# Aplicações de Treinamento Semi-Supervisionado em Casos Bipolares de Linguagem Natural

#### Renato Fabbri

Universidade de São Paulo, São Carlos, São Paulo, Brazil renato.fabbri@gmail.com

Julho de 2010

#### Resumo

Neste trabalho a técnica de aprendizado semi-supervisionado baseada grafos chamada Propagação de Rótulos foi aplicada à tarefa de detecção de polaridade (sentimento negativo/positivo) em fala e texto. Para trechos de fala, obtivemos grande exito com um aumento de 20% das classificações corretas em comparação com os resultados obtidos pela utilização de métodos de aprendizado supervisionado usuais. No caso dos textos, o ganho foi moderado, resultando em ganhos gerais de 2-3% em comparação com os métodos tradicionais de aprendizado supervisionado. São explicitados os métodos em utilização pelo nosso grupo de pesquisa e váriadas formas de aprimorar o resultados.

#### 1 Introdução

Transportando multiplas informações concomitantes, a fala é o recurso primário do ser humano para comunicação (Jusczyk, P. W., 2000; Teixeira, A., 2009). Nossa cultura é marcada pela escrita e sua reprodutilbilidade, e esta é uma versão reduzida e estática da fala natural - ao menos para os homens de cultura ocidental (McLuhan, M., 1962). As tecnologias de comunicação revelam esta natureza e primazia, em especial com os celulares e televisores, mas também pelo rádio, o qual não tem contrapartida da expressão visual. Seu estudo é essencial para práticas interpretativas e de oratória (Borges, C. de A.,

2004), sem o qual um ator ou orador dificilmente alcança o resultado que busca. Diversas patologias de natureza física, psíquica ou psicogênica, manifestam traços linguísticos e paralinguísticos discerníveis por especialistas e máquinas (Dalgalarrondo, P., 2008; Yacoub, S. and Simske, S. and Lin, X., and Burns, J., 2003; Tóth, S. L. and Sztahó, D. and Vicsi, K., ; McGilloway, S. and Cowie, R. and Douglas-Cowie, E. and Gielen, S. and Westerdijk, M. and Stroeve, S., 2000).

Características do locutor, de sua natureza e do seu estado, podem ser capturados pela análise do que ele apresenta pela fala. A idade, o gênero, origem geográfica, escolaridade, estado emocional e até condição física podem ser discernidos com base em características verbais e prosódicas (Yacoub, S. and Simske, S. and Lin, X., and Burns, J., 2003).

#### 2 Base Teórica

## 2.1 Processamento de Linguagem Natural

Por *linguagem natural* entende-se linguagens utilizadas por seres humanos para a comunicação cotidiana, linguagens como o português, o inglês ou o yorubá. Em contraste com linguagens artificiais como as linguagens de programação ou a linguagem matemática, as linguagens naturais se desenvolvem de geração em geração, e são difíceis de serem entendidas como uma série de regras bem definidas.

Processamento de Linguagem Natural (PLN) é uma ampla áerea de estudos que trata da análise e produção de textos através do computador. Envolve desde técnicas es-

tatísticas baseadas em ocorrências de palavras ou características destas¹ até o entendimento de colocações linguísticas no sentido de dar respostas adequadas (Bird, S. and Klein, E. and Loper, E., 2009). Nosso foco é posto na *análise* de estruturas textuais, mas vale ressaltar que a *produção* destas estruturas também é tema de estudo.

A utilização de tecnologias baseadas em PLN está bastante difundida. Tradutores automáticos, sumarizadores e corretores ortográficos e sintáticos são exemplos tipicamente vinculados à utilização do computador. Celulares e computadores de mão fazem uso de reconhecimento de escrita. Buscadores de web (como o Google ou o Altavista), fazem uso extensivo de técnicas de PLN para detectarem palavras-chave da página indexada.

Embora o processamento de fala possa ser entendido como uma sub-área de PLN, historicamente elas são tidas como áreas distintas que muitas vezes estão postas em conjunto. Nosso projeto abrange também o processamento de fala (o que envolve acústica), o que faz da pesquisa um estudo referente ao PLN no seu sentido mais amplo. Mesmo assim, preferimos, por razões de clareza e tradição das diferentes áreas, discernir, neste trabalho, as áreas de PLN e processamento de fala.

