

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

GUSTAVO BENNEMANN DE MOURA

**Redes Neurais Recorrentes para a Classificação de Estruturas
Retóricas**

Maringá

2018

GUSTAVO BENNEMANN DE MOURA

Redes Neurais Recorrentes para a Classificação de Estruturas Retóricas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientadora: Prof^a. Dr^a. Valéria Delisandra Feltrim

Maringá
2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

Moura, Gustavo Bennemann de

M929r Redes neurais recorrentes para a classificação de
 estruturas retóricas/ Gustavo Bennemann de Moura. --
 Maringá, 2018.
 62 f., il., color., figs.

 Orientadora: Prof.a. Dr.a. Valéria Delisandra
 Feltrim.

 Dissertação (mestrado) - Universidade Estadual de
 Maringá, Centro de Tecnologia, Programa de Pós-
 graduação em Ciência da Computação, 2018.

 1. Estruturas retóricas. 2. Processamento de
 linguagem natural. 3. Inteligência artificial. 4.
 Redes neurais recorrentes. 5. Redes neurais. I.
 Feltrim, Valéria Delisandra, orient. II.
 Universidade Estadual de Maringá. Centro de
 Tecnologia. Programa de Pós-Graduação em Ciência da
 Computação. III. Título.

CDD 22. ED.005.1

JLM-000189

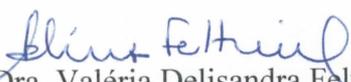
FOLHA DE APROVAÇÃO

GUSTAVO BENNEMANN DE MOURA

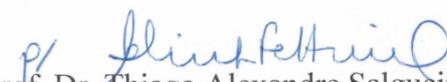
Redes neurais recorrentes para a classificação de estruturas retóricas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA


Profa.-Dra. Valéria Delisandra Feltrim
Universidade Estadual de Maringá – DIN/UEM


Prof. Dr. Marcos Aurélio Domingues
Universidade Estadual de Maringá – DIN/UEM


Prof. Dr. Thiago Alexandre Salgueiro Pardo
Universidade de São Paulo – ICMC/USP
participação por videoconferência

Aprovada em: 19 de fevereiro de 2018.

Local da defesa: Sala 120, Bloco C56, *campus* da Universidade Estadual de Maringá.

AGRADECIMENTOS

Aos meus pais, por sempre me apoiarem e serem exemplos de determinação, sinceridade e honestidade. Pela paciência e todas as oportunidades e possibilidades que me deram para que eu pudesse chegar até aqui. À minha esposa, Rita Mitushasi por ter me acompanhado e apoiado em todas as etapas do desenvolvimento deste trabalho. Por ter compartilhado momentos e me fazer uma pessoa feliz e melhor. À professora Dra. Valéria Delisandra Feltrim, por ter me incentivado, confiado e ajudado no desenvolvimento deste trabalho. A todos meus amigos de mestrado, que fizeram parte do meu dia a dia, compartilhando conhecimentos e descontração. Por fim, a todos que fazem parte da minha vida, que contribuíram para o indivíduo que sou hoje.

Redes Neurais Recorrentes para a Classificação de Estruturas Retóricas

RESUMO

A retórica trata de formas de persuadir, informar, ou motivar um interlocutor. Isso pode ser feito de diversas maneiras, como exemplo: por meio do uso de argumentações, ou apresentando definições. Uma característica importante da retórica é sua estrutura, ou seja, a organização de como tais argumentos são apresentados no texto. Este trabalho de mestrado teve como objetivo investigar o uso de redes neurais recorrentes para a classificação de estrutura retórica. Para isso, foi necessário primeiramente construir um *corpus* com uma grande quantidade de resumos de artigos científicos com suas estruturas retóricas anotadas para que fosse possível o treino da arquitetura de rede proposta. A rede neural recorrente (RNN) proposta utiliza camadas LSTM (Long short-term memory), para evitar os problemas que são comuns em RNNs, como a explosão e desvanecimento de gradiente. Também foi utilizada uma camada *encoder-decoder* para obter representações intermediárias das sentenças dos resumos, de modo que fosse possível transformar a sequência de palavras na sentença do resumo em um vetor de tamanho fixo que representasse essa sequência. Para comparação, foram implementadas duas *baselines* que utilizam os algoritmos de aprendizagem CRF e SVM, com características comprovadas por outros trabalhos na literatura. Os resultados obtidos pela rede proposta foram encorajadores, pois quando comparados os melhores resultados da rede com os melhores resultados das *baselines*, foi possível observar 3,18% de ganho na acurácia da classificação de estruturas retóricas em resumos como um todo. Na classificação por sentenças, foram observados ganhos de 1,02% na acurácia e 0,88% na medida de F1-macro. Assim, foi verificado que é possível o uso de redes neurais recorrentes para a classificação de papéis retóricos de forma eficaz.

