

APLICAÇÕES DE SISTEMAS MULTIAGENTES E APRENDIZAGEM AUTOMÁTICA NO PROCESSAMENTO DA LINGUAGEM NATURAL

Ricardo Annes
PUCRS – Campus II
BR 472 Km 7 - Sala 24
97.510-460 Uruguaiana, RS
Brasil
annes@pucrs.campus2.br

RESUMO

A proposta desse trabalho é o estudo de três áreas da Inteligência Artificial: Aprendizagem Automática, Processamento da Linguagem Natural e Sistemas Multiagentes. Além do estudo isolado de cada área, foi feita uma pesquisa das intersecções que existem entre essas áreas: Aprendizagem Automática para Processamento da Linguagem Natural, Aprendizagem em Sistemas Multiagentes e Sistemas Multiagentes no Processamento da Linguagem Natural. Nas conclusões, discutimos as potencialidades da aplicação de Sistemas Multiagentes no Processamento baseado em *corpus*.

Palavras chaves: Inteligência Artificial, Sistemas Distribuídos.

1. INTRODUÇÃO

Desde os primórdios da computação, o Homem precisou “aprender” a linguagem dos computadores, códigos binários, fórmulas matemáticas, linguagem gráfica, hoje o grande desafio é transformar essa comunicação no mais natural possível. A idéia do Processamento da Linguagem Natural é realizar um tratamento sistemático da Linguagem Natural, com o objetivo de os computadores entenderem a nossa linguagem.

A área é atuante e fecunda. Sistemas de computação inteligente, sistemas automáticos de tradução, sistemas de análise de textos, sistemas de instrução assistida por computador e consultas a Base de Dados em Linguagem Natural, são alguns exemplos.

Na seção 2 abordamos o Processamento da Linguagem Natural, seus conceitos básicos e o processamento baseado em *corpus*. Na seção 3 apresentamos um estudo dos Sistemas Multiagentes. Na seção 4 vamos descrever as principais técnicas da Aprendizagem Automática e na seção 5 apresentamos nossa conclusão.

2. PROCESSAMENTO DA LINGUAGEM NATURAL

O Processamento da Linguagem Natural (PLN), sub-área da Inteligência Artificial, é o encontro de duas áreas com enfoques distintos e problemas semelhantes. Por um lado a Computação necessita de interfaces mais naturais e por outro a Lingüística precisa dos métodos computacionais para desenvolver suas teorias.

O tratamento das bases de dados textuais que suportam o acervo de conhecimento da Humanidade, expressos em linguagem natural, a recuperação de informações, as interfaces Homem-

Máquina, traduções automáticas, a compreensão e geração de linguagem são determinantes à evolução do PLN.

Do ponto de vista da Computação, o Processamento da Linguagem Natural, pode ser sintetizado em [DE LIMA96]:

compreensão (reconhecimento exato) e

produção (geração natural e apropriada) da linguagem.

A linguagem natural, em contrapartida à linguagem artificial ou formal, é algo já existente que tem como objetivo a interação interpessoal. Já as linguagens artificiais (FORTRAN, COBOL, C) tem um propósito específico e são mais restritivas.

O tratamento computacional da linguagem natural envolve vários domínios de processamento: domínio léxico-morfológico, sintático e semântico-pragmático [DE LIMA97], nas próximas seções discutiremos sobre cada um deles.

O léxico ou dicionário, é a estrutura de dados que acompanha o processamento da análise e geração da linguagem natural, armazena as palavras e associa a elas informações relevantes. Sucintamente pode ser definido como uma lista de palavras com suas classes sintáticas.

Um dicionário de formas contém a estrutura completa das palavras. É utilizado para o reconhecimento das palavras, sem a preocupação do trato lingüístico. Um dicionário de bases contém apenas morfemas (partes das palavras, radical, prefixo, sufixo, ...), os quais através de modelos, são usados para compor as palavras. Nessa estrutura, normalmente existe uma maior organização e condensação.