Tarefas associadas à área de PLN abrangem desde contagem de palavras e hifenização até responstas automáticas e tradução em tempo real (Jurafsky, D. and Martin, J. H., 2000). O que distingue uma aplicação de processamento de linguas de outras aplicações de processamentos de dados é a utilização de conhecimentos linguísticos. Consideremos o programa comum em sistemas unix-like de nome wc. Quando invocado, ele retorna o número de bytes, palavras e linhas em um dado arquivo ou conjunto de arquivos. Quanto ao retornar bytes e linhas, isso é considerado uma tarefa ordinária de processamento de dados. Ao retornar o número de palavras a rotina envolve discernir o que é uma palavra, e aí o wc se torna um sistema de processamento de linguagem.

A seguinte divisão das categorias de conhecimento linguístico é feita tanto em teoria quanto na prática:

- Fonética e Fonologia O estudo dos sons linguísticos.
- Morfologia O estudo dos componentes das palavras (e seus significados).
- Sintáxe O estudo da relação estrutural entre palavras.
- Semântica O estudo dos significados.
- Pragmática O estudo do uso da linguagem para fins específicos.
- Discurso O estudo de unidades linguísticas que envolvem conjuntos de colocações, frases, etc.

#### 2.2 Processamento de Fala

A caracterização acústica de um som é usualmente feita em termos do espectro e do volume. Estaremos sempre lidando com a evolução temporal do espectro do trecho de fala, seja ele um fonema, uma palavra, um segmento associado a uma emoção ou uma gravação inteira associada a uma psicopatologia ou gênero discursivo. São possíveis inúmeras formas de retratar a evolução espectral de um som e podemos apontar questões chave através de um espectrograma.

A Figura 1 é um espectrograma usual. Escolhemos uma espectrograma com traços bastante claros para fins ilustrativos. O eixo vertical se refere ao espectro e o eixo horizontal às amostras. O significado real dos valores da figura depende da taxa de amostragem. Como exemplo, podemos supor que a gravação analisada tenha sido feita em 44.1kHz, portanto, pela lei de Nyquist a frequência máxima representada no sinal digitalizado é de  $\frac{44.1kHz}{2} = 22.05kHz$ . Notamos que a energia está concentrada abaixo de 0.7 no espectrograma, o que corresponde a  $0.7 \cdot 22.05kHz = 15.435kHz$ . Também podemos notar que o eixo horizontal se estende até 120k, portanto,

<sup>&</sup>lt;sup>1</sup>Por exemplo o tamanho e a classe gramatical de cada palavra.

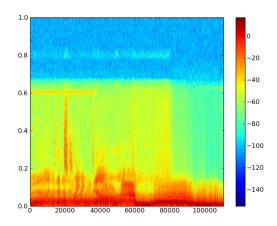


Figura 1: Espectrograma padrão

dada a taxa de amostragem de 44.1kHz, a duração do som é de  $\frac{120}{44.1} \approx 2.72$  segundos. Características mais elaboradas são obtidas pela análise das proporções entre os harmônicos mais frequentes em um dado momento (p.ex. para achar a frequência fundamental).

Desta forma, podemos apreender características locais como a intensidade, a frequência fundamental e a proporção de harmônicos de cada grão sonoro (duração típica entre 20-100ms), cada sílaba, cada palavra e cada formação contínua de fala ou construções maiores. Outras características são globais, referindo-se a cada trecho analisado. Embora existam bibliotecas disponíveis para estas tarefas, trabalhos relacionados a este Projeto muitas vezes utilizam programas como o Praat e o WaveSurfer, que agregam boa parte das rotinas necessárias para análises acústicas desta natureza (Toivanen, J. and Seppänen, T. and Väyrynen, E., 2009; Sidarova, J. and McDonough, J. and Badia, T., 2009). As características extraídas são utilizadas para alimentar algum tipo de classificador de padrões como os utilizados neste trabalho.

Em geral a altura (frequência fundamental F0) é vista como o parâmetro mais importante para a expressão vocal de emoções. A energia (intensidade), duração e taxa de articulação são outros parâmetros relevantes.

