

Análise Semântica de Sentimentos Utilizando Árvores de Decisão Adaptativa

A. M. Silva & R. L. A. Rocha & J. José Neto

Resumo— Devido ao crescimento das redes sociais na Web, análise de sentimentos e mineração de opiniões são temas de pesquisas que cada vez mais tem se tornado frequente. O objetivo deste trabalho de pesquisa é aprofundar o estudo de análise de sentimentos, e propor um modelo prático da Roda das Emoções de Plutchik juntamente com a utilização de Árvores de Decisão Adaptativa para classificação automática de sentimentos, de menções extraídas de textos da rede social Twitter. Neste sentido a classificação de sentimentos pode ser de grande ajuda, realizando a triagem de conteúdos de mensagens relevantes, descartando a informação desnecessária, vinculando mensagens com mesmo tema em comum, e até gerando métricas para classificação das emoções. Esta pesquisa é um exercício exploratório, com base no conjunto de Tweets de uma determinada ontologia, busca-se caracterizar indicadores linguísticos capazes de automaticamente, classificar e hierarquizar as emoções contidas nos sentimentos pelo método teórico Roda de Emoções de Plutchik utilizando os modelos explicitamente teóricos propostos: MFA – Modelo de Filtragem Adaptativa e o MCS – Modelo de Classificação de Sentimentos ambos dotados de ações adaptativas.

Keywords— Análise de Sentimentos, Árvores de Decisão Adaptativas; Análise Semântica, Linguística de Corpus, Redes Sociais, Adaptatividade.

I. INTRODUÇÃO

O crescimento de informações colocadas na Web por empresas de notícias, e usuários que compartilham suas ideias, opiniões pessoais e críticas, é evidente nas últimas décadas. A divulgação dessas informações se dá em blogs pessoais, microblogs, sites de relacionamento e principalmente em redes sociais como *Facebook* e o *Twitter*. Com esse crescimento de informações inseridas na Web pelos usuários, de forma que a informação divulgada é de opinião pessoal, isso despertou uma nova área de pesquisa chamada de análise de sentimentos ou mineração de opinião, que vem acompanhada da área de PLN - Processamento de Linguagem Natural e LC - Linguística de Corpus. Para extração, processamento e armazenamento das informações originadas nas redes sociais, utiliza-se de técnicas já empregadas dentro das áreas de PLN e LC.

O desafio consiste em descobrir padrões que possam resultar em uma consciência coletiva dos usuários de internet. E não simplesmente subjugar um ou outro usuário por determinada opinião divulgada. Mas sim realizar análise semântica dos sentimentos e classificá-los em um nível de assertividade.

Motivação

A pesquisa interdisciplinar da comunicação e análise de sentimentos em uma determinada ontologia é motivada não só

pela importância e efeito de causalidade da informação, mas também pela percepção de que o resultado das mensagens divulgadas na rede social gera um problema complexo de se lidar. O modo como a informação é distribuída afeta outros indivíduos, que por sua vez, são seguidores do indivíduo que iniciou a primeira divulgação na rede social. Os indivíduos influenciados podem ou não compartilhar da mesma opinião da mensagem inicial. Avaliar e extrair medidores deste tweets são desafios deste trabalho de pesquisa. Porém, a dimensão de informações disponíveis nas redes sociais, dificulta a análise e classificação das informações nelas contidas, que graças à pessoalidade, não existe uma uniformidade, um padrão entre as mensagens de diversos indivíduos, aos mais variados assuntos postados. Devido a isso, tem-se a necessidade de encontrar métodos que possam utilizar essas informações e prover resultados eficazes.

Objetivo

A partir de sentimentos exibidos nos tweets, o objetivo principal é concluir se a população achou um fato positivo ou negativo, e também a classificação das emoções, tomando como base a Roda das Emoções de Plutchik, contidas nos sentimentos pré-classificados.

Objetivos Específicos

Com base no objetivo central do trabalho, os objetivos específicos são:

- selecionar tweets referentes à ontologia em questão;
- estabelecer um critério computacional, de modo que as mensagens sejam de relevância dentro da ontologia, descartando informações desnecessárias;
- caracterizar indicadores linguísticos (léxico-semânticos) capazes de reconhecer se o tweet possui sentimentos/emoção dentro do domínio ontológico;
- descrever recursos computacionais adaptativos capazes de classificar os sentimentos e suas emoções de acordo com a Roda das Emoções de Plutchik;

Organização do Trabalho

A pesquisa será dividida em quatro grandes etapas, que serão desenvolvidas hierarquicamente. A primeira refere-se ao estudo da Análise de Sentimentos e das Ontologias. A segunda refere-se às tecnologias adaptativas que já foram empregadas em outros trabalhos de pesquisas e explorará a tecnologia adaptativa de árvores de decisão, e como ela pode ajudar no problema de classificação de sentimentos. A terceira refere-se à proposta de dois modelos encapsulados de ações adaptativas, o primeiro MFA – Modelo de Filtragem Adaptativa dotado de um automoto adaptativo simples e o segundo MCS – Modelo de Classificação de Sentimentos dotado de árvores de decisão

adaptativas. E a quarta etapa que irá explanar a conclusão deste trabalho de pesquisa.

Hipóteses

- Qual o tipo de emoção que o indivíduo mencionou ao deixar o seu tweet escrito?
- Como reconhecer automaticamente sentimentos de textos extraídos dos tweets?
- Como extrair automaticamente sentimentos de textos extraídos dos tweets?
- Como associar automaticamente se um tweet é positivo ou negativo?
- Como associar automaticamente os sentimentos contidos em um tweet na Roda das Emoções?

II. ANÁLISE DE SENTIMENTO

O foco da análise de sentimentos é identificar o sentimento que o usuário menciona a respeito de uma marca, uma entidade, um assunto de interesse, um lugar, um produto. Essa opinião explícita em formas de mensagens textuais divulgadas na internet pode ter sido formada a partir de conteúdos noticiosos também encontrados na Web, por *cookies* relacionados ao perfil do usuário, ou não. O usuário pode formar sua opinião a partir de conteúdos que buscou de forma não digital. Mas a maior parte das vezes, o usuário teve acesso à informação por meio digital, mas a continuidade e divulgação de forma liberal de sua opinião serão concebidas de forma digital, e estudos mostram que cada vez mais os indivíduos utilizam as redes sociais.