Palavras-chave: estruturas retóricas, redes neurais recorrentes, redes neurais

Recurrent Neural Networks for Rhetorical Structures Classification

ABSTRACT

Rhetoric is the art of using language effectively and persuasively. One important factor about rhetoric is its structure, i.e. the organization of how such arguments are presented in the text. This Master's work intend to evaluate if recurrent neural networks are able to contribute in the classification of rhetorical structure. To do this, it was necessary first to create a corpus with a large number of abstracts of scientific articles with their rhetorical structures annotated so that it was possible to train the proposed neural network architecture. The proposed recurrent neural network (RNN) uses LSTM (Long Short-Term Memory) layers to avoid the problems that are common in RNNs (vanishing and exploding gradients). An encoder-decoder layer was also used to obtain intermediate representations of the abstracts sentences so that it was possible to transform the word sequence into the abstract sentence into a fixed-length vector representing that sequence. For comparison, two baselines were implemented using the CRF and SVM machine learning algorithms, and with features that were used in other works in the literature. The results obtained by the proposed neural network were satisfactory. When comparing the best results of the network with the best results of the baselines, it was possible to observe 3.18% gain in the accuracy of the classification of rhetorical structures in abstracts as a whole. In the sentence classification, gains of 1.02% in accuracy and 0.88% in F1-macro measurement were observed. Thus, it was found that it is possible the use of recurrent neural networks for classification rhetorical roles effectively.

Keywords: rhetorical structure, recurrent neural networks, neural networks

LISTA DE FIGURAS

Figura 1.1	Exemplo da organização de um texto acadêmico	11
Figura 2.1	Modelo de estrutura argumentativa proposta por Weissberg e Buker (1990)	15
Figura 2.2	Modelo de estrutura argumentativa proposta por Swales (1994)	16
Figura 2.3	Exemplo de um resumo com suas sentenças atribuídas a papéis retóricos.	17
Figura 2.4	Neurônio Artificial	19
Figura 2.5	Rede neural de múltiplas camadas	21
Figura 2.6	Rede neural recorrente	23
Figura 2.7	Rede neural recorrente com passos expandidos	23
Figura 2.8	Rede neural recorrente <i>LSTM</i>	25
Figura 2.9	Arquitetura de um <i>encoder-decoder</i>	26
Figura 2.10	Exemplo de uma representação distribuída.	28
Figura 4.1	Exemplo de um resumo de artigo da literatura biomédica dividido em seções.	36
Figura 4.2	Exemplo do XML da MEDLINE/PubMed®	37
Figura 4.3	Exemplo de um resumo do <i>corpus</i>	37
Figura 4.4	Histograma do tamanho dos resumos no <i>corpus</i>	39
Figura 4.5	Histograma do tamanho das sentenças no <i>corpus</i>	40
Figura 4.6	Arquitetura de rede neural recorrente proposta	43
Figura 4.7	Histograma do tamanho das sentenças em palavras do <i>corpus</i>	46
Figura 5.1	F1-macro das execuções da validação cruzada	51
Figura 5.2	Acurácia das execuções da validação cruzada	52
Figura 5.3	Acurácia por resumo das execuções da validação cruzada	53
Figura 5.4	Gráfico da saída do <i>encoder</i>	55

LISTA DE TABELAS

Tabela 2.1	Tabela de características utilizadas por Teufel e Moens (1999) . . .	18
Tabela 3.1	Resumo das características e dos resultados principais dos trabalhos abordados nesta seção	33
Tabela 4.1	As 20 categorias de papéis retóricos mais frequentes das seções obtidas.	38
Tabela 4.2	Distribuição das sentenças no <i>corpus</i>	40
Tabela 4.3	Distribuição de transições de papéis retóricos no <i>corpus</i>	41
Tabela 4.4	As 10 mais frequentes sequências de transições no <i>corpus</i>	42
Tabela 5.1	Precisão (P), Revocação (R) e F1-score.	50

SUMÁRIO

1	Introdução	10
1.1	Motivação e justificativa	12
1.2	Objetivos e contribuições	12
1.3	Organização do texto	13
2	Fundamentação Teórica	14
2.1	Estrutura Retórica	14
2.1.1	Classificação de Papéis Retóricos	17
2.2	Redes Neurais	19
2.2.1	Neurônio Artificial	19
2.2.2	Multilayer Perceptron	20
2.2.3	Redes Neurais Recorrentes	22
2.2.4	Long short-term memory	24
2.2.5	Encoder-Decoder	26
2.2.6	Word Embeddings	27
3	Trabalhos Relacionados	29
3.1	Síntese dos Trabalhos Relacionados	32
4	Metodologia	34
4.1	Criação do corpus	34
4.1.1	Coleta	35
4.1.2	Estruturação	36
4.1.3	Análise	37
4.2	Redes neurais recorrentes para predição de estrutura retórica	42
4.2.1	Arquitetura	42
4.2.2	Dados de entrada	44
4.3	<i>Baselines</i>	46
4.3.1	Comparação com as <i>baselines</i>	48
5	Experimentos e Resultados	49
5.1	Comparação com as <i>baselines</i>	49
5.2	Análise do <i>encoder</i>	54
6	Conclusões	56
6.1	Trabalhos futuros	57