Segue uma breve descrição de cada uma destas dimensões da expressão vocal:

- A Frequência Fundamental (F0), medida em Hertz (Hz), é a taxa com que a forma de onda se repete por unidade de tempo. A F0 média é a soma do valor de F0 para cada trecho dividida pelo número de trechos. A mediana F0 é o valor da F0 do grão que fica no meio depois da ordenação dos grãos segundo a F0. A F0 modal é o valor de F0 mais ocorrente no trecho em questão. É preferencialmente utilizado os percentuais 5° e o 95° em vez dos limites inferior e superior, pois tais limites podem ser ocorrências pontuais de desvios como falsetes ou perda ocasional do tônus muscular. Tais valores podem ser dados em relação à escala de semitons musicais e então o termo pitch sigma é utilizado (a escala logarítmica é mais apropriada para pesquisas de acústica aplicadas a psicologia). A declividade da variação do F0 pode ser dada em semitons por segundo. É comum computar a máxima e a mínima variação de F0, assim como a variação média.
- O Jitter é definido como a variação de altura (novamente F0) entre ciclos consecutivos.
- O Shimmer é a magnitude da variação de intensidade entre ciclos consecutivos.
- A Amplitude é a magnitude de deslocamento de uma onda sonora, a Intensidade se refere à quantidade de energia transmitida por unidade de tempo e o Volume é a quantidade psicoacústica percebida da intensidade. O volume efetivo além de ser um mapeamento logarítmico da intensidade (ou da amplitude) varia também com a frequência de acordo com as curvas iso-audíveis de Fletcher-Munson. Na prática, geralmente se registra o volume somente mapeando a intensidade (ou a amplitude) para a escala

logarítmica, sem adequar o volume efetivo às frequências da onda emitida.

- A banda e a frequência média dos formantes (especialmente os primeiros três harmônicos F1, F2 e F3) portam informações valiosas do ponto de vista das emoções.
- A expressão de emoções podem estar ligadas a características espectrais tais como a distribuição de frequências mais agudas. Sabemos, por exemplo, que em uma fala raivosa ou bem tensa encontramos mais harmônicos agudos (acima de 1kHz). O oposto ocorre com fala bem relaxada ou relacionada ao tédio.
- Quanto à Articulação, é comum extrair os seguintes parâmetros: tempo total de fala, tempo total de pausa, tempo médio de fala, tempo médio de pausa, latência de resposta, taxa de articulação (fonemas/segundo), mínimo, máximo e desvio padrão da duração dos trechos de fala, mínimo, máximo e desvio padrão da duração dos trechos de silêncio.

Nosso trabalho está concentrado na exploração da frequência fundamental para detecção de polaridades. Salientamos que os resultados podem ser aprimorados de forma imediata simplesmente através da utilização de outras característica dos sons.

#### 2.3 Aprendizado Semi-supervisionado

O aprendizado semi-supervisionado (ASS) visa o aprendizado a partir da utilização tanto dos objetos rotulados quanto dos não rotulados. São exemplos de ASS a classificação semi-supervisionada (CSS) e a clusterização forçada. Este trabalho é um apanhado sobre CSS em grafos que é como geralmente é entendido o ASS com grafos.

O nome "semi-supervisionado" vem do fato de que os dados utilizados para o aprendizado está entre os dados utilizados no aprendizado supervisionado e não supervisionado<sup>2</sup>.

Podemos comparar a CSS ao processo de uma criança que entende o que é um animal através de alguns exemplos dados pelos adultos e de outros vários que ela mesma observa. Ou seja, para o aprendizado do que é (e o que não é) um animal, ela utiliza tanto o conhecimento sobre os animais que contaram para ela que é um animal quanto o conhecimento sobre outros seres que não falaram para ela o que é.

Um grande atrativo dos métodos de ASS é que ele pode ajudar a contornar o fato de que os objetos rotulados podem ser custosos. Comumente necessitam de um humano especializado para rotulá-los, outras vezes necessitam de equipamentos, condições ou processamentos especiais. Muitas vezes é simplesmente impossível obter novos objetos rotulados. Já os objetos não rotulados são geralmente imediatos e abundantes, podendo ser utilizados pelo ASS de imediato.

Uma distinção importante que se faz é quanto ao ASS *indutivo* e o ASS *transdutivo*. O ASS indutivo se assemelha ao aprendizado transdutivo. A diferença entre ambos é que o aprendizado transdutivo é otimizado para tratar somente os objetos utilizados para o treinamento, enquanto o ASS indutivo trata de objetos não presentes na fase de treinamento.