Grupos, empresas e pesquisadores querem saber a opinião coletiva de determinado assunto. Por exemplo, o lançamento de um novo produto por uma empresa, posteriormente essa empresa irá acompanhar as opiniões dos usuários e relacionar se aquele novo produto teve aceitação ou não, e qual foi o nível dessa aceitação. Será preciso monitorar essas opiniões divulgadas, realizar filtros específicos para encontrar exatamente as menções relacionadas aquele produto em específico. A análise de sentimentos é também uma forma de mineração de dados que irá prover essa demanda.

Para realização da análise de sentimentos, é preciso também realizar o uso de ontologias, valores numéricos estatísticos que ajudarão a quantificar o quanto uma palavra indica ou não determinado sentimento. Em análise de sentimentos, não serão analisados o conteúdo expresso no texto como um todo, mas sim a classificação da emoção predominante no assunto de pesquisa emergente, que poderá ser positivo, negativo.

Análise de sentimentos utilizando algoritmos de aprendizagem de máquina e redes bayesianas utilizando textos extraídos do Twitter em língua espanhola foi proposto por Grigori et al em 2012 [1]; Em 2014, KANAVOS et al. [2] propôs um modelo de análise de sentimentos que mede a influência do usuário na rede com os demais nós dessa rede; Em 2013, PORSHNEV et al. [3] propõem um modelo de previsão do estado psicológico dos usuários a partir dos textos do Twitter, e classifica as previsões utilizando o algoritmo DJIA e classifica-as em oito emoções.

Dificuldades Encontradas no Estudo de Análise de Sentimentos

Dentro da área de PLN, já é sabido que existem diversos problemas lexicais e sintáticos. Esses problemas que se referem à forma da língua natural não serão tratados neste momento, serão considerados como exceções. De forma que possamos focar no principal desafio, que é a análise semântica de sentimentos utilizando árvores adaptativas. Abaixo segue uma lista de dificuldades encontradas:

- textos escritos sintaticamente incorretos, dificultando as filtragens iniciais sólidas;
- distinção de textos opinativos, de textos factícios e de textos informativos;
- identificar dentro de um texto informativo, se existe opiniões embutidas;
- escrita com sarcasmos ou ironias;
- textos com um ou mais assuntos de interesse, e as várias opiniões descritivas no mesmo texto;
- escrita informal e abreviativa;
- conteúdo mentiroso, (ex: um “blogueiro” é pago para divulgar informações positivas de uma empresa e “contaminar” a rede social).

Aplicações Práticas

Abaixo cita-se alguns exemplos de aplicações práticas em que um analisador de sentimentos pode ser empregado:

Análise sobre pessoas: o analisador de sentimentos pode ser utilizado para verificar a opinião dos usuários sobre determinada pessoa, principalmente de pessoas que estão em destaque na mídia publicitária. Por exemplo, a aceitação ou rejeição de determinado político durante a época de eleições, e ofertando a oportunidade do próprio político utilizar as informações para alterar suas estratégias de campanha.

Análise de um produto: Essa análise é muito utilizada por empresas, para verificar as taxas de aceitação e rejeição no lançamento de um determinado produto, ou para acompanhar a curva de vendas de um produto carro chefe da linha.

Bolsa de Valores: Informações de determinada empresa que possui ações abertas na bolsa de valores são avaliadas para auxiliar na tomada de decisão dos próximos passos para investimentos, principalmente das empresas que estão em destaque na mídia.

Mecanismos de Busca

Abaixo cita-se alguns mecanismos de busca utilizados em extração de dados para análise de sentimentos:

Sites de Busca: é possível utilizar o Google ou o Bing, lembrando que essa busca se estenderia a qualquer site, o que torna difícil a padronização e extração dos textos, principalmente na indexação do corpus.

Facebook: Segundo SOCIAL TIMES (2014) [5], em dezembro de 2014 possuía 1,35 bilhões de usuários. Nesta rede social, os usuários podem compartilhar links e “likes” e demonstrar suas opiniões através de mensagens. Embora essa rede possa não ser tão trivial como o Twitter, pois existe uma barreira na obtenção dos dados.

Twitter: considerado um ótimo local, para obter informações sobre determinado assunto. Segundo o SOCIAL

TIMES (2014) [5], em dezembro de 2014 possuía 284 milhões de usuários. Os tweets em sua maioria são públicos, fornecendo uma base rica em informação e de fácil acesso. Uma extensão do Twitter, um site chamado Sentiment140 (<http://www.sentiment140.com>) [6], obtém informações de tweets recentes sobre determinado assunto e monta um gráfico analisando as opiniões de usuários. Na figura 1, é realizada a busca da palavra “dengue” na língua espanhola, o site classifica em 75% de menções negativas e 25% de menções positivas. Na figura 2, é realizada a busca da palavra “terrorism” na língua inglesa, o site classifica em 65% de menções negativas e 35% de menções positivas. Na figura 3, é realizada a busca da palavra “coca-cola” com letra minúscula na língua inglesa, o site classifica em 4% de menções negativas e 94% de menções positivas. Na figura 4, é realizada a busca da palavra “Coca-Cola” com letra maiúscula na língua inglesa, o site classifica em 16% de menções negativas e 84% de menções positivas.

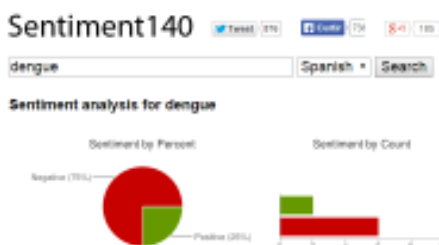


Figura 1: Análise de menções em língua espanhola com a palavra “dengue” (Sentiment140, 2015)



Figura 2: Análise de menções em língua inglesa com a palavra “terrorism” (Sentiment140, 2015)

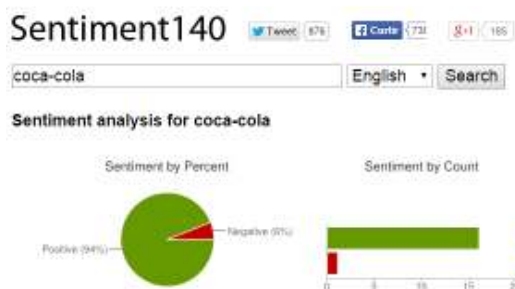


Figura 3: Análise de menções em língua inglesa com a palavra “coca-cola” em letra minúscula (Sentiment140, 2015)