Introdução

O Processamento de Linguagens Naturais (PLN) é uma área de pesquisa que explora formas de se utilizar computadores para entender ou manipular línguas naturais visando a aplicação em diversos domínios. Um desses domínios que está sendo estudado atualmente no PLN é o da retórica. Isso se dá ao fato de ela ser a forma com que argumentos e informações são apresentadas em uma comunicação. Com isso, a criação de programas e técnicas que conseguem extrair informações sobre a retórica é de extrema importância, pois a partir dessas informações seria possível gerar de forma artificial uma comunicação com resultados mais naturais.

É possível analisar a retórica de forma local, na qual cada palavra exerce uma função dentro de uma frase, ou de forma global na qual cada frase exerce uma função na comunicação. Neste trabalho trataremos a retórica de forma global, na qual cada frase (ou uma porção de texto) assume um valor abstrato, chamado de papel retórico e que corresponde a uma função desempenhada por esta frase em um contexto geral. Textos acadêmicos são normalmente divididos em seções de introdução, desenvolvimento e conclusão, e desta forma, podemos atribuir o papel retórico de introdução para o texto que pertence a seção de introdução e assim sucessivamente para as outras seções. A Figura 1.1 ilustra as seções normalmente presentes em uma dissertação acadêmica e consequentemente os possíveis papéis retóricos presentes no texto. A padronização na organização dos textos acadêmicos ocorre pois nessa área a comunicação é de extrema importância, e com isso, as formas que melhor expõem as informações acabam sendo utilizadas. Também é possível notar esse tipo de padronização em partes específicas de textos acadêmicos, como no caso dos resumos e no conteúdo presente na seção de introdução.

Figura 1.1: Exemplo da organização de um texto acadêmico

1. Introdução
...
2. Fundamentação Teórica
...
3. Trabalhos Relacionados
...
4. Metodologia
...
5. Resultados
...
6. Conclusões
...

Fonte: Elaborada pelo autor.

A organização (sequência) de como esses papéis retóricos são apresentados em um texto é chamada de estrutura retórica ou esquemática. Dijk e Kintsch (1983) realizaram experimentos nos quais obtiveram resultados que sugerem que a organização retórica possui uma importância na compreensão de textos. Outros linguistas como Weissberg e Buker (1990) e Swales (1994) propuseram modelos de estruturação, respectivamente, para resumos científicos e introduções científicas. Esses trabalhos sugerem que de fato existe uma estrutura para uma boa retórica.

Como tais estruturas são importantes para uma comunicação eficiente, a detecção automática dessas estruturas pode beneficiar aplicações que precisam gerar ou avaliar a qualidade de uma comunicação. Teufel e Moens (1999), procurando melhorar as técnicas de sumarização textual, propuseram que sentenças de artigos científicos fossem classificadas em papéis retóricos, visando gerar sumários informativos com uma boa estrutura retórica. A partir de um corpus criado para essa finalidade, Teufel e Moens (1999) investigaram se era possível que tais papéis retóricos pudessem ser detectados automaticamente, utilizando técnicas de aprendizagem de máquina e características superficiais extraídas das sentenças.

1.1 Motivação e justificativa

A classificação de estruturas retóricas é uma tarefa difícil. A função representada por cada parte da estrutura retórica, ou seja, os papéis retóricos, é algo abstrato e que depende da compreensão de todo um conjunto de informações, o que torna extremamente difícil criar um algoritmo que identifique os papéis retóricos de um texto.

Outros trabalhos já utilizaram a característica sequencial das estruturas retóricas com algoritmos de aprendizagem como CRF (*Conditional Random Fields*), SSVM (*Structured Support Vector Machine*), MEMM (*Maximum-Entropy Markov Model*) e obtiveram resultados melhores do que em trabalhos que fazem a classificação por sentenças individuais. Nos trabalhos que fazem a classificação de estrutura retórica por sentença, é utilizado algum tipo de característica para obter informações sobre as sentenças vizinhas, mostrando a importância da sequencialidade na estrutura retórica.

Redes neurais recorrentes têm sido aplicadas com sucesso em vários problemas de Processamento de Linguagem Natural (PLN), e principalmente no quais possuem uma estrutura sequencial, como por exemplo, na classificação de sentimentos (SOCHER et al., 2013).

Assim, espera-se que o uso de redes neurais recorrentes capazes de prever a estrutura retórica como um todo, possam obter melhores resultados do que classificadores que não utilizam informação da sequência para realizar a predição e também de outros algoritmos que utilizam predição estruturada.

