NeuralSumm: Uma Abordagem Conexionista para a Sumarização Automática de Textos[1]

Thiago Alexandre Salgueiro Pardo
Lucia Helena Machado Rino
Maria das Graças Volpe Nunes

Núcleo Interinstitucional de Lingüística Computacional (NILC/São Carlos)
CP 668 – ICMC-USP, 13560-970 São Carlos, SP, Brazil
http://www.nilc.icmc.usp.br
{thiago@nilc.icmc.usp.br; lucia@dc.ufscar.br; gracan@icmc.usp.br}

Sumário

Este artigo apresenta uma abordagem conexionista para a sumarização automática de textos. Por meio de uma rede neural do tipo SOM (*self-organizing map*) treinada para classificar as sentenças de um texto-fonte de acordo com seu grau de importância, o sistema NeuralSumm (*NEURAL network for SUMMarization*) produz o extrato correspondente. O artigo também apresenta a avaliação do NeuralSumm.

Abstract

This paper presents a conexionist approach for the automatic text summarization problem. By means of a SOM (self-organizing map) neural network trained for classifying the sentences of a source text according to their importance, the NeuralSumm (NEURAL network for SUMMarization) system generates the corresponding extract. The paper also presents the NeuralSumm evaluation.

1. Introdução

Sumarização automática de textos é o processo de se produzir uma versão mais curta de um texto-fonte (Mani and Maybury, 1999). O resultado desse processo pode ser um extrato (do termo *extract*, em inglês) ou um sumário (do termo *abstract*, em inglês): o primeiro corresponde à simples justaposição de sentenças do texto-fonte consideradas importantes; o segundo, por sua vez, altera a estrutura e/ou o conteúdo das sentenças originais, fundindo-as e/ou reescrevendo-as, para generalizar ou especificar as informações (Mani, 2001).

O sistema apresentado neste artigo, chamado NeuralSumm (*NEURAL network for SUMMarization*), produz extratos utilizando uma técnica de Aprendizado de Máquina, uma rede neural do tipo SOM (*self-organizing map*) (Kohonen, 1982), para identificar as sentenças importantes de um texto-fonte que irão compor seu extrato. A classificação das sentenças em graus de importância é feita pela rede neural com base em *features* (características) extraídas durante o processo de sumarização.

Outros trabalhos já fizeram uso de técnicas de Aprendizado de Máquina para fins de Sumarização Automática de Textos, mas não se sabe de nenhum que utilize redes neurais do tipo SOM até o presente momento. Essa rede neural organiza as informações aprendidas em *clusters* de similaridade em função das *features* apresentadas. Justamente por isso, diz-se que as redes neurais do tipo SOM são as que mais se aproximam da forma de funcionamento do cérebro humano. Desta forma, no

[1] Este trabalho conta com o apoio financeiro da FAPESP e do CNPq.

NeuralSumm, as sentenças de um texto-fonte recebem sua classificação de acordo com os clusters da rede em que se enquadram.

Dentre os trabalhos relacionados a este, destacam-se os trabalhos de Kupiec et al. (1995), Aretoulaki (1996), Larocca Neto et al. (2002) e Teufel e Moens (2002). Kupiec et al. e Teufel e Moens realizaram uma análise de *features* superficiais em sentenças de textos científicos que pudessem indicar sua relevância para compor um sumário/extrato. Na mesma linha, Larocca Neto et al. consideraram um conjunto mais amplo de *features* que abrangeu (a) a distribuição de palavras e medidas estatísticas, (b) as medidas de coesão das sentenças e (c) a estrutura argumentativa do texto a sumarizar. Os textos utilizados por eles eram artigos de revistas sobre computação em geral. Ambos os trabalhos utilizaram classificação *bayesiana*. Larocca et al. também avaliaram o desempenho do classificador C4.5 (Quinlan, 1993) para a mesma tarefa. Aretoulaki, em especial, fez uma análise detalhada de textos jornalísticos e científicos de vários domínios para reconhecer *features* genéricas que representassem bem o conteúdo das sentenças de qualquer gênero textual e que pudessem ser aprendidas por uma rede neural. Ela considerou *features* de diversas naturezas, desde superficiais até pragmáticas, que pudessem ser identificadas em uma análise textual superficial. Sua rede neural, entretanto, não era do tipo SOM. No NeuralSumm, as *features* consideradas são um subconjunto das dos trabalhos acima, sendo, em sua grande maioria, independentes de gênero textual e domínio. De fato, somente uma delas não é independente, mas pode ser facilmente adequada a outros contextos, conforme será discutido posteriormente.