Dados não rotulados podem ajudar de diversas formas, em especial quando as diferentes classes se mostram clusterizadas. Na Figura 2 é explicitado esquematicamente a forma com que os dados não rotulados podem auxiliar a determinar uma fronteira de decisão mais pertinente. Embora este seja o comportamento natural mais comum, devemos notar que o contrário também ocorre e que neste caso o princípio da clusterização das classes semelhantes pode nos levar a resultados enganosos. A Figura 3 demonstra este caso no qual o critério de clusterização levaria a uma pior classificação.

<sup>&</sup>lt;sup>2</sup>No aprendizado supervisionado são utilizados somente

dados rotulados. No aprendizado *não supervisionado* são utilizados somente dados não rotulados.

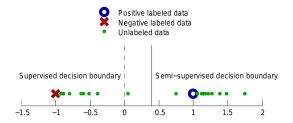


Figura 2: Dados não-rotulados auxiliando a criar uma melhor fronteira de decisão. Fonte: (Zhu, X. and Lafferty, J. and Rosenfeld, R., 2005).

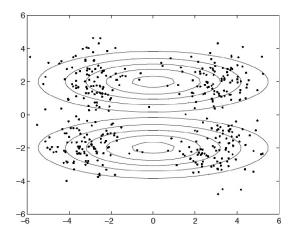


Figura 3: Dados não-rotulados atrapalhariam a correta separação das classes. Fonte: (Müller, K. R. and Zien, A., 2008).

# 2.3.1 Breve Histórico do Aprendizado Semi-Supervisionado

Os primeiros trabalhos em CSS assumiam duas classes, cada uma com uma distribuição normal. Desta forma, os dados como um todo resultam em um *modelo de mistura* (*mixture model*). São necessárias somente de duas amostras rotuladas, uma de cada classe, para determinar completamente o modelo de mistura.

Uma variante é o *self-training*: o classificador é treinado com os dados rotulados e depois aplicado aos não rotulados. Os pontos não rotulados que foram rotulados de forma mais confiante, são então adicionados ao conjunto de dados rotulados e então o classificador é treinado novamente. Este procedimento é chamado também de *self-teaching* e *bootstrapping*.

Ambos estes métodos são utilizados há muito tempo e continuam comuns devido à sua simplicidade conceitual e algorítmica.

Métodos mais recentes incluem o *co-training* (em que dois classificadores utilizam partes disjuntas dos atributos de cada objeto e utiliza os novos objetos rotulados um do outro) e *SVM transdutivo*.

Na última década, o aprendizado semisupervisionado baseado em grafos atraiu grande atenção. Estes métodos iniciam-se com grafos cujos nós são os objetos rotulados e não rotulados e cujas arestas tem pesos que refletem a similaridade entre os objetos. Assume-se que objetos conectados por arestas de peso alto tendem a ter o mesmo rótulo e que os rótulos podem propagar-se pelo grafo. A seguir daremos uma definição e introdução mais formal sobre os grafos e sobre redes complexas.

#### 2.4 Grafos e Redes Complexas

Está fora do escopo deste trabalho prolongarse com a definição de grafos e redes complexas. Nesta parte é dado uma breve introdução e histórico sobre o tema. O estudo dos grafos é matéria da *teoria dos grafos*.

Dito de forma direta, um grafo é uma representação abstrata de um conjunto de objetos no qual alguns destes objetos são ligados. Tais objetos são geralmente chamados de *nós* ou *vértices*. As ligações entre vértices são comumente chamadas de *arestas*. Tipicamente, um grafo é visualmente representado por um conjunto de pontos ou bolas (os vértices) ligados por linhas (arestas).

Embora uma grande parte da teoria dos grafos seja elaborada apenas sobre esta estrutura básica descrita acima, existem diversos possíveis acréscimos estruturais. Provavelmente o mais importante é o de especificar uma direção para as arestas, resultando em *arestas direcionadas*. A segunda característica diferencial de alguns grafos é a existência grandezas numéricas associadas às arestas, que resultam nas chamadas *arestas com pesos*.

Outras características dos grafos são: mais

de um tipo de aresta (possibilitando mais de uma aresta entre cada par de vértices); pesos associados aos nós; atributos diversos associados às arestas e aos nós (p.ex. cor); arestas incompletas; qualquer mistura de todas as características mencionadas acima.

Alguns grafos são clássicos, citamos alguns deles. Grafo regular é aquele no qual cada vértice tem o mesmo número de arestas. O grafo completo possui aresta entre qualquer par de vértices e também é chamado de *click*. Um grafo conectado é aquele que apresenta um caminho de sucessão de arestas entre qualquer par de vértices. Grafo randômico é um grafo construído com uma probabilidade de se ter uma aresta entre dois vértices. Similarmente, um grafo randômico é aquele que exibe *as mesmas características* que um grafo construído desta forma.