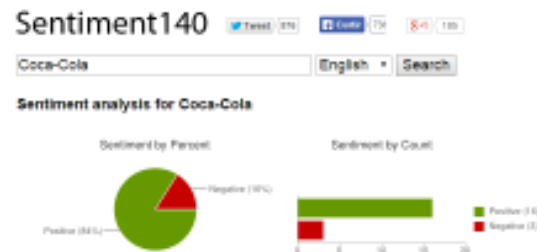


Figura 4: Análise de menções em língua inglesa com a palavra “Coca-Cola” em letra maiúscula (Sentiment140, 2015)

KWAK et al. [7], avalia em seu trabalho o potencial do Twitter como uma rede de informação, em seu estudo mostra que 85% dos tweets criados são relacionados a manchetes de jornais e notícias divulgadas na mídia. Também se pode considerar o fato do Twitter permitir em tempo real, o compartilhamento das mensagens, podendo então captar o sentimento dos indivíduos no momento em que ele passa a conhecer o assunto, a notícia, o anúncio em questão, o que faz com que o indivíduo demonstre a sua “emoção primária” antes mesmo que outros fatores possam vir a influenciá-lo, causando até mesmo uma diminuição do sentimento gerado [8].

Roda das Emoções de Plutchik

Uma emoção não é simplesmente um estado de sentimento. Emoção é uma cadeia de eventos frouxamente ligados que começam com um estímulo e incluem sentimentos, alterações psicológicas, impulso para a ação e comportamento específico, dirigido por objetivos [4]. O efeito do estado emocional é criar uma interação entre o indivíduo e o evento ou estímulo que precipitou a emoção. Analisar o que os indivíduos escrevem nas redes sociais é então uma forma de descobrir que existe sim um evento, algo que tenha estimulado a exteriorização do seu pensamento, de suas emoções, de forma textual utilizando uma rede social. As emoções não são simplesmente eventos lineares, mas são processos de *feedback* que se dá internamente dentro do indivíduo. A função da emoção é restaurar o indivíduo a um estado de equilíbrio quando eventos inesperados criam as disfunções.

PLUTCHIK [4] menciona que existem centenas de palavras para classificar as emoções, e essas palavras tendem a criar grupos/famílias com mesma similaridade semântica. PLUTCHIK [4] descobriu que as emoções primárias podem ser conceituadas de forma análoga a uma roda de cores, colocando as emoções semelhantes juntas ou opostas, como se fosse cores complementares. Ele afirma que existem emoções que são misturas de duas emoções primárias, assim como algumas cores, que são misturas de cores primárias. Ele propõe a análise das emoções com mais uma terceira dimensão que representa a intensidade daquela emoção, o modelo teria então uma forma de um cone.

O *modelo simplístico* consiste em fazer a análise lexical das frases e classificá-las em dois tipos: Positivos e negativos. Essa análise é bastante simples de ser realizada, pois há apenas duas classificações para cada palavra, variando apenas em intensidade. Porém, esse modelo apresenta também resultados muito simplísticos. Apesar de útil para uma análise inicial,

perde a sua utilidade quando são buscadas informações mais precisas quanto às frases. Além disso, esse modelo já usado no âmbito das próprias redes sociais e também como visto nos exemplos do site Sentiment140 [6].

O *modelo de Plutchik* aplica a roda das emoções de Plutchik, mostrada abaixo. Esse modelo apresenta oito sentimentos básicos em formato circular e sentimentos e emoções mais complexas à medida que se afasta do centro da roda. Esse modelo, apesar de apresentar uma complexidade maior para ser trabalhado, se corretamente realizado pode informar com precisão o tom geral da mensagem. Para isso, pode ser usada uma classificação vetorial de cada palavra e esses vetores seriam somados, retornando no sentido da frase.

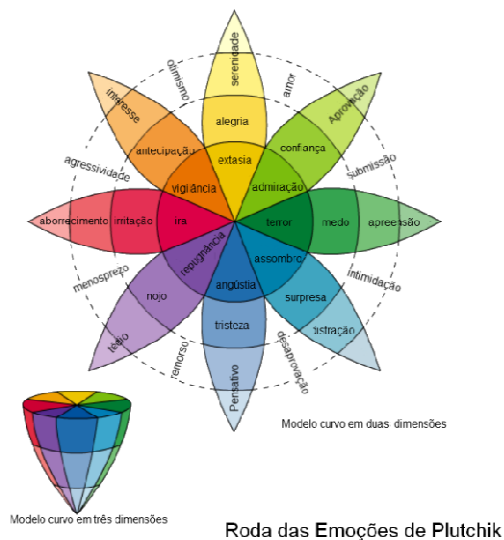


Figura 5: Roda das Emoções de Plutchik (PLUTCHIK, 2001)

III. TECNOLOGIAS ADAPTATIVAS

“O termo “tecnologia” denota o emprego de métodos e conhecimentos científicos com finalidades práticas. Assim, pode-se dizer que se entende por “Tecnologia Adaptativa” o conjunto das aplicações práticas do conceito fundamental da Adaptatividade, sempre que esta for utilizada como instrumento na resolução de problemas oriundos das mais variadas áreas de interesse” [9].

Destacam-se alguns Dispositivos Adaptativos: Dispositivos de reconhecimento, da classe dos autômatos, baseado na sucessão de mudanças de estado; Dispositivos para a representação de sistemas assíncronos, tais como statecharts, que incorporam mecanismos responsáveis pela representação dos fenômenos de sincronização; Dispositivos estocásticos, como as redes de Markov, capazes de representar fenômenos de caráter aleatório; Dispositivos de auxílio à tomada de decisões, representados principalmente pelas tabelas de decisão e pelas árvores de decisão.

Os dispositivos adaptativos são formados de duas camadas. Contêm um dispositivo convencional não adaptativo e uma camada externa envolvente, que leva consigo o mecanismo de

adaptatividade. A camada externa irá atuar na camada interna modificando a estrutura, o comportamento e principalmente o conjunto de regras. A maior vantagem deste dispositivo é a facilidade em poder utilizar dispositivos não adaptativos com funções adaptativas, ocorrendo então à inclusão de camada adaptativa.