A próxima seção descreve o NeuralSumm e suas características, enquanto a Seção 3 apresenta a avaliação do sistema. As conclusões são apresentadas na Seção 4.

2. Descrição do NeuralSumm

O NeuralSumm utiliza uma rede neural do tipo SOM[2] para classificar cada sentença de um texto-fonte de acordo com seu grau de importância. Neste trabalho, considerou-se que as sentenças podem ser classificadas como *essenciais*, *complementares* ou *supérfluas*: sentenças *essenciais* são aquelas que transmitem a idéia principal de um texto; sentenças *complementares* são aquelas que acrescentam conteúdo à idéia principal, complementando-a; sentenças *supérfluas*, por sua vez, não acrescentam conteúdo algum. Idealmente, assume-se que sentenças essenciais devem sempre estar no extrato, que sentenças complementares podem ou não ser incluídas nele e que sentenças supérfluas devem ser omitidas.

Uma vez classificadas as sentenças de um texto-fonte, o extrato correspondente é produzido pelo NeuralSumm observando-se (a) a classificação obtida pelas sentenças, (b) a taxa de compressão desejada para o extrato e (c) a pontuação das sentenças dada pela distribuição de palavras do texto, que é uma das *features* consideradas pela rede neural. A pontuação das sentenças é utilizada para decidir por sentenças para compor o extrato em casos de empate na classificação. A seguir, é descrita a forma como os dados de treinamento foram obtidos, as *features* utilizadas e o treinamento da rede.

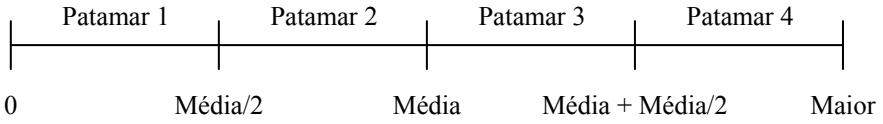
2.1. Treinamento da rede neural e descrição das features

A rede neural do NeuralSumm foi treinada com sentenças de um corpus de 10 textos científicos (introduções de teses e dissertações com aproximadamente 530 palavras e 19

[2] A rede utilizada, em particular, possui uma arquitetura de 14x14 neurônios e foi treinada com precisão de 1 milionésimo.

sentenças cada) do domínio da Computação em Português do Brasil, chamado *CorpusDT*[3] (Feltrim et al., 2001). Os textos desse corpus tiveram suas sentenças classificadas de acordo com sua importância (ou seja, segundo os valores *essencial*, *complementar* ou *supérflua*) por 10 juízes lingüistas computacionais e falantes nativos do Português do Brasil. Para cada sentença, extraiu-se um conjunto de 8 *features*, associando a ele a classificação indicada pela maioria dos juízes. Para o treinamento, propriamente dito, cada conjunto de *features* foi apresentado à rede, associando-se sua classificação ao neurônio em que se enquadrou. Desta forma, formaram-se *clusters* ("marcados" com as classes) representando as três classificações possíveis. As *features* utilizadas, suas descrições e valores possíveis são:

1. *Tamanho da sentença*: sentenças longas normalmente apresentam maior conteúdo informativo, sendo, portanto, relevantes para o texto (Kupiec et al., 1995). No NeuralSumm, as sentenças de um texto são enquadradas em uma escala de tamanho de 4 patamares. Considerando que a variável *Média* representa a média do tamanho das sentenças do texto e que a variável *Maior* representa o tamanho da maior sentença do texto, a escala é dada por:



2. *Posição da sentença no texto*: a posição da sentença pode indicar sua relevância (Baxendale, 1958; Aretoulaki, 1996). No NeuralSumm, uma sentença pode estar no *início* (primeiro parágrafo), no *fim* (último parágrafo) ou no *meio* (parágrafos restantes) do texto;
3. *Posição da sentença no parágrafo a que pertence*: a posição da sentença no parágrafo também pode indicar sua relevância (Baxendale, 1958). No NeuralSumm, uma sentença pode estar no *início* (primeira sentença), no *fim* (última sentença) ou no *meio* (posições restantes) do parágrafo;
4. *Presença de palavras-chave na sentença*: as palavras-chave são comumente utilizadas para expressar a idéia principal do texto, tendendo a se repetir no decorrer do texto (Luhn, 1958). Assim, uma sentença pode conter (*True*) ou não (*False*) palavras-chave do texto. As palavras-chave de um texto são as palavras significativas com mais alta freqüência.
5. *Presença de palavras da "gist sentence" na sentença*: sentenças que possuem palavras da *gist sentence*, isto é, a sentença que melhor expressa a idéia principal do texto[4], tendem a ser relevantes, pois contribuem para a idéia principal do texto (Pardo, 2002a). Uma sentença pode conter (*True*) ou não (*False*) palavras da *gist sentence*.
6. *Pontuação da sentença com base na distribuição das palavras do texto*: sentenças com alta pontuação normalmente são relevantes para o texto (Black and Johnson, 1988). A pontuação de uma sentença, neste trabalho, é a soma das freqüências de suas palavras dividida pelo número de palavras da sentença. Essa pontuação também é enquadrada em uma escala similar à da *feature* 1, indicando o patamar a que a sentença pertence (*1*, *2*, *3* ou *4*).

---

[3] http://www.nilc.icmc.usp.br/nilc/tools/corpora.htm

[4] A *gist sentence* de um texto é determinada pela aplicação do método GistKey, que também utiliza a distribuição de palavras no texto.

7. *TF-ISF da sentença*: sentenças com alto valor de TF-ISF (*Term Frequency - Inverse Sentence Frequency*) são sentenças representativas do texto (Larocca Neto et al., 2000). A medida TF-ISF é derivada da medida TF-IDF (*Term Frequency - Inverse Document Frequency*) (Salton, 1989) da área de Recuperação de Informação. Para cada palavra de uma sentença, a medida TF-ISF é calculada pela fórmula abaixo, onde F(w) é a freqüência da palavra w na sentença, n é o número de palavras da sentença a que w pertence e S(w) é o número de sentenças em que w aparece:

$$TF - ISF\,(w) = F(w) \times \frac{\log n}{S(w)}$$

   O valor TF-ISF de uma sentença será a média dos valores TF-ISF de cada uma de suas palavras. O valor TF-ISF de cada sentença também é enquadrado em uma escala similar à da *feature* 1, indicando o patamar a que a sentença pertence (*1*, *2*, *3* ou *4*).
8. *Presença de palavras indicativas na sentença*: palavras indicativas (*clue words*) normalmente indicam a importância do conteúdo das sentenças (Paice, 1981). Uma sentença pode conter (*True*) ou não (*False*) palavras indicativas. Essa *feature* é a única dependente de língua e do gênero e domínio do texto. No NeuralSumm, atualmente customizado para textos científicos em Português, as palavras indicativas consideradas são *avaliação*, *conclusão*, *método*, *objetivo*, *problema*, *propósito*, *resultado*, *situação* e *solução*.

2.2 Arquitetura

Para produzir o extrato de um texto-fonte, o NeuralSumm realiza os seguintes passos (vide Figura 1):

1. segmentação do texto-fonte em sentenças;
2. extração de *features* das sentenças:
   a) conforme sugerido por Witten et al. (1994), para cada sentença, são realizados os processos de (i) remoção de *stopwords* (palavras muito comuns ou sem conteúdo que são consideradas irrelevantes), (ii) transformação das palavras em suas formas-base (similar ao processo de *stemming*) utilizando-se um léxico e (iii) *case folding*, ou seja, transformação de todas as letras das palavras para a mesma caixa, para uniformização
   b) com base nas sentenças modificadas pelos processos acima, realiza-se, então, a extração das *features*
3. classificação do conjunto de *features* de cada sentença como *essencial*, *complementar* ou *supérfluo* pela rede neural, de acordo com o *cluster* da rede em que cada conjunto se enquadrou;
4. produção do extrato.

A seleção de sentenças para a produção do extrato (passo 4) acontece da seguinte forma:

- são selecionadas somente sentenças classificadas como *essenciais* e *complementares*;
- caso todas as sentenças do texto-fonte sejam classificadas como *supérfluas*, elas são ordenadas pela pontuação obtida pela distribuição das palavras (*feature* 6), selecionando-se, então, aquelas com pontuação mais alta.