1.2 Objetivos e contribuições

O objetivo deste trabalho foi investigar o uso de redes neurais recorrentes para a classificação de estrutura retórica. Para isso, foi necessário primeiramente construir um *corpus* com uma grande quantidade de resumos de artigos científicos com suas estruturas retóricas anotadas para que fosse possível o treino da arquitetura de rede proposta. Em seguida, tendo como base as características utilizadas no trabalho de Hirohata et al. (2008), foram implementadas duas *baselines* utilizando os algoritmos de aprendizagem CRF e SVM.

As contribuições deste trabalho são uma arquitetura de rede neural recorrente que pode ser utilizada para a classificação de estrutura retórica em resumos de artigos científicos, bem como um novo *corpus* e *baselines* que podem ser utilizados na avaliação de classificadores de estrutura retórica. Adicionalmente, os resultados obtidos por esse trabalho podem melhorar o desempenho de ferramentas de auxílio a escrita científica.

1.3 Organização do texto

Esta dissertação está dividida em seis capítulos, sendo este o primeiro deles. No Capítulo 2 são apresentadas a definição de estrutura retórica e uma fundamentação teórica de redes neurais, redes neurais recorrentes e *embeddings*. O Capítulo 3 apresenta os métodos e resultados de trabalhos que já realizaram classificação de estruturas retóricas. No Capítulo 4 é apresentada a metodologia para a criação de um corpus, a arquitetura de rede desenvolvida e também as características extraídas para as *baselines*. Os experimentos realizados e os resultados obtidos são apresentados no Capítulo 5, na qual são comparados os resultados obtidos pela rede proposta com os obtidos pelas *baselines*. Nesse Capítulo também é feita a análise de uma parte da rede para avaliar o seu desempenho. Por fim, no Capítulo 6, as considerações finais são feitas e são apresentadas propostas para trabalhos futuros.

Fundamentação Teórica

Neste capítulo é apresentada a fundamentação teórica necessária para o desenvolvimento e entendimento do trabalho.

2.1 Estrutura Retórica

O estudo de como a língua é utilizada é chamado de análise de discurso. Um dos diversos tópicos dentro da análise de discurso é o estudo da retórica. A retórica trata de formas de persuadir, informar, ou motivar um interlocutor. Isso pode ser feito de diversas maneiras, como por exemplo: por meio do uso de argumentações, ou apresentando definições. Uma característica importante da retórica é sua estrutura, ou seja, a organização de como tais argumentos são apresentados no texto.

Em certos gêneros textuais é comum o uso de estruturas retóricas que são recorrentes, ou seja, estruturas que são utilizadas como um padrão para a escrita. Um dos gêneros estudados por linguistas é o científico, pois sabemos que a linguagem é de extrema importância para a ciência, e de modo geral, a partir dela conseguimos divulgar e adquirir novos conhecimentos. Mas para isso é necessário que a comunicação seja feita de forma adequada. Linguistas como Weissberg e Buker (1990) e Swales (1994) investigaram a presença de padrões nas estruturas de argumentação para criarem um modelo de escrita, de modo a ajudar na escrita de textos científicos. A Figura 1.1 mostra um exemplo de uma estrutura que é comumente utilizada em dissertações acadêmicas. Mais especificamente, Weissberg e Buker (1990) estudaram e propuseram um modelo de estruturação para resumos científicos, ilustrado na Figura 2.1. Swales (1994) propôs um modelo para

estruturção de introduções científicas, no qual divide cada componente em estratégias mais específicas. Cada componente é chamado de movimento, que por sua vez possui estratégias chamadas de passos, conforme ilustrado na Figura 2.2.

Figura 2.1: Modelo de estrutura argumentativa proposta por Weissberg e Buker (1990)