No domínio léxico-morfológico o tratamento é realizado sobre as palavras de uma sentença e obtêm as diferentes categorias de cada palavra e outras informações disponíveis no léxico. A quantidade de informações para cada entrada do léxico é proporcional à riqueza da análise léxico-morfológica [DE LIMA96].

Para os mesmos tipos de palavras (categorias gramaticais), existem registros que caracterizam o comportamento de um subconjunto de vocábulos da linguagem.

A morfologia estuda o tratamento da estrutura das palavras (forma, flexões, classificação) em relação a cada tipo de palavra.

Um fenômeno, muito estudado, que ocorre nesse domínio é a ambigüidade léxica. Quando a mesma palavra pode ter mais de uma representação ou situa-se em diferentes categorias gramaticais, dizemos que ocorre uma ambigüidade léxica [SILVA97]. O exemplo a baixo descreve esse fenômeno:

Maria ocupou a casa.

a - pode ser:
artigo,
preposição ou
pronomes.

ambigüidade léxica categorial

casa - pode ser:

residência ou

divisão do tabuleiro de xadrez.

ambigüidade léxica semântica

A **sintaxe** estuda a disposição das palavras para formação da estrutura da sentença. O *parser* é um programa que a partir de uma gramática, um léxico e um texto, determina quais construções estão de acordo com as regras gramaticais e gera a descrição da estrutura. Se a sentença for ambígua, o *parser* descreve todas as estruturas possíveis para a sentença. Essa estrutura hierárquica de informações é chamada de árvore de derivação; na sua construção é feita a verificação da adequação das seqüências de palavras ou termos que compõem a frase, período ou oração, às regras de formação impostas pela linguagem, tais como:

- concordância nominal e/ou verbal;
- regência nominal e/ou verbal;
- posicionamento dos termos na frase.

Da decomposição realizada pelo analisador sintático surgem elementos funcionais da oração, denominados **sintagmas** (unidades sintáticas): termos essenciais (sujeito e predicado), termos integrantes (complementos verbal e nominal) e termos acessórios (adjunto adverbial, adjunto adnominal e aposto). A análise sintática do período deve considerar também o tipo de período (simples ou composto), sua composição (por subordinação, por coordenação) e a classificação das orações (absolutas, principal, coordenada ou subordinada).

Com esse fim o analisador sintático utiliza dois componentes:

1 - uma representação gramatical onde estão declarados os fatos sintáticos da linguagem, as estruturas de composição dos sintagmas;

2 - um procedimento de análise, responsável por verificar se a frase de entrada está de acordo com as regras gramaticais e gerar uma estrutura hierárquica, na forma de uma árvore, representando a estrutura da frase analisada.

Formalmente, uma linguagem é um conjunto de sentenças, onde cada sentença é a concatenação de um ou mais símbolos (palavras) do vocabulário desta linguagem. Uma gramática é uma especificação finita deste conjunto.

A prototipação de analisadores para reconhecer sentenças, é freqüentemente feita usando uma notação conhecida como *Definite Clause Grammar* (DCG) que é disponível em PROLOG. Com essa notação podemos definir gramáticas livres do contexto analisar sentenças (*parsing*) [BEARD91].

No domínio léxico-morfológico, a categorização gramatical das palavras provê um tratamento semântico inicial. No domínio sintático, a dependência (concordância verbal, por exemplo), ao ser tratada gramaticalmente, com formalismos mais complexos, impõe restrições semânticas para determinar se uma sentença é bem formada ou não.

Nesse aspecto, a análise sintática/semântica da linguagem natural é semelhante a linguagem de programação (linguagem artificial) onde a sintaxe é uma expressão da forma das sentenças válidas, e a semântica invalida construções bem formadas mas sem sentido.

O processamento da associação de um ou mais representações semânticas às árvores de derivação, denominado análise semântica, pode ocorrer à medida que estas árvores vão sendo produzidas, ou pode ser feita em etapa posterior. Há tendência de resolução da análise sintática e da análise semântica em cooperação.