“Non-Adaptive” ou dispositivo orientado a regras é qualquer máquina formal, cujo comportamento depende de um conjunto de regras já conhecidos, essas regras devem ser finitas e sequenciais de modo que se possam mapear as próximas ações. Esses dispositivos podem ser determinísticos ou não determinísticos. Um dispositivo é chamado de adaptativo quando ocorre uma ação adaptativa que altere seu conjunto de regras, e somente se a ação adaptativa for não nula. Os dispositivos adaptativos devem ser capazes de reconhecer todos os mecanismos ao longo de toda sua extensão, e saber detectar quais são as possíveis situações em que pode ocorrer efetivamente a mudança e de reagir adequadamente. No modelo proposto por JOSÉ NETO [10], ele enfatiza que o uso de regras em sistemas “Non-Adaptive” já evoluiu muito no que se diz respeito, a notação, sintaxe e semântica. É preciso utilizar-se dos sistemas não adaptativos já projetados e bem documentados, se familiarizar com os dispositivos adaptativos e começar a produzir novas notações se preocupando apenas com a parte que realmente exigirá do projetista a adaptabilidade.

Diversidade do conceito de adaptatividade

Para demonstrar que o conceito de adaptatividade pode ser empregado em diversos dispositivos computacionais, utilizaremos como exemplo a dissertação de mestrado de BASSETO [11], onde demonstra a inserção de sistemas computacionais na composição musical automatizada, modelando o conhecimento musical através de gramáticas sensíveis ao contexto. BASSETO [11] propõem em sua dissertação um modelo de formalismo adaptativo, mediante uma variação dos modelos não determinísticos em Rede de Markov.

A música é um fenômeno contínuo que se constitui de uma sequência temporal de eventos discretizada no tempo. Os seus elementos que a compõem, o conjunto de símbolos nos remete a noção de linguagem, onde pode ser empregado o conceito de tradução musical, para uma linguagem assim chamada de partitura. Dado o conceito de que música pode ser representada por uma linguagem, pode-se então manipular essa linguagem utilizando sistemas computacionais.

O conceito de composição musical se difere completamente do conceito de caos, e sim a uma semântica musical. Comumente um computador não pode ter a sensibilidade de criação, de compor músicas. Porém guiados por programas pré-definidos, poderiam auxiliar na criação de novas combinações, novas composições musicais. Se os computadores forem sujeitos a criação a partir de escolhas aleatórias, seria praticamente submetê-los a geração de ruídos, não haveria um som harmonioso. Porém se programados a partir de regras, de uma linguagem com estruturas sintáticas. “A possibilidade de codificação desses símbolos gráficos em

elementos computacionais pode ser inferida pela existência de um conjunto finito de símbolos e de sua ordenação lógica. A existência de denominações para cada um dos símbolos da partitura já é um ponto de partida para este tipo de codificação” [11].

A música é tratada como uma linguagem musical, e a forma da escrita são transpostas para uma linguagem computacional, ou seja, conversão de regras, onde ele chama de Representação em Listas de Eventos; mais adiante BASSETO [11] trata a questão da interpretação da partitura musical como uma Representação Procedimental, onde um algoritmo, passo a passo, irá interpretar os símbolos, assim como um músico quando lê a partitura. Essas duas representações são semelhantes à forma de representação musical tradicional, ou seja, um processo computacional simplório de execução imediata do material musical. A evolução do trabalho de BASSETO [11] propõem mais duas representações que incluem elementos semânticos sofisticados, onde a proposta seja uma aplicação amigável, mais inteligente e compostas por mais recursos computacionais, onde seja possível através a concepção do usuário realizarmos a representação de modelos musicais semânticos. A Representação Estruturada que engloba o conceito de contexto no aninhamento de estruturas e também o conceito de abstração dos dados associados a elementos musicais; A Representação Semântica, que é a mais interessante para esse projeto de pesquisa, capaz de adaptar modelos mentais e efetuar processamentos muito mais específicos sobre o material musical. Nesta representação os eventos musicais não são tratados de maneira isolada. “A implementação da semântica de um dado objeto musical consiste normalmente na definição de propriedades ou atributos, específicos para este dado objeto musical, bem como através do estabelecimento de relações determinadas entre os objetos musicais que foram definidos” [11].

Segundo PISTORI [12], as árvores de decisão são uma forma de representar hierarquicamente funções discretas sobre múltiplas variáveis. Combinam as características discretas de um dispositivo adaptativo, cujo mecanismo subjacente é uma árvore de decisão, com estratégias para tratamento de valores contínuos, incorretos e inconsistentes.

“Uma árvore de decisão não-determinística é uma árvore de decisão que incorpora o conceito de não-determinismo, bastante estudado na teoria dos autômatos. A introdução do não-determinismo possibilita um tratamento elegante de problemas relacionados com informação incompleta ou ausente, além de propiciar uma distinção conceitual entre decisões exatas e decisões aproximadas” [12].

IV. CLASSIFICANDO SENTIMENTOS COM O APOIO DE ÁRVORES ADAPTATIVAS

Para o emprego da classificação de sentimentos de textos extraídos de uma rede social, neste caso o Twitter, é preciso que exista uma base de conhecimento que primeiramente reconheça se o texto, no caso do Twitter, se o tweet (menção) pertence à ontologia a ser estudada. Sabe-se que o Twitter é uma rede social pública e disponível para qualquer pessoa que

queira minerar dados de sua base. O único padrão que o Twitter obrigatoriamente limita aos seus usuários, é a publicação de tweets com no máximo 140 caracteres. Além dessa limitação, não existe nenhuma outra exigência, de escrita, na forma da escrita, do uso de caractere especial, espaçamentos, fotos e links. Ou seja, os textos podem ser lexicalmente e sintaticamente corretos, como podem ser textos incorretos. A partir dessa observação tornam-se como premissa do projeto proposto, que não serão resolvidos problemas léxicos e sintáticos da língua a ser estudada, mas que o principal problema é o reconhecimento de uma palavra que semanticamente representa um sentimento, uma emoção. Mais adiante será mostrado como o coletor de “sentimentos” tratará cada palavra que representa um sentimento como um único símbolo dentro do autômato adaptativo.