1. Contextualização
2. Propósito e escopo
3. Metodologia
4. Resultados importantes
5. Conclusão e recomendações

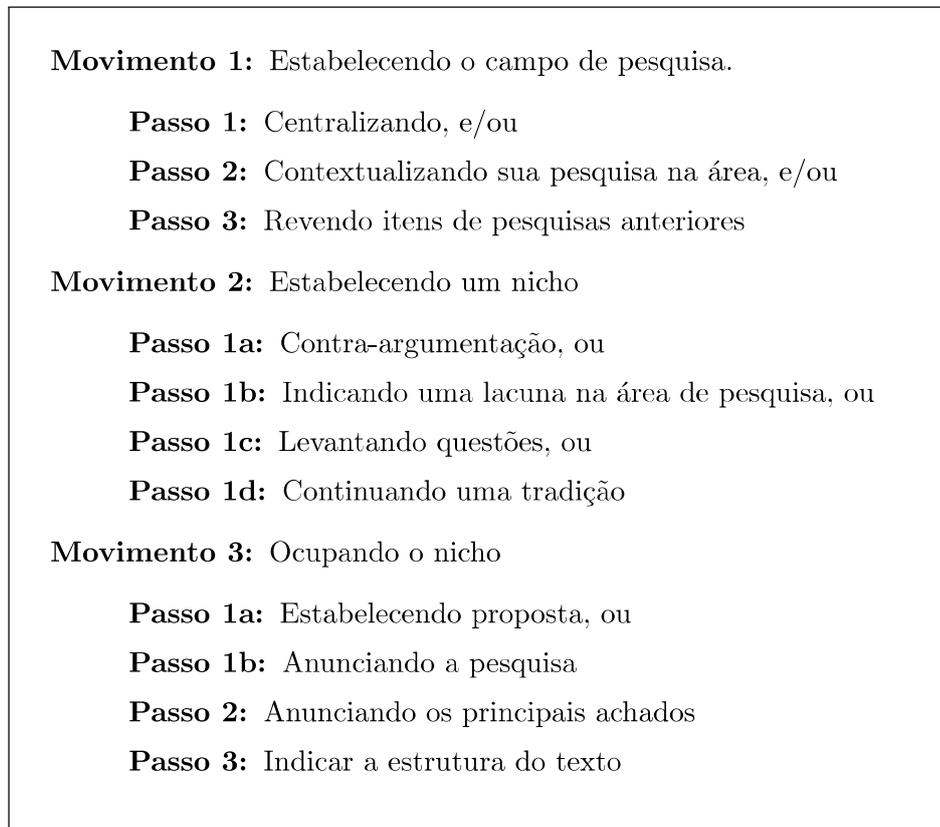
Fonte: Adaptado de Weissberg e Buker (1990).

Apesar de serem modelos de estruturação para diferentes partes de um texto científico, é possível observar que existe uma sobreposição entre as componentes propostas. A componente de contextualização de Weissberg e Buker (1990) pode ser encaixada no primeiro movimento de Swales (1994), propósito e metodologia no segundo movimento, e por fim resultados importantes e conclusão e recomendação no terceiro movimento. Em ambos os casos, a principal motivação dos estudos foi para facilitar a escrita de textos científicos, uma tarefa que é considerada difícil principalmente para escritores iniciantes. Isso evidencia que, embora os modelos sejam distintos, uma boa argumentação é composta por uma certa estrutura retórica que fica aparente nos textos de um gênero.

Segundo Dijk e Kintsch (1983), as estruturas retóricas são utilizadas para aumentar a eficiência da comunicação e são independentes da língua utilizada. Para testar essa afirmação, Dijk e Kintsch (1983) realizaram um experimento no qual textos com uma estrutura retórica considerada boa era apresentada para um grupo de pessoas. Esses textos tiveram sua estrutura alterada e foram apresentados para outro grupo de pessoas. Após a leitura dos textos, eram realizadas perguntas sobre informações contidas neles. Foi notado que o grupo para o qual foi apresentado o texto sem a alteração da estrutura retórica obteve um percentual melhor de acerto do que o outro grupo, sugerindo a importância da estrutura retórica para a compreensão de textos.

Chamamos de papel retórico a abstração da função desempenhada na argumentação de um certo segmento de texto, por exemplo: quando um segmento do texto estabelece conceitos necessários para o entendimento do restante da argumentação, atribuímos a

Figura 2.2: Modelo de estrutura argumentativa proposta por Swales (1994)



Fonte: Adaptado de Swales (1994).

tal segmento o papel retórico de contextualização. A Figura 2.3 exemplifica um resumo no qual para cada sentença foi atribuído um papel retórico que corresponde a função desempenhada na argumentação pela sentença.

A atribuição de um papel retórico além de ser abstrata não é necessariamente única (DAYRELL et al., 2012). Por exemplo, na quarta sentença da Figura 2.3 também poderia ser designado um papel retórico de "Propósito", que exerce uma função de mostrar o propósito do trabalho em seu contexto.

Figura 2.3: Exemplo de um resumo com suas sentenças atribuídas a papéis retóricos.

1. Aplicações científicas atuais têm produzido volumes de dados cada vez maiores. (Contexto)
2. O processamento, a manipulação e a análise desses dados requerem infraestruturas computacionais de larga escala tais como aglomerados e grades de computadores. (Contexto)
3. No entanto, uma das principais lacunas dessas pesquisas está na falta de emprego de conhecimento sobre aplicações com objetivo de realizar essa otimização. (Contexto)
4. Essa lacuna motivou esta tese que visa empregar comportamento histórico e preditivo de aplicações a fim de otimizar suas operações de leitura e escrita sobre dados distribuídos. (Contexto)
5. Os estudos foram iniciados empregando-se informações previamente monitoradas de aplicações a fim de tomar decisões relativas à replicação, migração e manutenção de consistência. (Metodologia)
6. Observou-se, por meio de uma nova heurística, que um conjunto histórico de eventos auxilia a estimar o comportamento futuro de uma aplicação e otimizar seus acessos. (Resultado)
7. Esse requisito pode limitar aplicações reais que apresentam mudanças comportamentais ou que necessitam de longos períodos de execução para completar seu processamento. (Conclusão)

Fonte: Adaptado de Ishii (2010).