Por exemplo, a sentença:

João comprou um Fusca. Ele gosta de carros.

exige interação entre essas duas etapas para ser resolvido o problema da anáfora (**Ele**).

Enquanto a análise semântica se restringe a lidar com os significados das sentenças a partir dos significados de suas partes, a pragmática preocupa-se com a interpretação do todo, que não é somente a união das partes. É necessário a resolução de fenômenos de referência, tais como anáforas e elipses.

2.1. PROCESSAMENTO BASEADO EM CORPUS

O emprego de heurísticas para a escolha de alternativas para a análise sintática, tem sido muito utilizada. Uma dessas abordagens, o processamento baseado em *corpus*, usa técnicas baseadas na Teoria da Probabilidade.

Nos últimos anos, grandes bases de dados textuais (*Brown Corpus*, *Folha de São Paulo*, *AP News*, *Wall Street Journal*) tornaram-se disponíveis para o PLN. Nessas bases são encontrados vários milhões de palavras que podem ser processadas automaticamente.

Os primeiros esforços da utilização de métodos empíricos e estatísticos no processamento da linguagem, datam da década de 1950. A célebre frase “You shall know a word by the company it keeps” [Firth apud CHUR93] já bem demonstrava essa abordagem.

Um dos maiores atrativos das abordagens baseadas em estatística é a habilidade da efetiva aprendizagem de parâmetros a partir do processamento do *corpus*. Esses algoritmos começam com uma estimativa inicial das probabilidades e então processam o *corpus* para calcular a melhor estimativa, repetindo o procedimento até que não hajam mais melhorias. A técnica garante a convergência, embora não necessariamente descubra um valor ótimo.

A abordagem baseada em conhecimento pura emula conhecimento da fala humana usando técnicas oriundas dos Sistemas Especialistas. Sistemas baseados exclusivamente em regras tem obtido somente sucesso limitado. Muitos dos sistemas bem sucedidos, atuais, usam a abordagem estocástica [CHUC93].

Um *corpus* é uma coleção de peças (pedaços, fragmentos) da linguagem que são selecionados e ordenados de acordo com um critério lingüístico explícito para ser usado como um exemplo da linguagem. O critério lingüístico pode ser externo (quando diz respeito aos participantes, a ocasião, classe social ou a função comunicativa das peças da linguagem) ou interno, (quando está relacionado à ocorrência de padrões da linguagem dentro das peças da linguagem).

2.2. ETIQUETAÇÃO

Um das técnicas estatísticas muito utilizadas no PLN é o **Modelo de Markov**, usado para a etiquetagem de *corpus*. Esse modelo é uma máquina de estados finitos que possibilita regular transições entre estados e controlar a emissão de sinais de saída.

A etiquetagem de categorias das palavras, envolve a seleção das seqüências das categorias sintáticas mais parecidas com as palavras numa sentença.

Dada uma sentença com palavras ambíguas, a tarefa é determinar a categoria léxica mais provável de cada palavra. Na implementação desse algoritmo deve-se sempre optar por um conjunto de treinamento [ALLEN94]. Essa técnica frequentemente obtém mais de 90% de sucesso, principalmente porque mais da metade das palavras que aparecem no *corpus* não são ambíguas.

O processo pode ser modelado por uma configuração especial de máquina de estados finitos probabilística. Cada nodo representa a possível categoria léxica e a possibilidade de transições que indica a probabilidade de uma categoria seguir outra. Com uma rede como essa, pode-se computar a probabilidade de qualquer seqüência de categorias simplesmente pela descoberta do caminho através da rede indicada pela seqüência e multiplicando pelas probabilidades de transições. Essa representação só é exata se a probabilidade de cada categoria ocorrida depende só de uma categoria anterior. Na Teoria da Probabilidade isso é chamado de **Suposição de Markov** (*Markov Assumption*) e essa rede chamada de **Cadeia de Markov** (*Markov Chains*).

Segundo Biber [BIBER93], os principais usos das técnicas estatísticas no Processamento da Linguagem Natural são:

- * Gramaticais - prover descrições gramaticais de características lingüística particulares;
- * Etiquetagem probabilística e *Parsing*;
- * Lexicografia - uso e significado das palavras e padrões lexicais.