Formalmente falando, um grafo é um par ordenado G=(V,E) composto de um conjunto de vértices V e um conjunto E de arestas, que são subconjuntos de V com dois elementos. Estes são os grafos simples (com somente uma aresta possível entre cada par de vértices e sem aresta entre um vértice e ele mesmo) e não direcionado. Caso o conjunto E seja composto de pares ordenados de elementos de V, o grafo é direcionado.

A medida mais comum de um grafo é o grau, que é a quantidade de arestas ligadas a um vértice. Caso o grafo seja direcionado, podemos dividir o grau em duas quantidades: grau de entrada e grau de saída de um vértice. Caso o grafo tenha arestas com peso, pode-se falar em força de um vértice que é a soma dos pesos das arestas ligadas ao vértice. Similarmente, podemos num grafo direcionado com peso, falar em força de entrada e força de saída de um vértice.

Antes de continuar com as medidas de um grafo, para o nosso caso é pertinente citar o que é uma *rede complexa*. Dito de forma direta, uma rede complexa é um grafo com uma grande quantidade de vértices e arestas. Muitos pequisadores definem as redes complexas como sendo grafos de grandes dimensões que exibem características não triv-

iais, isto é, que diferem de um grafo regular ou randômico. Esta definição é invalidada pelo fato de que se fala em rede complexa randômica e em rede complexa regular. Podese sim entender que a área de estudo das redes complexas se interessa pelos casos de redes complexas (comumente) encontrados na natureza que exibem características não triviais.

Outra diferença entre as áreas de grafo e redes complexas é que enquanto a área de grafos tradicionalmente é vinculada à extração de propriedades matemáticas de grafos em geral ou com construções particulares, a área de redes complexas é marcada mais acentuadamente pela observação das redes complexas que representam fenômenos humanos ou naturais e suas divergências se comparadas às redes mais triviais (regulares e randômicas). Exemplos destas características peculiares são as conhecidas *lei livre de escala* e *lei de pequeno mundo*.

## 2.4.1 Breve Histórico da Teoria dos Grafos e das Redes Complexas

A estrutura de pontos e retas é tão simples que é difícil precisar quando foi utilizada pela primeira vez, se é que isso é possível. O artigo escrito por Leonhard Euler chamado "Konigsberg Bridge problem" é geralmente descrito como o primeiro trabalho tanto de topologia quanto de teoria dos grafos.

Mais de um século depois, Cayley estudou *árvores* através de formas analíticas que surgiam do cálculo diferencial. Tais estudos tiveram grandes consequências para a química teórica e esta fusão entre química e grafos está no centro da nomenclatura comum utilizada para grafos.

Em particular, o nome *grafo* foi introduzido por Sylvester em 1878 através de um artigo publicado na Nature.

No século XX a teoria dos grafos teve focos diferentes, passando pelo problema de coloração de grafos, topologia (em particular em conexão com a álgebra, p.ex. *Lei de Kirchoff* para circuitos), e, por fim, por teorias probabilísticas.

Já adentrando o terreno das chamadas re-

des complexas, os artigos que demonstravam as características *livre de escala* e de *pequeno mundo* para grafos grandes representando estruturas da natureza (Ferrer i Cancho and Sole, 2001; Albert and Barabasi, 2002) marcaram este novo grande interesse pela área.

Atualmente, as redes complexas são aplicadas com sucesso a diversas áreas (Costa et al., 2008), o que inclui neurociências (Sporns, 2002), física (Gfeller, 2007), linguística (Dorogovtsev and Mendes, 2001) e ciência da computação (Moura et al., 2003), para citar simplesmente alguns casos.

## 2.5 Aprendizado Semi-Supervisionado em Grafos

Diversas técnicas de CSS baseada em grafos tem sido desenvolvidas e utilizadas. Citamos os seguintes métodos, de forma ilustrativa e sem a pretensão de obtermos uma lista completa: algorítmo de regularização (Belkin, M. and Matveeva, I. and Niyogi, P., 2004), mincut (Blum, A. and Chawla, S., 2001), propagação de rótulo (Zhu, X. and Ghahramani, Z., 2003), modelos de mistura e maximização de entropia (Zhu, X. and Lafferty, J. and Rosenfeld, R., 2005).