2.1.1 Classificação de Papéis Retóricos

A classificação de papéis retóricos trata de encontrar a melhor classe que represente um certo segmento de um texto. Os papéis retóricos não são exclusividade de um único gênero textual. Existem trabalhos que abordam a classificação em *essays* (BURSTEIN; MARCU; KNIGHT, 2003), notas clínicas (LI; GORMAN; ELHADAD, 2010), textos jurídicos (GROVER; HACHEY; KORYCINSKI, 2003), e textos científicos (TEUFEL; MOENS, 1999).

Como visto na seção anterior, a atribuição de um papel retórico é algo subjetivo e abstrato, o que torna a classificação automática um desafio. Para superar esse desafio é necessário encontrar as melhores características que possam representar as diversas classes de papéis retóricos. Outro desafio é encontrar quais os papéis retóricos que melhor definem

o problema a ser classificado, e para isto, é necessário analisar o conteúdo do problema que será classificado. Por exemplo, Teufel e Moens (1999) utilizaram como base os modelos propostos por Swales (1994) de modo que foi possível obter uma base sólida de quais papéis retóricos poderiam ser utilizados.

Encontrar características que representem um problema também não é uma tarefa trivial. Não existe uma fórmula para seleção de quais características devem ser utilizadas e, muitas vezes elas são propostas por especialistas da área da qual fazem parte, no caso de papel retórico, linguistas. Essas características por sua vez são testadas e avaliadas empiricamente. Um exemplo é apresentado na Tabela 2.1 na qual é possível visualizar as características que foram utilizadas por Teufel e Moens (1999) para a classificação de papéis retóricos em artigos científicos.

Tabela 2.1: Tabela de características utilizadas por Teufel e Moens (1999)

Tipo	Nome	Descrição
Estrutura explícita	Estrutura-1	Tipo do cabeçalho da seção
	Estrutura-2	Posição relativa da sentença dentro do parágrafo
	Estrutura-3	Posição relativa da sentença dentro da seção
Posição relativa	Localização	O artigo é segmentado em 10 partes de tamanhos iguais.
Citações	Citação-1	A sentença contém uma citação ou um nome de um autor que está na lista de referências?
	Citação-2	A sentença contém uma citação própria?
Sintática	Sintática-1	Tempo (associado ao primeiro verbo da sentença)
	Sintática-2	Modais Auxiliares
	Sintática-3	Voz
	Sintática-4	Negação
Semântica	Semântica-1	Tipo da ação do primeiro verbo da sentença
	Semântica-2	Tipo do agente
	Semântica-3	Tipo da expressão formulaica que ocorre na sentença
Conteúdo	Conteúdo-1	A sentença contém uma palavra chave determinada pela medida <i>tf-idf</i> ?
	Conteúdo-2	A sentença contém alguma palavra que também aparece no título ou cabeçalho?

Fonte: Adaptado de Teufel e Moens (1999).

As principais abordagens à classificação de estrutura retórica utilizam aprendizagem de máquina supervisionada, portanto, necessitam de um conjunto de dados no qual a informação sobre os papéis retóricos está disponível. Como não existem muitos conjuntos publicamente disponíveis, grande parte dos trabalhos acabam criando seus próprios conjuntos (TEUFEL; MOENS, 1999; ANTHONY; LASHKIA, 2003; HIROHATA et al., 2008).

2.2 Redes Neurais

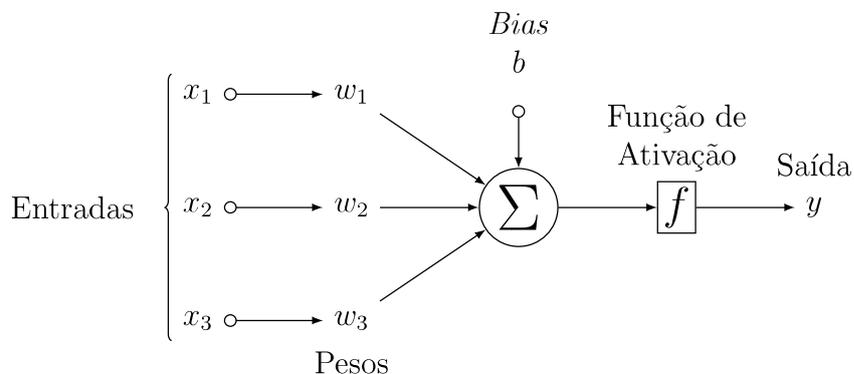
Redes neurais artificiais ou simplesmente redes neurais são modelos inspirados no funcionamento biológico dos neurônios naturais. O primeiro modelo artificial de um neurônio foi proposto por McCulloch e Pitts (1943) o qual possui um funcionamento simples. Como se tratava de apenas um neurônio, era difícil de utilizá-lo em tarefas mais complexas. Portanto, a solução mais direta foi encadear vários neurônios de modo a conseguir resolver problemas mais complexos, porém isso introduziu uma dificuldade para realizar o treinamento de tais redes. Para resolver esse problema, foi proposto o algoritmo de *backpropagation* (RUSSELL; NORVIG, 2016). Portanto, redes neurais são uma série de funções encadeadas, que são chamadas de camadas, e são aplicadas a uma entrada de dados.