E ainda, prevê como futuras pesquisas e aplicações:

- * Predição automática de Registros de Categoria e
- * Comparação inter-lingüística;

ambas consideradas etapas preliminares às tarefas de Recuperação de Informações (*Information Retrieval*) e Tradução Automática (*Machine Translation*).

3. SISTEMAS MULTIAGENTES

A Inteligência Artificial Distribuída e os Sistemas Multiagentes obtiveram um acentuado crescimento de importância como modelo na Ciência da Computação, bem como em outras ciências, Lingüística, Lógica Matemática, Teoria da Organização, Biologia, Sociologia e Psicologia [WERN92]. Isso demonstra o forte entrelaçamento da Inteligência Artificial Distribuída com disciplinas de outras áreas.

O interesse pela Inteligência Artificial Distribuída advém da grande ênfase no entendimento da cooperação e organização de grupos e sociedades de agentes que são naturalmente

atividades sociais (sentido humano), apoiando atividades complexas, interdisciplinares e distribuídas (lógica e geograficamente). Tais atividades incluem: simulação de capacidades inteligentes, raciocínio, comunicação em linguagem natural e aprendizagem.

“A Inteligência Artificial Distribuída (IAD) é a subárea da Inteligência Artificial concernente à concorrência em computadores de Inteligência Artificial, em diversos níveis” [BOND88]. O enfoque da IAD está nos aspectos de interação, cooperação e no fluxo de conhecimento entre unidades logicamente distintas. A IAD caracteriza-se como um outro paradigma, onde o comportamento inteligente é visto como resultado das interações entre agentes.

A Inteligência Artificial Distribuída, tradicionalmente é dividida em subáreas, caracterizadas pela abordagem na definição dos agentes e sociedades:

* DPS (Distributed Problem-Solving) - A Resolução Distribuída de Problemas considera que o trabalho de resolver um problema particular pode ser dividido entre um número de módulos que cooperam para dividir e compartilhar conhecimento sobre o problema e sobre o desenvolvimento da solução. Todas as interações (cooperação, coordenação) estratégicas são incorporadas como parte integral do sistema.

* MAS (Multi-Agent Systems) - Os Sistemas Multiagentes dizem respeito à coordenação de um comportamento inteligente entre uma coleção de agentes (normalmente já existentes) com o qual podem coordenar seus conhecimentos, objetivos, capacidades e planos com o intuito de resolver um problema ou tomar uma ação.

4. APRENDIZAGEM

Os métodos de Aprendizagem Automática (Machine Learning) desempenham um papel muito importante na área da Inteligência Artificial, nos dias de hoje, Resolução de Problemas, Prova de Teoremas, Processamento da Linguagem Natural, Robótica e Sistemas Especialistas são sub áreas que utilizam tais métodos.

A habilidade de aprender, de adaptar-se, de modificar o comportamento é um componente inalienável da inteligência humana. Como poderemos construir artefatos verdadeiramente inteligentes, capazes de auto melhora ?

Considerações filosóficas à parte, a área da Aprendizagem Automática tem obtido grandes progressos. Indução em arvores de decisão aplicadas ao controle de processos industriais [CARB89], modelos de Markov escondido aplicados ao reconhecimento da fala são alguns exemplos.

Vamos considerar aprendizagem como mudanças adaptáveis no sistema, no sentido de que o permitam, da próxima vez, realizar a mesma tarefa ou tarefas do mesmo grupo com mais eficiência e eficácia [Simon apud RICH93].

Atualmente podemos agrupar em cinco os paradigmas de Aprendizagem Automática:

- * Aprendizagem Indutiva (Inductive Learning),
- * Aprendizagem Analítica (Analytic Learning),
- * Aprendizagem por Reforço (Reinforcement Learning),
- * Algoritmos Genéticos (Genetic Algorithms) e

* Métodos de Aprendizagem Conexionistas (Connectionist Learning Methods).