A primeira decisão prática a ser tomada é: que língua será utilizada e parametrizada para ser utilizada como filtro inicial da base de dados que se deseja extrair? Para o modelo proposto optou-se pela Língua Portuguesa, por se tratar de uma língua que muito tem a contribuir em PLN – Processamento de Linguagem Natural. A segunda decisão a ser tomada refere-se a qual banco de dados deve-se utilizar? Facebook? Um jornal regional? Transcrever áudios de sessões entre um paciente e um terapeuta? Para o modelo proposto optou-se pela rede social Twitter, utilizada mundialmente, que desperta grande interesse coletivo sobre assuntos temporais. A terceira decisão é: quais mensagens enquadram-se na Ontologia desejada? Para o modelo proposto a Ontologia pode variar a qualquer momento, pois são sentimentos ligados a Ontologia escolhida que será a saída, indicadores e classificadores que de certa forma medirão o cunho de emoções de um assunto dentro da Ontologia analisada. Pode-se optar também por escolher mais de uma ou várias Ontologias para treinamento inicial e filtragem dos sentimentos que estão na figura 9, representados pelo retângulo “Filtragem *Top Down*”. Nesta etapa que serão reconhecidos os sentimentos, ou seja, quais palavras contidas nos textos extraídos do Twitter devem ser mineradas? Na ausência de um corpus notado em língua portuguesa, que diferencie uma palavra e que a reconheça e a categorize como um sentimento (não importando neste momento qual seja ele) é preciso ensinar a máquina a aprender esse reconhecimento. Nesta fase optou-se por um modelo de automoto adaptativo que reconheça o léxico e o classifique como um sentimento.

V. RESULTADOS

Modelo Proposto 1 – MFA – Modelo de Filtragem Adaptativa

Na figura 6 é proposto um modelo de filtragem adaptativo que irá auxiliar no reprocessamento e enriquecimento da base de dados dos sentimentos que serão utilizados como entrada no MCS – Modelo de Classificação de Sentimentos da figura 9.

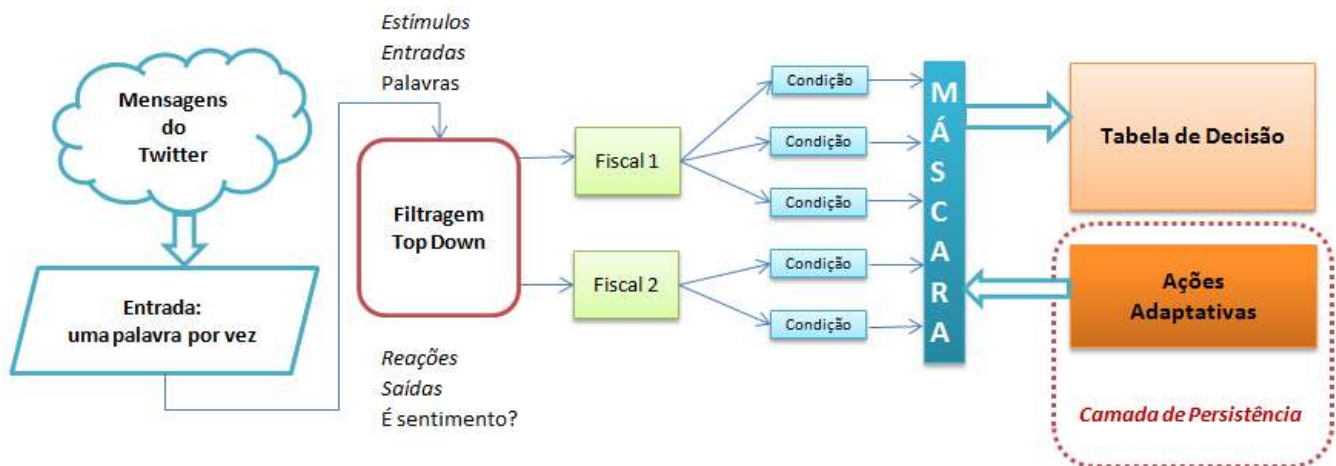


Figura 6: MFA - Modelo de Filtragem Adaptativa

Segundo CEREDA [13], Persistência é o nome dado para o intervalo de disponibilidade do conteúdo de um determinado objeto. Em soluções computacionais essa característica é fundamental em diversos problemas. Ao longo do tempo as informações coletadas necessitam de uma área de preservação do estado, de atualização ou até mesmo uma eventual remoção. “A persistência é uma funcionalidade que pode contribuir para a extensão dos dispositivos adaptativos, conferindo-lhes a capacidade de preservar estados internos e recupera-los conforme a necessidade [13]”.

O MFA possui os seguintes componentes:

- Nuvem extratora de informações do Twitter;
- Módulo de entrada: onde cada palavra é processada individualmente umas das outras;
- Módulo de Filtragem Top Down: onde será enriquecido a cada novo processamento para que se tenha um conhecimento pré-existente (isso é imprescindível para a alimentação do MCS), este módulo receberá os estímulos/entradas que são as palavras uma a uma de cada texto e enviará as saídas que são chamadas de reações, ao próximo módulo.
- Módulos Fiscais: que tem a função de monitorar/vigiar o módulo anterior e diagnosticá-lo;
- Condições: onde as saídas serão analisadas e validadas dando subsidio na tomada de decisão;
- Máscara: que tem a função de interfacear às condições no devido preenchimento da tabela de decisão. A máscara também terá o papel de incorporar as mensagens de retorno das ações adaptativas e incrementar a tabela de decisão.
- Tabela de Decisão: avalia os diagnósticos do bloco anterior, e atua sobre todos os modelos do sistema para adequá-los.
- Ações Adaptativas: que possui a capacidade de aprender e incorporar novos conhecimentos, associando a cada conhecimento adquirido, a informação da procedência da nova informação, e retornando essa informação para a Máscara. A manipulação das informações dentro do módulo de Ações Adaptativas se dá devida a capacidade dos dispositivos adaptativos poderem interpretar a informação e manipulá-la. Este módulo é responsável por todo tipo de decisão tomada, e

também se responsabiliza pela mudança no entorno do modelo a cada vez que uma nova informação é acrescentada.

No MFA o módulo de Ações Adaptativas é composto de uma camada de persistência para facilitar na manutenção e atualização da base de conhecimento que deve ser produzida. A persistência ajuda na configuração de ações léxico-sintáticas em dispositivos adaptativos. A estrutura proposta em camadas, onde a persistência esta implementada em uma camada léxico-sintática, permite, que se caso for preciso realizar alguma alteração nas regras não seria preciso excluir, e treinar a minha base de informações já armazenadas, bastaria apenas excluir e consertar o item que apresenta a inconsistência. Isso mostra que não preciso modificar a estrutura completa de um dispositivo adaptativo e sim envolvê-lo com essa camada de persistência (regras e conhecimento).

Abaixo temos dois exemplos, o primeiro constando a seguinte menção “eu to bem triste com esse negócio da redução da maioria penal ter sido aprovada q odio” e o segundo com a seguinte menção “Feliz por terem aprovado a maioria penal 😊”. Ambos os exemplos foram extraídos do Twitter no dia 02 de Jul. de 2015, e ambos referem-se ao tema de importância nacional “redução da maioria penal”.