Ambos os casos, entretanto, são restringidos pela taxa de compressão[5], especificada pelo usuário do sistema. No primeiro caso, quando a taxa de compressão restringe o número de sentenças selecionadas, as sentenças *essenciais* têm prioridade sobre as *complementares*, das quais as que possuem pontuação mais alta obtida pela distribuição das palavras (*feature* 6) têm prioridade sobre as demais.

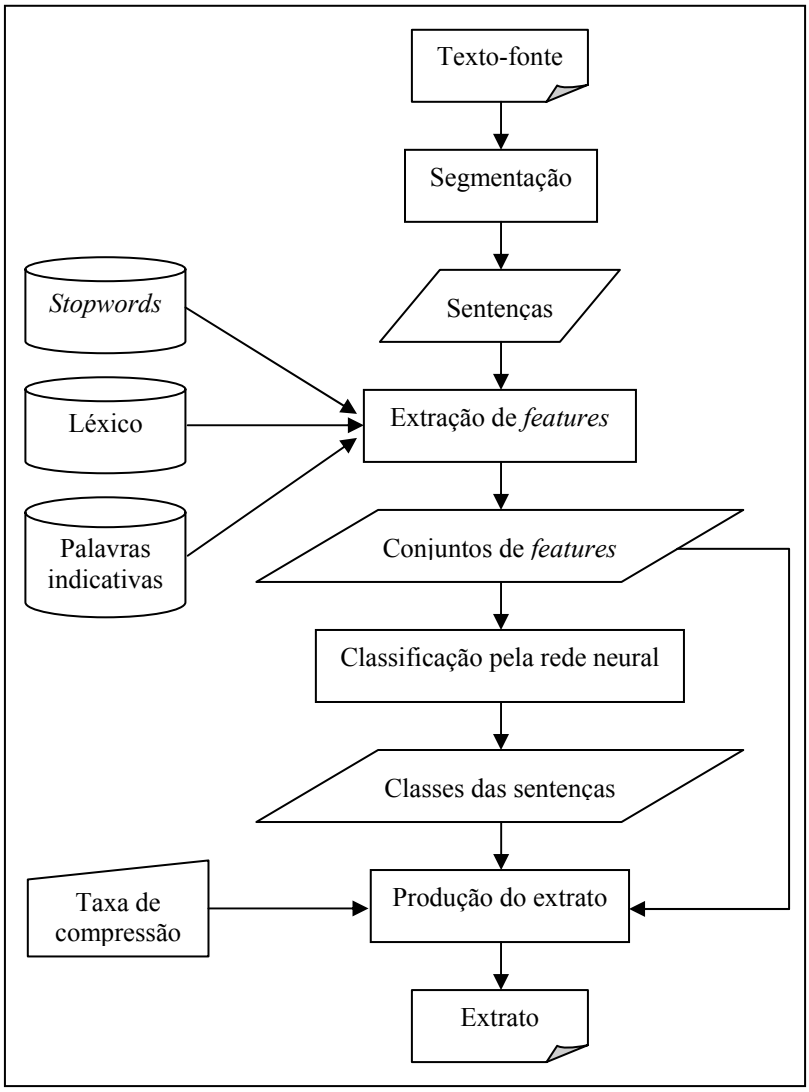


Figura 1 – Arquitetura do NeuralSumm

Como exemplo de sumarização, a Figura 3 mostra o extrato produzido pelo NeuralSumm para o texto da Figura 2, com uma taxa de compressão de 60%. Analisando-se o extrato produzido, nota-se que ele realmente contém as duas sentenças mais importantes do texto-fonte. Entretanto, o extrato apresenta uma abreviatura ("AF") sem sua definição ("Autômatos Finitos"). É importante dizer que métodos extrativos normalmente não tratam referências anafóricas de qualquer tipo. Quando se aumenta a taxa de compressão para 70%, é produzido um extrato que contém somente a primeira sentença do extrato da Figura 3, ou seja, somente a sentença mais importante do texto-fonte.