Grandes esforços, com relativo sucesso, tem-se alcançado com abordagens híbridas, unindo métodos de mais de um paradigma.

4.1. APRENDIZAGEM INDUTIVA

A Aprendizagem Indutiva [CARB89] é um dos métodos de aprendizagem simbólicos mais estudados, partindo de uma seqüência de instâncias de conceitos e opcionalmente contra-exemplos, induz-se descrições gerais de conceitos.

A tarefa é construir a descrição de um conceito a partir do qual instancias de exemplos positivos podem ser derivadas mas as instancias de exemplos negativos não possam ser derivadas.

Algumas dimensões importantes, relativas a Aprendizagem Indutiva são:

* Linguagem de descrição: a linguagem na qual as instâncias de entrada e os conceitos de saída serão expressas podem variar em poder de representação (calculo proposicional, lógica de primeira ordem, ou outras). Muitos sistemas adotam um vocabulário fixo de hipóteses no qual todos os descritores relevantes precisam estar presentes no conjunto de saída.

* Classificação e ruído das instâncias: as instâncias de entrada não podem ser simplesmente classificadas de forma binária, em positivas e negativas, normalmente nas aplicações do mundo real, encontramos instâncias inexatas e até não nomeadas (classificadas), criando um ruído de entrada no processo.

* Indução incremental ou não-incremental: nos modelos não-incremental são consideradas todas as instâncias positivas e negativas que serão usadas como dados do treinamento de uma só vez e produz-se a descrição de um conceito não aberto para futuras modificações. A técnica incremental produz o conceito de melhor suposição (*best-guess*) ou uma faixa de conceitos consistentes com os dados que mais tarde podem interagir com novas instâncias. Esse modelo reflete maior fidelidade com situações do mundo real nos quais a aprendizagem é um processo contínuo.

Uma árvore de decisão [RUSS95] é uma estrutura simples de aprendizagem indutiva, dada uma instância de um objeto ou situação a qual é especificada por um conjunto de propriedades, a árvore retorna a decisão "sim" ou " não" sobre essa instância. Cada nodo interno na árvore representa um teste de uma dessas propriedades e os nodos ramos dos nodos são rotulados com as possíveis respostas dos testes. Cada nodo folha é uma classificação Booleana da instância de entrada.

Um exemplo clássico de sistema de aprendizagem indutiva é o ID3 de Quinlan, que desenvolveu em 1979, um sistema de aprendizagem com árvore de decisão usando ganho de informação, no qual o atributo escolhido é calculado a fim de obter-se o maior ganho na divisão desses atributos. Atualmente existe uma versão industrial do ID3, o C4.5 e o ID5 de Utgoff que estendeu o ID3 permitindo classificação multi valorada, incorporando o conceito de aprendizagem incremental ao modelo.

4.2. APRENDIZAGEM ANALÍTICA

Esse paradigma é baseado numa aprendizagem analítica a partir de poucos exemplos (frequentemente um só) mais uma teoria inicial básica, também chamada de Teoria do Domínio (ou background knowledge). Os métodos envolvidos são mais dedutivos do que indutivos.

Os métodos analíticos estão focados para a melhoria da eficiência dos sistemas sem sacrificar a exatidão ou generalidade.

EBL (explanation-based learning) é um método para extração de regras gerais a partir de observações individuais [RUSS95]. A idéia básica do EBL é primeiro construir uma explanação das observações usando conhecimento inicial, e então estabelecer uma definição das classes dos casos para os quais a mesma estrutura da explanação pode ser usada. Essa definição prove a base para a regra cobrir todos os casos da classe.

4.3. ALGORITMOS GENÉTICOS

Os algoritmos genéticos [CARB89] que foram inspirados diretamente da analogia das mutações nas reproduções biológicas, representam uma posição bastante empírica da Aprendizagem Automática.

A partir das Teorias de Darwin sobre a seleção natural das espécies (sobrevivência dos mais ajustados em cada nicho ecológico).