O MFA processa cada palavra por vez, e interpreta os símbolos e a composição deles. No primeiro processamento ele aprende, incrementa, e atualiza a base de conhecimento. Nos exemplos acima a base seria incrementada com os seguintes sentimentos: triste / ódio / feliz. Nos processamentos seguintes a comunicação entre a máscara e as condições também serão atualizados com as novas regras, e caso entre uma palavra como, por exemplo, “tristeza” os dispositivos adaptativos podem analisar a terminologia da palavra e já classificá-la também como um sentimento. É importante destacar que no MFA não esta sendo analisada a proximidade semântica dos sentimentos. O principal objetivo do MFA é dizer se a palavra é ou não é um sentimento e enriquecer a base a cada vez que chega um novo sentimento na escrita textual em língua portuguesa.

O MFA é de extrema importância, pois nesse módulo já

será realizado todo o descarte de informações não relevantes, permanecendo apenas as palavras categorizadas como



Figura 7 - Exemplo de mensagem “feliz” (extraído do TweetDeck em 02 de Jul. de 2015)

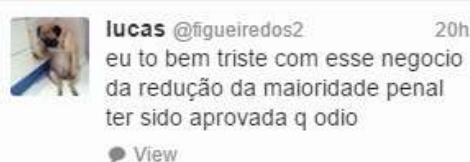


Figura 8 - Exemplo de mensagem “ódio/triste”(extraído do TweetDeck em 02 de Jul. de 2015)

sentimentos, formando então uma base enriquecida, que por sua vez servirá de entrada para o MCS na sessão seguinte. Recomenda-se que primeiro seja treinada uma quantidade razoável de dados oriundos de diversas Ontologias separadamente do funcionamento do MCS. E somente depois de treinado o MFA deve acopla-lo ao MCS. Isso ajuda na performance e evita o processamento desnecessário de informações. Pois o tempo que se gasta esperando o MFA processar cada palavra a mencionar se é ou não um sentimento, em fase de aprendizagem, e mandar essa informação para o MCS, é desnecessário. Então primeiramente deve ocorrer o treinamento do MFA e depois acopla-lo no MCS. E quando entrar um sentimento novo é permissível o tempo de espera do MCS pelo MFA trabalhando em conjunto. Lembrando que as ações adaptativas do MFA realizam modificações somente em seu próprio modelo. Em seguida será explorado o objetivo do MCS e como ocorrem suas ações adaptativas.

Modelo Proposto 2 – MCS – Modelo de Classificação de Sentimentos

Abaixo é proposto o Modelo de Classificação de

Sentimentos utilizando o conceito de árvores adaptativas, para realizar a classificação por proximidade semântica do sentimento mais próximo ao sentimento primário encontrado na Roda das Emoções de Plutchik [4].

Na sessão anterior foi proposto o MFA – Modelo de Filtragem Adaptativa que está ilustrado no MCS – Modelo de Classificação de Sentimentos, pelo quarto componente no fluxo acima proposto, o retângulo “Filtragem Top Down” com uma linha pontilhada em vermelho, onde indica que este é um módulo que possui recurso de adaptatividade e que executa sua função independente dos outros componentes do MCS.

As saídas que foram geradas no MFA serão as entradas e estímulos do MCS. Neste passo os sentimentos são apenas palavras, cada um com seu significado, porém o modelo ainda não é capaz de diferenciá-las em sua forma semântica.

Na figura 10 estão ilustradas as oito emoções primárias que compõem a Roda das Emoções de Plutchik. A partir das emoções primárias foram ilustradas as demais emoções utilizando a notação gráfica de árvores de decisão.

Abaixo esta ilustrada um exemplo da árvore de decisão que pode compor a emoção primária “Ira”. A classificação foi feita a partir de sinônimos de palavras e por proximidade semântica. Na Roda das Emoções “Ira” esta representada na cor rosa, e vem acompanhada das emoções: “irritação” e “aborrecimento”. Essas duas emoções estão representadas na árvore de decisão cada uma por um nó, mas isso não necessariamente significa que elas são opostas ou contrárias, pelo contrário, são emoções sintaticamente escritas de forma diferentes, porém com proximidade semântica. As que são aninhadas de forma hierárquica pai e filho, são as emoções semanticamente mais próximas, exemplo as emoções fúria (pai) e furor (filho); aborrecimento (pai) e abatimento (filho); irritação (pai) e raiva (filho). A relação de proximidade semântica neste exercício se dá pelo significado de cada palavra, e não pela geração de uma emoção primária, que por sua vez gera suas emoções secundárias e assim por diante. E emoções como impaciência (filho de ansiedade) e desânimo (filho de abatimento) não possuem grande proximidade semântica, mas são originadas da mesma emoção primária. Devida a essa sinergia entre emoções semanticamente

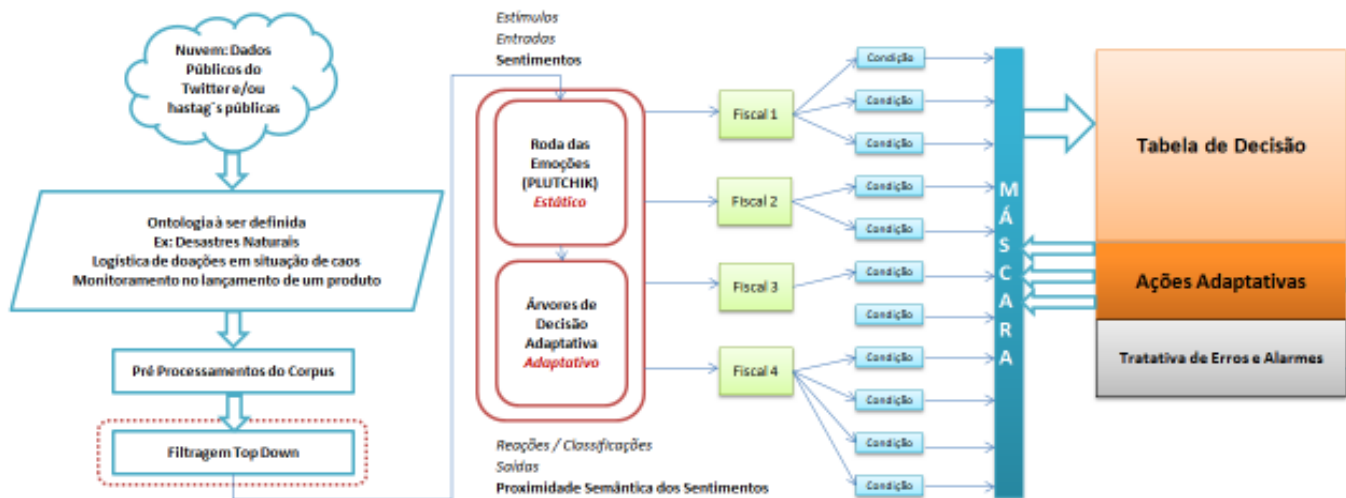


Figura 9 - MCS - Modelo de Classificação de Sentimentos

próximas e não próximas justifica-se a utilização do mecanismo de árvore de decisão.