Nessa seção apresentaremos as noções básicas sobre redes neurais, desde o primeiro modelo de um neurônio artificial, até formas mais complexas de redes neurais, chamadas de redes neurais recorrentes.

2.2.1 Neurônio Artificial

Como dito anteriormente, o primeiro modelo artificial de um neurônio foi proposto por McCulloch e Pitts (1943). Como ilustrado na Figura 2.4, o neurônio artificial funciona multiplicando pesos (w_1, w_2, w_3) pelos valores da entrada (x_1, x_2, x_3), em seguida os resultados dessas operações, incluindo um valor de *bias*, são somados. Por fim, o valor resultante do somatório é aplicado a uma função de ativação e seu resultado é o valor da saída desse neurônio.

Figura 2.4: Neurônio Artificial



Fonte: Elaborada pelo autor.

2.2.2 Multilayer Perceptron

É possível fazer com que neurônios possam ser ligados uns aos outros de forma que um neurônio tenha sua saída conectada na entrada de outro, formando então uma arquitetura (rede) em múltiplas camadas. Uma das arquiteturas mais simples, são as redes neurais de múltiplas camadas, que comumente são chamadas de *multilayer perceptron* (MLP).

A Figura 2.5 apresenta uma arquitetura de uma MLP com duas camadas ocultas. Como é possível observar, chamamos de "camada de entrada" a primeira camada de uma MLP e a última camada de "camada de saída". As camadas intermediárias entre a camada de entrada e a camada de saída são chamadas de "camadas ocultas". As redes neurais também possuem uma unidade chamada de *bias*, ela se comporta como um neurônio na qual sua função de ativação sempre retorna o valor 1. É possível observar que o *bias* está presente na camada de entrada, nas camadas ocultas, porém não é presente na camada de saída, já que uma saída com o valor sempre 1 não é útil.

Considerando o funcionamento entre duas camadas, tendo um vetor $X \in \mathbb{R}^n$ com valores das ativações da primeira camada (no caso de uma camada de entrada a ativação é o próprio valor de entrada), em que n é a quantidade de valores de entrada, a matriz $\theta \in \mathbb{R}^{m \times n}$ com os pesos (conexões) entre os neurônios das camadas, sendo m a quantidade de neurônios na segunda camada, e por fim um vetor $Y \in \mathbb{R}^m$ como o resultado da computação entre as duas camadas, considerando uma função de ativação f na segunda camada. Temos que $Y = f(\theta^\top X)$, ou seja, o resultado da conexão entre duas camadas é um vetor com os resultados da aplicação da função f no resultado entre o produto da matriz dos pesos das conexões e as entradas para cada neurônio.