Algoritmos Genéticos [RUSS95] buscam num espaço de indivíduos por bons candidatos. A boa qualidade de um indivíduo é medida por algumas funções de ajuste (função objetivo). A busca é feita em paralelo com muitos indivíduos e gerações. A abordagem é *hill-climbing*, visto que em cada geração a descendência dos melhores candidatos são preservados. Cada indivíduo é um conjunto de bits que codifica suas características, cada elemento desse conjunto é chamado de **gene** e o conjunto de bits é como uma **cadeia de DNA**.

O algoritmo consiste de um *loop* através de gerações e em cada geração, um sub conjunto da população é selecionada para reproduzir, a seleção é feita por substituição (os indivíduos ajustados podem reproduzir mais vezes). A reprodução ocorre por acasalamento randômico de todos os pares do conjunto selecionado e gerando dois novos indivíduos por *cross-over*. Também, existe uma pequena chance de haver **mutações** nos genes de um indivíduo.

4.4. APRENDIZAGEM CONEXIONISTAS E MÉTODOS ESTATÍSTICOS

Os Métodos de Aprendizagem Conexionistas, [CARB89] também chamados de Redes Neurais, distinguem entre classes de padrões a partir de um domínio de entrada de maneira holística. Elas recebem como conjuntos de instâncias representativas de cada classe, etiquetadas (com algum ruído tolerado), aprendem a reconhecer essas e outras instâncias de cada classe representativa. A aprendizagem consiste em ajustar os pesos em uma rede de topologia fixa através de diferentes algoritmos de aprendizagem tais como *Bolzmans* ou *backpropagation*.

Uma tendência recente para a aprendizagem em processamento da linguagem natural é a utilização de métodos estatísticos em grandes massas de dados. Sobre esses *corpus* são feitas análises estatísticas das características do texto sem inicialmente adicionar significado lingüístico ou do conhecimento do mundo.

Anotações de *corpus* com etiquetas categoriais ou árvores de *parse* são tido muita atenção nas análises lingüísticas baseadas em *corpus*. Também são de importância atual as técnicas estatísticas para construção de dicionários de sinônimos, aprendizagem probabilística de gramáticas, técnicas de n-gram e modelos escondidos de Markov.

4.5. APRENDIZAGEM EM SISTEMAS MULTIAGENTES

Sistemas Multiagentes tipicamente tem estrutura e funcionamento complexos, sendo difícil ou até impossível determinar corretamente o repertório de comportamentos em tempo de projeto. Essa complexidade pode ser reduzida ou evitada dotando os agentes com capacidades de aprendizagem [WEIB95].

Não existe uma distinção explícita entre adaptação e aprendizagem, no entanto, é assumido que adaptação é coberta por aprendizagem.

4.6. CATEGORIAS DE APRENDIZAGEM

Num Sistemas Multiagentes a aprendizagem pode realizar-se de forma isolada ou interativa, quando os agentes aprendem de forma mútua, coletivamente. Na aprendizagem isolada os agentes aprendem de forma solitária e completamente independentes uns dos outros. Na aprendizagem coletiva, ou interativa, os agentes trocam informações, compartilham hipóteses, geralmente desenvolvem pontos de vista de seus ambientes, normas que regulam e restringem seus comportamentos e interações. A aprendizagem simples e a coletiva constituem-se a principal categoria de aprendizagem em Sistemas Multiagentes e é de difícil delinear a fronteira entre ambas, usualmente quando se refere a aprendizagem em SMA, estamos nos referindo a aprendizagem simples. Também são encontrados outros termos para a expressão, como: a mútua, a cooperativa, a colaborativa, co-aprendizagem, aprendizagem compartilhada, aprendizagem social, aprendizagem pluralista, aprendizagem organizacional.

4.7. APRENDIZAGEM PARA PROCESSAMENTO DA LINGUAGEM NATURAL

Nos últimos anos floresceu, um grande interesse e atividades no desenvolvimento de novas abordagens de aprendizagem para o Processamento da Linguagem Natural: Redes Conexionistas/Redes Neurais, Métodos Estatísticos, Algoritmos de Aprendizagem Simbólica, Métodos Genéticos, e várias abordagens híbridas.