De especial importância é a forma de construção do grafo. Uma discussão detalhada sobre o assunto é encontrada no terceiro capítulo de (Zhu, X. and Lafferty, J. and Rosenfeld, R., 2005).

Para um detalhamento das técnicas *mincut* e *propagação de rótulo* nos referimos ao outro trabalho apresentado nesta disciplina (Fabbri, R., 2010). Cabe lembrarmos aqui que, dadas as convenções envolvidas e explicitadas no referido artigo, a técnica de propagação de rótulo pode ser entendida pela seguinte fórmula:

$$f_u = (I - P_{UU})^{-1} P_{UL} Y_L \tag{1}$$

Isso nos permite resolver a propagação de rótulo diretamente sem a propagação iterativa.

Tabela 1: Pré-processamento da sentença "The projection of the dataset into three dimensions is essential for very large datasets". Note que as *stopwords* são removidas e as palavras restantes são lematizadas.

Original sentence	Pre-processed sentence	
The projection of the dataset	projection dataset	
into three dimensions is es-	three dimension be essen-	
sential	tial	
for very large datasets	very large dataset	

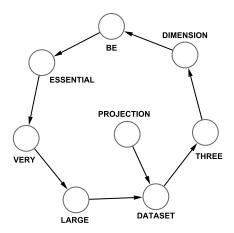


Figura 4: Exemplo de formação de rede resultante da senteça "The projection of the dataset into three dimensions is essential for very large datasets".

#### 3 Material e Métodos

#### 3.1 Redes de Palavras

#### 3.1.1 Regra de Formação

O método utilizado consiste em duas partes. A primeira é o processamento das palavras através da remoção das chamadas *stopwords* e de toda pontuação do texto. Em sequência efetuamos a lematização das palavras restantes. Tal procedimento pode ser visualizado na Figura 1.

A segunda parte da formação da rede consiste em tomar cada palavra como um vértice e em estabelecer arestas entre palavras consecutivas. Quando uma mesma palavra é encontrada novamente, ela é entendida como o

mesmo vértice. Este procedimento pode ser visto na Figura 4. Sabe-se que redes complexas assim formadas são do tipo livre de escala e pequeno mundo.

#### 3.1.2 Formação das Redes de Palavras

O corpus utilizado consiste em textos da Folha de São Paulo coletados durante dez anos. A seção do citado jornal do qual foram coletados os textos consistia em uma opinião positiva e outra negativa a respeito de determinado assunto.

Os textos foram divididos entre positivos e negativos. Em cada grupo, redes eram formadas aglomerando-se os textos até atingirem 1200 vértices. Como resultado obtemos diversas redes com 1200 vértices, cada rede formada somente por textos positivos ou negativos. Obtivemos 50 redes positivas e 50 redes negativas.

Na sequência realizamos as seguintes medidas: graus, coeficientes de clusterização, caminhos mais curtos, eficiência global, proximidade(closeness) e acessibilidade. Com cada uma destas medidas foram obtidas a média, desvio padrão, mediana e quartis.

#### 3.2 Processamento de Fala

Na seção 2.2 foram expostas as diferentes medidas usuais para o processamento de fala. Em nosso caso, utilizamos somente medidas de frequência fundamental. Utilizando o programa livre Praat (Boersma, P. and Weenink, D., 2010) e scripts próprios em Praat e Python, realizamos medidas de frequência fundamental a cada 10 milisegundos nos arquivos de áudio envolvidos.

# 3.2.1 Arquivos de Áudio e Medidas Derivadas da Sucessão de Frequências Fundamentais

Foram utilizados cinquenta trechos de fala de duração livre entre três e sete segundos. Todos foram retirados de um mesmo depoimento encontrado no projeto Iboruna (Gonçalves, S. C. L., ). Metade dos trechos foram trechos que, quando ouvidos no depoimento original, foram tidos como positivos. A outra metade são trechos que quando ouvidos foram

Tabela 2: % de textos classificados corretamente. O melhor classificador foi o de propagação de rótulos com 67,65 % de acerto (na média).

Método	Acerto
Bayesian Decision	60,39 %
Decision Tree	61,27 %
Decision Rules	64,08 %
Naive Bayes	64,79 %
Propagação de Rótulos	67,65 %

classificados como negativos. Isso segundo um único ouvinte, ou seja, os dados são de um único locutor e classificados segundo a paercepção de uma única pessoa (o autor deste trabalho).