“ousadia”. Esta classificação resulta em um desdobramento de 5 nós para realizar essa classificação.



Figura 10 - Emoções Primárias

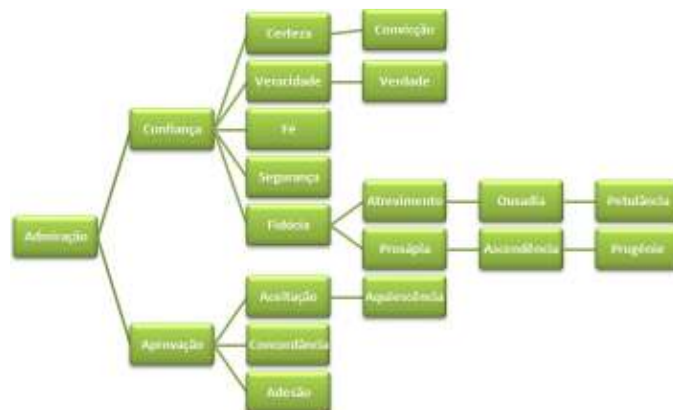


Figura 12 - Árvore de Decisão da Palavra "Admiração"

A figura 12, também representa um exercício de como seria o comportamento da Árvore de Decisão Adaptativa caso receba-se como entrada as emoções nela contida. Caso esse componente receba como entrada uma emoção ainda não mapeada, ou seja uma emoção nova, automaticamente seriam acionados os outros módulos do sistema. Primeiramente os “fiscais” com o objetivo de vigiar e diagnosticar o módulo anterior caso manifeste algum comportamento diferente, neste caso, o módulo anterior ainda não saberia identificar a nova emoção. Em seguida, passaria pelas componentes de “condição”, que analisam, filtram e dão os subsídios necessários para a tomada de decisão. Na tabela de decisão então é avaliado os diagnósticos do bloco anterior, e atuará sobre todos os modelos do sistema para adequá-los. Caso necessário então ações adaptativas operam transformando as regras e modificando o modelo atual, principalmente na criação de uma nova categoria para a nova emoção inputada no modelo de classificação de sentimentos. O componente de ações adaptativas retroalimenta o modelo. É importante destacar também os componentes de erro e alarme que registram o comportamento do modelo, caso não exista nenhum caminho a seguir ou nenhuma decisão a ser tomada.

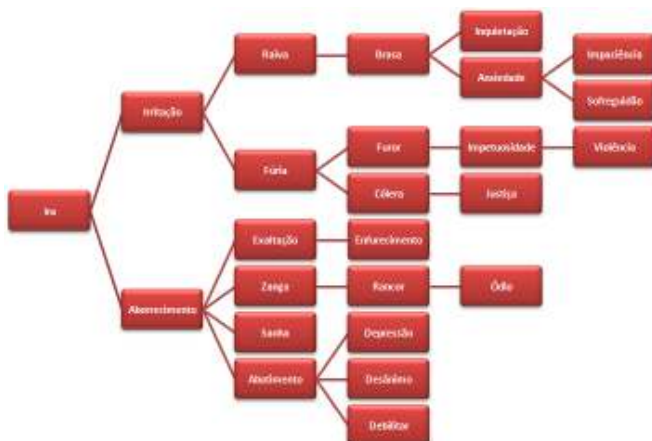


Figura 11 - Árvore de Decisão da Emoção "Ira"

A árvore de decisão apresentada com a emoção primária “ira” é um exercício prático de como seria o comportamento do componente modelo representado pelo retângulo em vermelho na figura 9. Esse componente é formado pela união do quadrado “Roda das Emoções” mais o quadrado “Árvore de Decisão Adaptativa”, sendo o primeiro uma pré decisão da entrada do quadrado seguinte, ambos trabalham em conjunto, o primeiro fazendo a seleção de uma emoção e verificando se a mesma já existe na Roda das Emoções, caso contrário, a emoção é enviada para o segundo onde será realizada uma classificação a partir da árvore de decisão adaptativa. Por exemplo, se a árvore de decisão recebe-se a palavra “Petulância”, como ilustrada na figura 12, a árvore realizaria os testes iniciando da emoção mais primária “admiração”, em seguida faria um teste de sinônimo e varredura histórica, para verificar em qual dos filhos ela melhor se encaixa, passando pelas emoções “confiança”, em seguida “fidúcia”, em seguida “atrevimento”, e finalmente seria aninhada a emoção

Problemas em Aberto

No exercício praticado para compor a árvore de decisão adaptativa dotada de emoções, secundárias, terciárias, quartanárias e assim por diante, foi utilizado um raciocínio a partir dos sinônimos das palavras que já vinham mapeadas como sentimentos no MFA – Modelo de Filtragem Adaptativa. Um dos problemas que fica em aberto neste trabalho para desenvolvimento futuro é a relação automática para aninhar os sentimentos no nó mais próximo semanticamente de sua emoção primária. Poder-se-ia explorar mais a fundo as técnicas de frequência, probabilidade e estática, e principalmente de semântica latente.

Como um ponto de atenção para estudos futuros é a exploração de qual emoção gera outra emoção no âmbito psicológico de formação dessa emoção, propriamente dita, dentro do ser humano. E se pode haver uma relação dessa

formação natural de emoções e passagens por cada uma dessas emoções com a forma de classificação adaptativa utilizando árvore de decisão para classificar as emoções a partir de um modelo raiz, que é a Roda das Emoções.