A utilização dos métodos de aprendizagem para o PLN é especialmente importante porque a aprendizagem incorpora avanços tecnológicos em muitas tarefas relacionadas a linguagem, tais como reconhecimento da fala, compreensão da linguagem falada, tradução automática e recuperação de informação. Outro fator é o acréscimo de flexibilidade, adaptabilidade, escalabilidade e portabilidade aos sistemas de processamento de linguagem natural.

Podemos dividir os métodos de aprendizagem para o PLN em dois grupos: Abordagem Simbólica e Abordagem Conexionista.

Os métodos de aprendizagem simbólica já citados, algoritmos de árvore de decisão, aprendizagem baseada em explanação, métodos de raciocínio analógico tem sido utilizados no processamento da linguagem natural.

Os avanços alcançados com os métodos de aprendizagem indutiva são usados com sucesso por pesquisadores em PLN em resolução de anáforas, problemas de ambigüidade, categorização de textos e conhecimento semântico.

As arquiteturas conexionistas, redes neuronais, são mecanismos muito utilizados no processamento da linguagem natural, por terem mostrado resultados de muito sucesso.

Algoritmos genéticos de aprendizagem trabalham em expressões lógicas as quais podem representar gramáticas livres do contexto sem recursão [Smith apud WERM96].

5. CONCLUSÕES

Esse estudo abordou três áreas:

- * Aprendizagem Automática,
- * Processamento da Linguagem Natural e
- * Sistemas Multiagentes ,

todas abrigadas no seio da Inteligência Artificial e amplamente estudadas, investigadas e com farto material disponível.

Nossa proposta inicial era o estudo da intersecção dessas três áreas, conforme ilustra a figura 1:

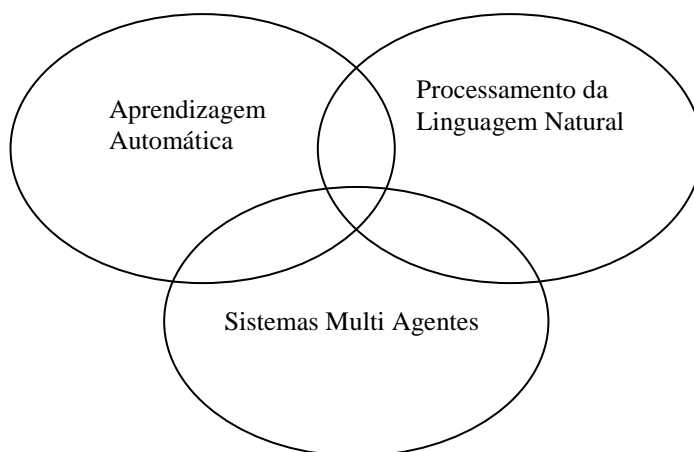


Figura 1 - Áreas estudadas.

Tomadas duas a duas, as áreas tem uma intersecção bastante forte, formando campos de pesquisa fecundos e agrupando grande quantidade de pesquisadores em projetos, *workshop* e conferências:

- Aprendizagem Automática para Processamento da Linguagem Natural,
- Aprendizagem em Sistemas Multiagentes e
- Sistemas Multiagentes no Processamento da Linguagem Natural.

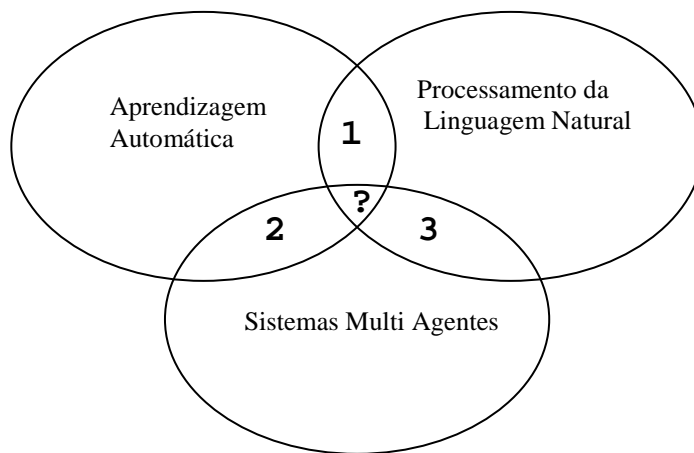


Figura 2 - Intersecção da áreas estudadas.