[5] Taxa de compressão é calculada como 1 - (tamanho do extrato / tamanho do texto-fonte)

A representação de grandes dicionários de língua natural, principalmente nos casos em que se trabalha com vários milhões (ou dezenas de milhões) de palavras, é um interessante problema computacional a ser tratado dentro da área de Processamento de Língua Natural (PLN). Autômatos Finitos (AFs), largamente usados na construção de compiladores, são excelentes estruturas para representação desses dicionários, permitindo acesso direto às palavras e seus possíveis atributos (gênero, número, grau, etc).

Um dicionário contendo mais de 430.000 palavras da língua portuguesa sem atributos, cuja representação em formato texto ocupa mais de 4,5Mb, pode ser convertido em um AF compactado de apenas 218Kb, conforme será apresentado (o processo de conversão para esse caso levou 4:17Seg, rodando em um Pentium166/Linux).

Esse trabalho tem por objetivo investigar métodos eficientes de construção, representação e minimização de AFs de grande porte (dezenas de milhões de estados), a fim de propor um conjunto de estruturas e algoritmos para criação de um sistema eficiente e versátil de representação de grandes léxicos de língua natural.

Outro objetivo é o de investigar métodos de representação para os atributos das entradas léxicas, de forma a permitir a criação de um formato otimizado capaz de ser inserido diretamente no AF, com o menor impacto possível no tamanho final desse último.

Será também desenvolvido um protótipo do sistema, para efeito de testes e confirmação de eficiência do sistema.

Figura 2 – Texto-fonte

Esse trabalho tem por objetivo investigar métodos eficientes de construção, representação e minimização de AFs de grande porte (dezenas de milhões de estados), a fim de propor um conjunto de estruturas e algoritmos para criação de um sistema eficiente e versátil de representação de grandes léxicos de língua natural.

Outro objetivo é o de investigar métodos de representação para os atributos das entradas léxicas, de forma a permitir a criação de um formato otimizado capaz de ser inserido diretamente no AF, com o menor impacto possível no tamanho final desse último.

Figura 3 – Extrato

A avaliação do NeuralSumm é descrita a seguir.

3. Avaliação

A avaliação do NeuralSumm objetivou (a) medir o desempenho da rede neural do tipo SOM em classificar sentenças corretamente como *essenciais*, *complementares* e *supérfluas* e (b) verificar a proximidade dos extratos gerados automaticamente com seus sumários autênticos, isto é, aqueles produzidos pelos próprios autores dos textos-fonte. As subseções seguintes descrevem as duas avaliações.

3.1. Desempenho da rede neural

Aqui, considera-se que, se a classificação da rede tiver um bom desempenho, os extratos produzidos também serão satisfatórios, já que as sentenças terão sua importância satisfatoriamente calculada.

Para treinamento e teste, foram utilizadas as sentenças do *CorpusDT* (já descrito na subseção 2.1). Devido ao seu pequeno tamanho, realizou-se a *10-fold cross-validation*. Para fins de comparação, o mesmo procedimento foi realizado para os classificadores Naive-Bayes e C4.5 (tanto para árvores quanto para regras de decisão). As taxas de erro obtidas são mostradas na Tabela 1. Como é possível notar, a rede neural do NeuralSumm obteve a menor taxa de erro em relação aos outros classificadores. Esse resultado corrobora Larocca Neto et al. (2002), que afirmam que a

busca de classificadores mais adequados para a Sumarização Automática de Textos é tão importante quanto a seleção de *features* representativas para o problema em questão.

Tabela 1 – Desempenho dos classificadores

Classificadores	**Taxa de erro**
Rede SOM	43%
C4.5 – Regras de decisão	51%
C4.5 – Árvores de decisão	53%
Naive-Bayes	57%