VI. CONCLUSÃO

Nos modelos teóricos propostos neste trabalho de pesquisa o MFA e o MCS, foi possível esboçar o uso de funções Adaptativas que podem auxiliam no tempo de processamento, na tomada de decisão e principalmente no tamanho do caminho que o sistema irá percorrer para chegar a uma solução coerente, e certamente correta. O uso da camada de persistência pode se mostrar eficaz no MFA, comunicando-se diretamente com a Máscara que por sua vez, retroalimenta o sistema e as gravações nas tabelas de decisões. Tendo os sentimentos já separados pelo MFA, e trabalhando em conjunto com o MCS pode-se perceber a eficiência dos modelos em funcionamento paralelo. O uso de recursos adaptativos no MCS pode se mostrar eficaz e coerente para classificação de sentimentos fundamentada nas oito emoções primárias da Roda das Emoções.

O mapeamento deste trabalho de pesquisa foi fundamental para realizar o detalhamento das atividades de próximos passos. Classificação de sentimentos é uma área multidisciplinar que envolve o conhecimento em ferramentas de processamento de linguagem natural, e o uso de ferramentas adaptativas foi fundamental no processo de aprendizagem do sistema de classificação de sentimentos.

AGRADECIMENTOS

Ao CNPQ, aos colegas de trabalho do LTA – Laboratório e Linguagens e Técnicas Adaptativas, aos organizadores deste Workshop e ao programa de Pós-Graduação em Engenharia da Computação.

REFERÊNCIAS

- [1] GRIGORI, S. et al. **Empirical study of machine learning based approach for opinion mining in tweets**. In *Proceedings of the 11th Mexican international conference on Advances in Artificial Intelligence*, 2012. Springer-Verlag, Berlin, Heidelberg, 1-14
- [2] KANAVOS, A. et al. **Conversation Emotional Modeling in Social Networks**, Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on , vol., no., pp.478,484, 10-12 Nov. 2014
- [3] PORSHNEV, A.; REDKIN, I.; SHEVCHENKO, A., "Machine Learning in Prediction of Stock Market Indicators Based on Historical Data and Data from Twitter Sentiment Analysis," Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on , vol., no., pp.440,444, 7-10 Dec. 2013.
- [4] PLUTCHIK, R; **The Nature of Emotions**. In: American Scientist Magazine, vol. 89, number 4, pp. 344-350, 2001.
- [5] SOCIAL TIMES. **The 10 Biggest Social Networks Worldwide**. Disponível em: <<http://www.adweek.com/socialtimes/largest-social-networks-worldwide/504044>>. Acesso em 19 abr. 2015.
- [6] SENTIMENT 140. Disponível em: <<http://www.sentiment140.com>>. Acesso em 19 abr. 2015.
- [7] KWAK, H., LEE, C., PARK, H., and MOON, S. (2010). "What is Twitter, a social network or a news media?", Proceedings of the 19th International Conference on World Wide Web, (New York, NY, USA: ACM), pp. 591-600.

[8] SLOMAN, A., CHRISLEY, R., SCHEUTZ, M. (2005). "The architectural basis of affective states and processes", Who Needs Emotions?: The Brain Meets the Machine, v. 3, pp. 203-244.

[9] JOSÉ NETO, J. **Um glossário sobre adaptatividade**. In: **Terceiro Workshop de Tecnologia Adaptativa - WTA 2009**. EPUSP, 2009.

[10] JOSÉ NETO, J. **Adaptive Rule-Driven Devices - General Formulation and Case Study**, in Bruce W. Watson & Derick Wood, ed., 'CIAA', Springer, pp. 234-250, 2001.

[11] BASSETO, B. A. **Um Sistema de Composição Musical Automatizada, Baseado em Gramáticas Sensíveis ao Contexto, Implementado com Formalismos Adaptativos**. São Paulo 2000, 143p. Dissertação de Mestrado. Escola Politécnica, Universidade de São Paulo.

[12] PISTORI, H. **Tecnologia em Engenharia da Computação: Estado da Arte e Aplicações**. Tese de Doutorado, Escola Politécnica da Universidade de São Paulo, 2003.

[13] CEREDA, Paulo Roberto Massa, JOSÉ NETO, João. **Persistência em dispositivos adaptativos**. Em: Memórias do VIII Workshop de Tecnologia Adaptativa – WTA 2014. EPUSP, São Paulo, ISBN: 978-85-86686-76-4, pp. 120-125. 06 e 07 de Fevereiro de 2014.



Ariana Moura da Silva, graduada em Sistemas de Informação pela FAENAC – Faculdade Editora Nacional e Mestre em Engenharia da Informação pela UFABC – Universidade Federal do ABC. Atualmente, é doutoranda do Programa de Pós-Graduação em Engenharia de Computação do Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo, atuando como aluna pesquisadora no LTA – Laboratório de Linguagens e Técnicas Adaptativas – do PCS. Tem experiência na área de Processamento de Linguagem Natural, com ênfase em Análise Semântica, atuando principalmente nos seguintes temas: classificação de emoções, mineração de dados de redes sociais, redes de comunicação em desastres naturais. Curriculum Lattes: <http://lattes.cnpq.br/2930741552082859>.



Ricardo Luis de Azevedo da Rocha, natural do Rio de Janeiro-RJ e nasceu em 29/05/1960. Graduou-se em Engenharia Elétrica modalidade Eletrônica na PUC-RJ, em 1982. É Mestre e Doutor em Engenharia de Computação pela Escola Politécnica da USP (1995 e 2000, respectivamente). Suas áreas de atuação incluem Tecnologias Adaptativas, Fundamentos de Computação e Modelos Computacionais. Dr. Rocha é membro da ACM, IEEE e SBC. Curriculum Lattes: <http://lattes.cnpq.br/5660360751410581>.



João José Neto é graduado em Engenharia de Eletricidade (1971), mestre em Engenharia Elétrica (1975), doutor em Engenharia Elétrica (1980) e livre-docente (1993) pela Escola Politécnica da Universidade de São Paulo. Atualmente, é professor associado da Escola Politécnica da Universidade de São Paulo e coordena o LTA – Laboratório de Linguagens e Técnicas Adaptativas do PCS – Departamento de Engenharia de Computação e Sistemas Digitais da EPUSP. Tem experiência na área de Ciência da Computação, com ênfase nos Fundamentos da Engenharia da Computação, atuando principalmente nos seguintes temas: dispositivos adaptativos, tecnologia adaptativa, autômatos adaptativos, e em suas aplicações à Engenharia de Computação, particularmente em sistemas de tomada de decisão adaptativa, análise e processamento de linguagens naturais, construção de compiladores, robótica, ensino assistido por computador, modelagem de sistemas inteligentes, processos de aprendizagem automática e inferências baseadas em tecnologia adaptativa. Curriculum Lattes: <http://lattes.cnpq.br/4091709928353457>.