Já a intersecção das três áreas, embora pareça clara e de óbvia aplicabilidade, não foi encontrada. Pesquisas bibliográficas, revistas, *proceeding* de conferências e *workshop* e a Internet foram consultadas sem sucesso.

Diante desse quadro, põe-se diante de nós um grande desafio atingirmos a intersecção das três áreas na busca de elementos já existentes ou até na proposta de uma arquitetura que aproveite as vantagens de cada área numa abordagem global.

Dessa forma pretendemos prosseguir no estudo da intersecção das três áreas para tentar descobrir a aplicabilidade da **Aprendizagem Automática** em **Sistemas Multiagentes** no **Processamento da Linguagem Natural**.

Como visto nas seções 2.1, 4.4 e 4.7, a aplicação de Aprendizagem Automática no Processamento da Linguagem Natural e principalmente no Processamento baseado em *corpus*, já vem sendo pesquisada com sucesso. A utilização do paradigma dos Sistemas Multiagentes sobre o Processamento da Linguagem Natural, também é uma realidade exitosa.

A nossa proposta é no sentido de descobrir como aproveitar as potencialidades do modelo de Multiagentes na Aprendizagem Automática do processamento baseado em corpus.

O objetivo maior do PLN é a concepção e a construção de sistemas computacionais que analisem, entendam e gerem aplicações em linguagem natural inclusive sistemas de tradução que transcrevam textos de uma língua para outra. Muitos esforços tem sido feito para atingir esse objetivo, novos modelos, equipes interdisciplinares, projetos grandiosos mas metas de alta complexidade ainda devem ser atingidas e certamente através da aplicabilidade da **Aprendizagem Automática** em **Sistemas Multiagentes** no **Processamento da Linguagem Natural** grandes passos devem ser atingidos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [ALLE94] ALLEN, J. **Natural language understanding**. The Benjamin/Cumming Company. 1994.
- [BEAR91] BEARDON, C; LUMSDEN, D.; HOLMES, G. **Natural Language and Computational Linguistics**. Ellis Horwood Ltd. Melsham-Wiltshire, England, 1991.
- [BIBER93] BIBER, D. **Using Register-Diversified Corpora for General Language Studies**. *Computational Linguistic* Vol. 19 N.2. 1993.
- [BOND88] BOND, A H.; GASSER, L (eds) **Readings in Distributed Artificial Intelligence**. San Mareo, Morgan Kaufmann, 1988.
- [CARB89] CARBONELL, J. (eds.) **Machine Learning - Paradigms and Methods**. MIT/Elsevier. 1989.
- [CHUR93] CHURCH, K.; MERCER, R. **Introduction to the Special Issue on Computational Linguistic Using Large Corpora**. *Computational Linguistic* Vol. 19 N.1. 1993.
- [DE LIMA96] DE LIMA, V.L.S. - Processamento da Linguagem Natural - premissas e desafios -IV Escola Regional de Informática - SBC. 1996
- [DE LIMA97] DE LIMA, V.L.S. - Projeto NALAMAS - Atividades setembro 1996/ agosto 1997 - PUCRS - Relatório Técnico -1997
- [RUSS95] RUSSELL, S.; NORVING, P. **Artificial Intelligence - A Modern Approach**. Prentice Hall. 1995.
- [WEIß95] WEIß, G. ; SEN, S. **Adaptation and Learning in Multi-Agent**. Springer, 1995.
- [WERN92] WERNER, E. "The Design of Multi-Agent Systems" in **Decentralized AI3**. WERNER, E. & Demazeau, Y., (eds). Elsevier Science Publishers B.V.,1992.