Sabe-se que a classificação humana de sentenças de acordo com sua importância para compor um extrato é uma tarefa bastante subjetiva, pois depende, principalmente, (a) do modelo de mundo e do conhecimento prévio do leitor sobre o assunto abordado no texto-fonte (Sparck Jones, 1999), (b) do "tempo disponível" para a leitura (Hovy, 1988) e (c) da intenção comunicativa percebida pelo leitor (Moore and Paris, 1993; Rino, 1996; Pardo, 2002b). Um leitor com "pressa" e com bom conhecimento do assunto do texto pode achar que o extrato deva conter somente sua informação nova, enquanto um leitor com pouco conhecimento pode precisar de um extrato com mais informação, para contextualizá-la. De acordo com a intenção comunicativa percebida pelo leitor, extratos de diferentes conteúdos podem, assim, ser produzidos. De fato, Mitra et al. (1997) constataram uma concordância na seleção de sentenças para formar extratos de textos, por juizes humanos, de somente 46%. Em face disso, foi realizada uma nova avaliação do desempenho da rede neural do NeuralSumm, permitindo-se uma certa flexibilidade em sua classificação. Assim, assumiu-se que (a) sentenças da classe *complementar* também podem ser classificadas como *essenciais* e (b) sentenças da classe *supérflua* também podem ser classificadas como *complementares*, buscando reproduzir, na medição do desempenho da rede, um pouco da subjetividade da classificação humana.

Embora se tenha permitido que as sentenças tivessem um maior grau de importância, o contrário não foi permitido, isto é, as sentenças não puderam ser rebaixadas, pois não se queria correr o risco de excluir do extrato sentenças consideradas *essenciais*. Com isso, a taxa de erro da rede neural caiu para 27%. Em termos práticos, essa não é uma taxa ruim, pois não se sabe em que nível tal taxa influi na qualidade do extrato. Por exemplo, caso algumas sentenças *essenciais* tenham sido mal classificadas como *supérfluas* e não tenham sido incluídas no extrato, este provavelmente perderá qualidade; por outro lado, caso algumas sentenças *supérfluas* tenham sido mal classificadas como *essenciais* e tenham sido incluídas no extrato, isto pode ou não influir na qualidade do extrato. Podem-se imaginar, ainda, casos em que sentenças *complementares* que não seriam incluídas no extrato (devido, por exemplo, a uma alta taxa de compressão) possam ser mal classificadas como *supérfluas*, casos estes que não alteram a qualidade do extrato. De fato, para verificar a influência real da má classificação das sentenças de um texto-fonte na qualidade do extrato, faz-se necessária uma avaliação por juízes humanos, a qual está prevista nos trabalhos futuros, conforme se discute na Seção 4.

3.2. Proximidade com os sumários autênticos

Como os sumários autênticos geralmente não preservam as sentenças dos textos-fonte da forma que elas aparecem (devido à reescrita usual realizada por escritores humanos), realizar a comparação desses sumários com os extratos produzidos automaticamente se

torna uma tarefa difícil. Para resolver esse problema, tem-se adotado os sumários "ideais" (os chamados *gold-standards*) (Teufel and Moens, 1999; Mani and Bloedorn, 1998) para comparação, os quais consistem na versão extrativa dos sumários autênticos. Dessa forma, a verificação entre sentenças do sumário ideal e sentenças do extrato gerado automaticamente se torna possível. Para se produzir o sumário ideal, costuma-se utilizar a medida do co-seno (Salton, 1989): para cada sentença do sumário autêntico, procura-se a sentença correspondente no texto-fonte mais semelhante. Isto é feito pela co-ocorrência de palavras: quanto mais palavras do sumário autêntico uma sentença do texto-fonte tem, maior sua chance de apresentar o mesmo conteúdo da sentença do sumário, podendo, portanto, ser utilizada para compor o sumário ideal.