Para cada arquivo, foi efetuada as seguintes medidas sobre a sequência de frequências fundamentais: média, desvio padrão, âmbito, mediana, porcentagem de medições dentre as localizadas no quinto inferior do âmbito, porcentagem de medições dentre as localizadas no quinto superior do âmbito.

#### 4 Resultados

#### 4.1 Redes de Palavras

Escolhemos 50% das redes para treinamento. Cada vez que os algorítmos eram rodados, as redes eram escolhidas ao acaso. O treinamento semi-supervisionado por propagação de rótulos deu a maior variação entre cada rodada, atingindo marcas de 80% de acerto. Na média, conseguimos um ganho de 2-3% em relação aos métodos usuais que utilizamos. O resultado está sumarizado nas Tabelas 2 e 3.

### 4.2 Trechos de Fala

Escolhemos também 50% das redes para treinamento. Cada vez que os algorítmos eram rodados, as redes eram escolhidas ao acaso. Aqui houve grande surpresa, com um aumento de 20% nas classificações corretas. Os resultados estão sumarizados nas Tabelas 4 e 5.

Tabela 3: Precisão e Cobertura para opiniões positivas e negativas.

Método	Precisão	Cobertura	Precisão	Cobertura
	Pos.	Pos.	Neg.	Neg.
Bayesian	70,2 %	49,3 %	77,3 %	71,4 %
Decision				
Decision	71,1 %	38,0 %	57,7 %	84,5 %
Tree				
Decision	70,0 %	49,3 %	60,9 %	78,9 %
Rules				
Naive	70,6 %	50,7 %	61,5 %	78,9 %
Bayes				
Propagação	60,0 %	83,3 %	72,7 %	44,4 %
de Rótulo				

Tabela 4: % de textos classificados corretamente. O melhor classificador foi o de propagação de rótulos com 67,65 % de acerto (na média).

`	
Método	Taxa de Ac- erto
RIP	57,78 %
Naive Bayes	53,33 %
Propagação de Rótulos	75,71 %

Tabela 5: Precisão e Cobertura para opiniões positivas e negativas.

Método	Precisão Pos.	Cobertura Pos.	Precisão Neg.	Cobertura Neg.
RIP Naive Bayes	52,4 % 47,1 %	55,0 % 40,0 %	62,5 % 57,1 %	60,0 % 64,0 %
Propagação de Rótulo	66,6 %	80,0 %	83,3 %	71,4 %

#### 5 Conclusão

Embora na literatura seja bastante enfatizado que a utilização de dados não rotulados para algoritmos de aprendizado de máquina é de grande utilidade quando os dados rotulados são poucos ou custosos, mostramos a utilidade do algorítmo de propagração de rótulos para dois casos diferentes em que os objetos rotulados eram metade dos objetos. Isso tanto desmistificou o uso do aprendizado semisupervisionado para o caso de uma grande quantidade de objetos rotulados que os testes com poucos objetos rotulados não chamou atenção e foi inclusive omitido deste trabalho.

Dada a supresa, estamos começando já a utilizar o método de propagação de rótulos nas pesquisas que estamos realizando no grupo. Os próximos passos são claramente conseguir mais exemplos para atigir resultados mais relevantes. É necessário obter mais trechos e analizar fala de outros locutores.

Outra via que queremos explorar e que tem com os conteúdos desenvolvidos nesta disciplina é a utilização de redes formadas a partir das sequências de medidas das falas. Desenvolvemos um método de criação de redes que exibe lei livre de escala para a distribuição de grau dos vértices. Também estamos estudando redes criadas pelo método conhecido como *redes de visibilidade*. Há ainda um terceiro método, este não explorado, que constrói a rede a partir da matriz de correlação relacionada à série temporal de origem.

A disponibilização dos algoritmos na rede<sup>3</sup> permite não só a troca facilitada de conhecimentos e tecnologias, mas também que possamos acessar o algoritmo de qualquer lugar. Esta disponibilidade também nos incentiva a arrumar o código de forma a deixá-lo facilmente utilizável.

#### Referências

Gonçalves, S. C. L. Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista.. http:://www.alip.ibilce.unesp.br/iboruna. .

