016. Living with robots: Interactive environmental knowledge acquisition. *Robotics and Autonomous Systems* 78, 1 – 16.
- Haykin, S., 1998. *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Helton, J. C., Davis, F. J., Johnson, J. D., 2005. A comparison of uncertainty and sensitivity analysis results obtained with random and latin hypercube sampling. *Reliability Engineering and System Safety* 89, 305–330.
- Hemachandra, S., Walter, M. R., Tellex, S., Teller, S., May 2014. Learning spatial-semantic representations from natural language descriptions and scene classifications. Em: International Conference on Robotics and Automation. pp. 2623–2630.
- Jensen, F. V., Nielsen, T. D., 2009. *Bayesian networks and decision graphs*. Springer Science and Business Media.
- Karaoğuz, H., Bozma, H., 2016. An integrated model of autonomous topological spatial cognition. *Autonomous Robots* 40 (8), 1379–1402.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43 (1), 59–69.
- Kollar, T., Perera, V., Nardi, D., Veloso, M., May 2013. Learning environmental knowledge from task-based human-robot dialog. Em: International Conference on Robotics and Automation. pp. 4304–4309.
- Kostavelis, I., Gasteratos, A., 2013. Learning spatially semantic representations for cognitive robot navigation. *Robotics and Autonomous Systems* 61 (12), 1460 – 1475.
- Kostavelis, I., Gasteratos, A., 2015. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems* 66, 86 – 103.
- Kostavelis, I., Gasteratos, A., 2017. Semantic maps from multiple visual cues. *Expert Systems with Applications* 68, 45 – 57.
- Krishnan, A. K., Krishna, K. M., Oct 2010. A visual exploration algorithm using semantic cues that constructs image based hybrid maps. Em: International Conference on Intelligent Robots and Systems. IEEE, pp. 1316–1321.
- Kunze, M., Steffens, J., 1995. Growing cell structure and neural gas - incremental neural networks. Em: Proceedings of the Fourth AIHEP Workshop. World Scientific.

- Lauritzen, S., Richardson, T., 2002. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3), 321–348.
- Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Em: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. pp. 2169–2178.
- Marsland, S., Shapiro, J., Nehmzow, U., 2002. A self-organising network that grows when required. *Neural Networks* 15 (8–9), 1041 – 1058.
- Müller, E., Günnemann, S., Assent, I., Seidl, T., Agosto 2009. Evaluating clustering in subspace projections of high dimensional data. *Proc. VLDB Endow.* 2 (1), 1270–1281.
- Oh, J., Suppe, A., Duvallet, F., Boularias, A., Vinokurov, J., Navarro-Serment, L., Romero, O., Dean, R., Lebiere, C., Herbert, M., Stentz, A., 2015. Toward mobile robots reasoning like humans. Em: *AAAI Conference on Artificial Intelligence*.
- Pronobis, A., 2011. Semantic mapping with mobile robots. Tese de Doutorado, KTH, School of Computer Science and Communication.
- Pronobis, A., Caputo, B., 2009. Cold: The cosy localization database. *The International Journal of Robotics Research* 28 (5), 588–594.
- Pronobis, A., Jensfelt, P., May 2012. Large-scale semantic mapping and reasoning with heterogeneous modalities. Em: *International Conference on Robotics and Automation. IEEE*, pp. 3515–3522.
- Pronobis, A., Mozos, O. M., Caputo, B., Jensfelt, P., 2010. Multi-modal semantic place classification. *International Journal of Robotics Research* 29 (2-3), 298–320.
- Rubio, F., Martínez-Gómez, J., Flores, M. J., Puerta, J. M., 2016. Comparison between bayesian network classifiers and svms for semantic localization. *Expert Systems With Applications* 64, 434 – 443.
- Saltelli, A., Chan, K., Scott, E. M., 2009. *Sensitivity Analysis*. Wiley, New York, NY, USA.
- Sunderhauf, N., Dayoub, F., McMahon, S., Talbot, B., Scultz, R., Corke, P., Wyeth, G., Upcroft, B., Milford, M., May 2016. Place categorization and semantic mapping on a mobile robot. Em: *International Conference on Robotics and Automation (ICRA)*. pp. 5729–5736.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the inception architecture for computer vision. *CoRR abs/1512.00567*.
- Tellex, S., Kollar, T., Dickerson, S., 2011. Understanding natural language commands for robotic navigation and mobile manipulation. Em: *National Conference on Artificial Intelligence (AAAI)*.
- Walter, M. R., Antone, M., Chuangsuwanich, E., Correa, A., Davis, R., Fletcher, L., Frazzoli, E., Friedman, Y., Glass, J., How, J. P., Jeon, J., Karaman, S., Luders, B., Roy, N., Tellex, S., Teller, S., 2015. A situationally aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments. *Journal of Field Robotics* 32 (4), 590–628.
- Walter, M. R., Hemachandra, S., Homberg, B., Tellex, S., Teller, S., 2014. A framework for learning semantic maps from grounded natural language descriptions. *The International Journal of Robotics Research* 33 (9), 1167–1190.