Para a avaliação em questão, foram coletados 10 textos científicos (introduções de teses e dissertações) da Computação (não contendo nenhum dos textos anteriores do *CorpusDT*) com seus respectivos sumários autênticos, para os quais foram gerados automaticamente os sumários ideais[6]. Estes foram comparados aos extratos gerados pelo NeuralSumm com uma taxa de compressão de 80%. Os resultados obtidos para cobertura e precisão foram, respectivamente, de 32% e 41%. A cobertura, neste caso, indica o número de sentenças do extrato que coincidem com as do sumário ideal; a precisão indica a razão entre o número de sentenças coincidentes com as do sumário ideal e o total de sentenças do extrato. Estes resultados, aparentemente insatisfatórios, não indicam que os extratos produzidos pelo NeuralSumm sejam ruins, pois: (a) a utilização da medida do co-seno pode incluir sentenças inadequadas nos sumários ideais, pois ela faz uso da simples co-ocorrência de palavras; (b) mais de uma sentença pode apresentar o mesmo conteúdo e, assim, pode-se gerar mais de um extrato, variando-se as sentenças e levando a uma incompatibilidade aparente entre o sumário ideal e o extrato, do ponto de vista de proximidade semântica; (c) a taxa de compressão afeta o número de sentenças que serão incluídas no extrato, sendo que diferentes taxas acarretam diferentes resultados em termos de cobertura e precisão. Neste contexto, acredita-se que essas medidas podem apenas dar um vislumbre do desempenho do sumarizador, sendo necessária uma avaliação dos resultados por juízes humanos. Na realidade, os valores de cobertura e precisão obtidos estão relativamente próximos dos valores que outros trabalhos obtiveram quando fizeram avaliação similar: Teufel e Moens (2002) atingiram cobertura de 44% e precisão de 65%[7]; Larocca Neto et al. (2002) atingiram cobertura e precisão de 51% (no melhor dos casos); Kupiec et al. (1995) atingiram cobertura e precisão de 42%; Saggion e Lapalme (2002) atingiram cobertura de 23% e precisão de 20%. De fato, acreditamos que os resultados desta avaliação são bastante satisfatórios em vista do pequeno tamanho do corpus de treinamento (somente 10 textos) e que, ao aumentarmos o tamanho desse corpus, podemos melhorar bastante os resultados.

4. Conclusões e comentários finais

Neste artigo, foi apresentado o NeuralSumm (*NEURAL network for SUMMarization*), um Sumarizador Automático de Textos que utiliza uma rede neural do tipo SOM (*self-organizing map*) para classificar em graus de importância as sentenças de um texto-

---

[6] O programa utilizado para gerar os sumários ideais encontra-se disponível para *download* em http://www.nilc.icmc.usp.br/~thiago/NeuralSumm.html

[7] Teufel e Moens não utilizaram a medida do co-seno para a produção dos sumários ideais. Os autores fizeram uso de uma medida mais sofisticada, que também considerava a ordenação das palavras além da co-ocorrência somente.

fonte a sumarizar. Neste contexto, sentenças classificadas como *supérfluas* serão descartadas, enquanto sentenças classificadas como *essenciais* e *complementares* serão selecionadas para compor o extrato correspondente.

Para a classificação das sentenças, são utilizadas *features* que, quase na totalidade, são independentes de língua, gênero textual e domínio. Apenas uma das *features* é dependente (que verifica a presença de palavras indicativas nas sentenças), mas pode ser facilmente customizada para novos contextos de aplicação do NeuralSumm, o qual, para este artigo, foi treinado, testado e avaliado com textos científicos da Computação em Português do Brasil. De fato, imagina-se que o NeuralSumm possa ser facilmente customizado para qualquer língua ocidental. Para isso, os repositórios de informação lingüística, isto é, o léxico (ou um processo de *stemming*) e a lista de *stopwords*, também devem ser adaptados.

Durante a avaliação do NeuralSumm, comprovou-se que o desempenho da rede neural utilizada foi bem melhor que o desempenho dos classificadores bayesiano e C4.5, indicando, como em outras pesquisas relacionadas, que a busca de classificadores mais adequados para tarefas de sumarização é tão importante quanto a seleção de *features* para a classificação. Até onde se sabe, este foi o primeiro trabalho a fazer uso de uma rede neural do tipo SOM para a Sumarização Automática de Textos. Em outra avaliação, verificou-se que a proximidade dos extratos produzidos pelo NeuralSumm com os sumários ideais está relativamente próxima dos resultados obtidos em outras pesquisas recentes. Nesta avaliação, o NeuralSumm foi treinado com um corpus de apenas 10 textos. Acreditamos que, ao aumentar o corpus de treinamento, poderemos melhorar bastante os resultados.

Os trabalhos futuros apontam, primeiramente, para a realização de uma avaliação subjetiva, com juízes humanos, da qualidade dos extratos gerados pelo NeuralSumm. Além disso, há planos para a realização de uma avaliação comparativa do NeuralSumm com outros Sumarizadores Automáticos disponíveis. Devem-se investigar, também, a utilização de outras *features*, mesmo que dependentes de gênero textual e domínio, que possam aprimorar o processo de classificação, e as possíveis combinações das *features* que produzam as melhores classificações e, portanto, os melhores extratos.