<sup>&</sup>lt;sup>3</sup>svn co http://svn.assembla.com/svn/audioexperiments

- Zhu, X. and Lafferty, J. and Rosenfeld, R. 2005. Semisupervised learning with graphs. *Unpublished doctoral dissertation*.
- Zhu, X. 2006. Semi-supervised learning literature survey. *Citeseer*.
- Zhu, X. and Ghahramani, Z. 2002. Learning from labeled and unlabeled data with label propagation. *Citeseer*.
- Blum, A. and Chawla, S. 2001. Learning from labeled and unlabeled data using graph mincuts.
- Belkin, M. and Matveeva, I. and Niyogi, P. 2004. Regularization and Semi-supervised Learning on Large Graphs.
- Müller, K. R. and Zien, A. 2008. Semi-Supervised Learning.
- Chapelle, O. and Sch "olkopf, B. and Zien, A. 2006. Semi-supervised learning. *Citeseer*.
- Boersma, P. and Weenink, D. 2010. Praat: doing phonetics by computer [Computer program]. http://www.praat.org/.
- R. Ferrer i Cancho and R. V. Sole. 2001. The small world of human language. *Proceedings of the Royal Society of London B*, 268:2261.
- M. E. J. Newman. 2003. The Structure and Function of Complex Networks. *SIAM Review*, 45:167–256.
- R. Z. Albert and A.L. Barabasi. 2002. Statistical Mechanics of Complex Networks. Rev. Modern Phys., 74:47–97.
- L. F. da Costa, O. N. Oliveira Jr., G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, L. E. C. da Rocha. 2008. Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications. arXiv 0711.3199.
- O. Sporns. 2002. Network analysis, complexity, and brain function. *Complexity*, 8(1):56–60.
- D. Gfeller, P. LosRios, A. Caflisch and F. Rao. 2007. Complex network analysis of free-energy land-scapes. *Proceedings of the National Academy of Science USA*, 104 (6):1817–1822
- S. N. Dorogovtsev and J. F. F.Mendes. 2001. Language as an evolving word web. *Proceedings of the Royal Society of London B*, 268:2603.
- A. P. S. de Moura, Y. C. Lai and A. E. Motter. 2003. Signatures of small-world and scale-free properties in large computer programs. *Physical Review E*, 68(1):017102.

- Belkin, M. and Matveeva, I. and Niyogi, P. 2004. Regularization and semi-supervised learning on large graphs. *Learning theory, Springer*, 624–638.
- Jusczyk, P. W. 2000. The Discovery of Spoken Language. *The MIT Press*.
- Teixeira, A. 2009. Fala e Emoção. Tese de Doutoramento, Universidade de Aveiro.
- McLuhan, M. 1962. The Gutenberg Galaxy: The Making of Typographic Man. *University of Toronto Press*.
- Borges, C. de A. 2004. Dando corpo a palavra: um exercicio cenico sobre a voz. *Tese de Mestrado, Universidade de Campinas*.
- Dalgalarrondo, P. 2008. Psicopatologia e Semiologia dos Transtornos Mentais. Artmed.
- Yacoub, S. and Simske, S. and Lin, X., and Burns, J. 2003. Recognition of emotions in interactive voice response systems. Eighth European conference on speech communication and technology.
- Tóth, S. L. and Sztahó, D. and Vicsi, K. Speech Emotion Perception by Human and Machine. *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. LNCS (LNAI), Volume 5042, p. 213-224.*
- McGilloway, S. and Cowie, R. and Douglas-Cowie, E. and Gielen, S. and Westerdijk, M. and Stroeve, S. 2000 Approaching automatic recognition of emotion from voice: a rough benchmark. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.
- Fabbri, R. 2010 Aprendizado Semi-Supervisionado Baseado em Grafos. *Trabalho da matéria Matéria* SCC5882 sob orientação dos Profs. Dr. Zhao Liang e Dr. Alneu de Andrade Lopes.
- Toivanen, J. and Seppänen, T. and Väyrynen, E. 2009 Automatic Discrimination of Emotion from Voice: A Review of Research Paradigms. *Emotions in the Human Voice*, v. 3, cap. 5, p. 67-78.
- Sidarova, J. and McDonough, J. and Badia, T. 2009 Automatic Recognition of emotive Voice and Speech. *Emotions in the Human Voice, v. 3, cap. 12, p. 205-237.*
- Bird, S. and Klein, E. and Loper, E. 2009 Natural Language Processing with Python. *O'Reilly Media*.
- Jurafsky, D. and Martin, J. H. 2000 Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics. *Prentice Hall*.