Referências

Aretoulaki, M. (1996). *COSY-MATS: A Hybrid Connectionist-Symbolic Approach To The Pragmatic Analysis of Texts For Their Automatic Summarisation*. PhD. Thesis. University of Manchester.

Baxendale, P.B. (1958). Machine-made index for technical literature – an experiment. *IBM Journal of Research and Development*, Vol. 2, pp. 354-365.

Black, W.J. and Johnson, F.C. (1988). A Practical Evaluation of Two Rule-Based Automatic Abstraction Techniques. *Expert Systems for Information Management*, Vol. 1, No. 3. Department of Computation. University of Manchester Institute of Science and Technology.

Feltrim, V.D.; Nunes, M.G.V.; Aluísio, S.M. (2001). *Um corpus de textos científicos em Português para a análise da Estrutura Esquemática*. Série de Relatórios do NILC. NILC-TR-01-4. Available for download in www.nilc.icmc.usp.br/nilc/pessoas/valeria.htm

Hovy, E. (1988). *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, Vol. 43, pp. 59-69.

Kupiec, J.; Pedersen, J.; Chen, F. (1995). A trainable document summarizer. *ACM SIGIR*, pp. 68-73.

Larocca Neto, J.; Santos, A.D.; Kaestner, A.A.C; Freitas, A.A. (2000). Generating Text Summaries through the Relative Importance of Topics. In the *Proceedings of the International Joint Conference IBERAMIA/SBIA*, Atibaia, SP.

Larocca Neto, J.; Freitas, A.A; Kaestner, A.A.C. (2002). Automatic Text Summarization Using a Machine Learning Approach. In the *Proceedings of the 16th Brazilian Symposium on Artificial Intelligence*, pp. 205-215. Porto de Galinhas/Recife.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, Vol. 2, pp. 159-165.

Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Co., Amsterdam.

Mani, I. and Bloedorn E. (1998). Machine Learning of Generic and User-Focused Summarization. In the *Proceedings of the Fifteenth National Conference on Artificial Intelligence* – AAAI'98, pp. 821-826.

Mani, I. and Maybury, M. T. (1999). *Advances in automatic text summarization*. MIT Press, Cambridge, MA.

Mitra, M.; Singhal, A.; Buckley, C. (1997). Automatic text summarization by paragraph extraction. In the *Proceedings of the ACL/EACL Workshop on Intelligent Scalable Text Summarization*. Madrid.

Moore, J.D. and Paris, C. (1993). Plannig Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*, Vol. 19, No. 4, pp. 651-694.

Paice, C. D. (1981). The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. Information Retrieval Research. Butterworth & Co. (Publishers).

Pardo, T.A.S. (2002a). *GistSumm: Um Sumarizador Automático Baseado na Idéia Principal de Textos*. Série de Relatórios do NILC. NILC-TR-02-13. Available for download in www.nilc.icmc.usp.br/nilc/~thiago

Pardo, T.A.S. (2002b). *DMSumm: Um Gerador Automático de Sumários*. Dissertação de Mestrado. Departamento de Computação. Universidade Federal de São Carlos. São Carlos - SP. Available for download in www.nilc.icmc.usp.br/nilc/~thiago

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Rino, L.H.M. (1996). *Modelagem de Discurso para o Tratamento da Concisão e Preservação da Idéia Central na Geração de Textos*. Tese de Doutorado. IFSC-USP. São Carlos - SP.

Saggion, H. and Lapalme, G. (2002). Generating Indicative-Informative Summaries with SumUM. *Computational Linguistics*, Vol. 28, N. 4, pp. 497-526.

Salton, G. (1989) *Automatic Text Processing. The Transformation, Analysis and Retrieval of Information by Computer*. Addison-Wesley.

Sparck Jones, K. (1999). Automatic Summarizing: factors and directions. In I. Mani and M. Maybury (eds.), *Advances in automatic text summarization*, pp. 1-12, The MIT Press.

Teufel, S. and Moens, M. (1999). Argumentative classification of extracted sentences as a first step towards flexible abstracting. In I. Mani and M. Maybury (eds.), *Advances in automatic text summarization*, MIT Press.

Teufel, S. and Moens, M. (2002). Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, Vol. 28, N. 4, pp. 409-445.

Witten, I.H.; Moffat, A.; Bell, T.C. (1994). Managing Gigabytes. *Van Nostrand Reinhold*. New York.