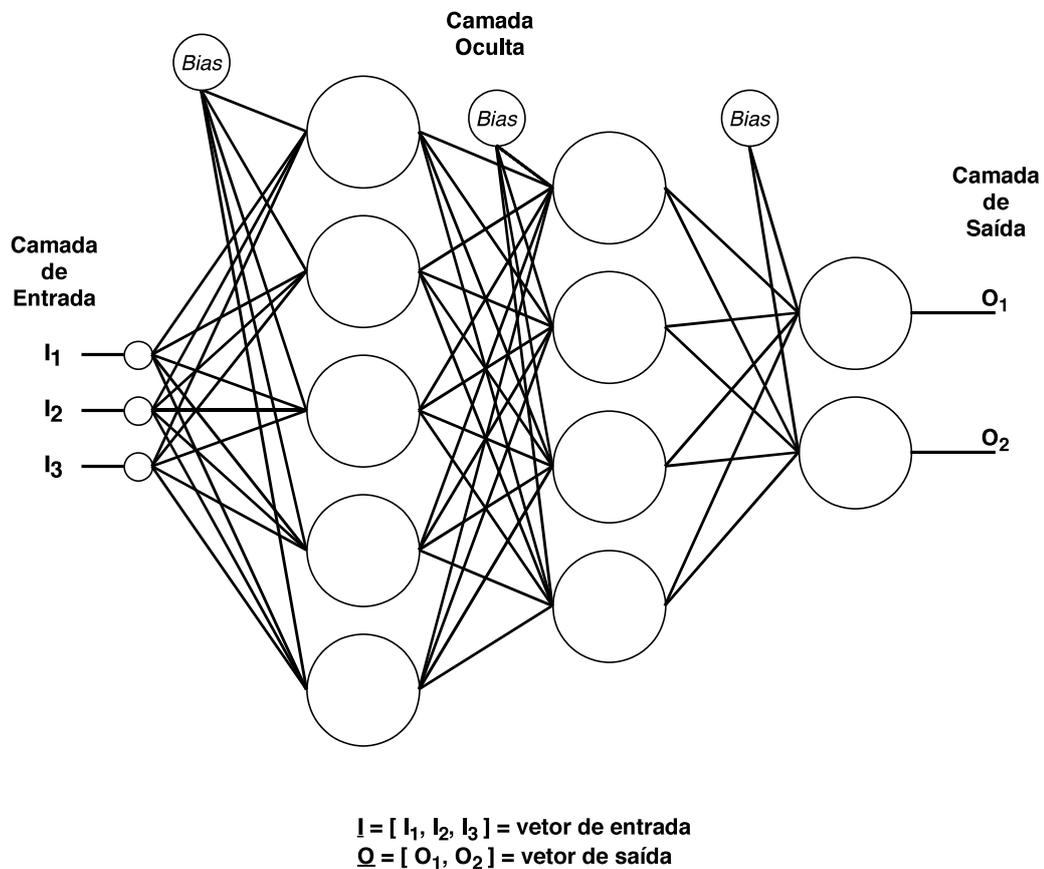
A função de ativação é comumente uma função não-linear, e como dito anteriormente, ela é aplicada sobre o resultado do produto entre os pesos e as entradas. Existem diversas funções de ativação que são utilizadas em redes neurais, alguns exemplos comuns são as funções: sigmoide (Equação 2.1), tangente hiperbólica (Equação 2.2) e *Rectified Linear Unit* (ReLU) (Equação 2.3).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.2)$$

$$\text{relu}(x) = \max(0, x) \quad (2.3)$$

Figura 2.5: Rede neural de múltiplas camadas



Fonte: Adaptado de Gardner e Dorling (1998)

Krizhevsky, Sutskever e Hinton (2012) mostraram que ReLUs podem acelerar o treinamento de redes neurais, fazendo com que seja necessário menos tempo de treinamento para as redes convergirem. A sua implementação é simples e sua execução requer pouco processamento computacional.

É conhecida na literatura como *Forward Propagation* a etapa na qual é calculada todos os valores de ativações de uma rede, ou seja, da camada de entrada até a sua camada de saída. O número de neurônios na última camada da rede, e por consequência a quantidade de ativações, define a quantidade de classes que a rede pode classificar.

Na classificação de problemas binários a última camada da rede apresenta somente um neurônio e normalmente utiliza uma função de ativação sigmoide. De modo que é possível considerar o resultado da ativação como a probabilidade de pertencer a uma classe, e o inverso dessa probabilidade como a probabilidade de pertencer a outra classe.

Em problemas de múltiplas classes a última camada da rede apresenta N neurônios, em que N é a quantidade de classes a serem classificadas. Nesse tipo de problema é utilizada a função de ativação *softmax* (Equação 2.4), que retorna as probabilidades das classes ocorrerem dada uma entrada.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{n=1}^N e^{z_n}} \quad (2.4)$$

Para realizar o treinamento de uma rede neural de múltiplas camadas, ou seja, encontrar os pesos ideais de uma rede, é necessário encontrar um valor de erro entre o resultado da camada de saída da rede e o valor real esperado. A função que mensura esse erro é chamada de função custo. Como normalmente as redes neurais retornam a probabilidade de uma classe ocorrer, uma função custo que é normalmente utilizada em redes neurais é a de entropia cruzada, descrita na Equação 2.5.

$$\mathcal{L}(X, Y) = - \sum_{n=1}^N x_n \log y_n \quad (2.5)$$

O *backpropagation* (RUSSELL; NORVIG, 2016) é utilizado para minimizar a função de custo em redes neurais. Para isso, ele utiliza um método chamado de gradiente descendente (CAUCHY, 1847) para atualizar os pesos da rede, no qual o algoritmo segue a direção inversa do *Forward Propagation* para computar os gradientes de pesos em relação a função custo \mathcal{L} . A Equação 2.6 mostra a regra para a atualização dos pesos da rede em um instante n .

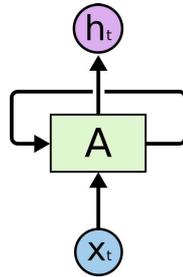
$$\theta_{n+1} = \theta_n - \gamma \nabla \mathcal{L}(\theta_n) \quad (2.6)$$

Na qual γ é a taxa de aprendizagem, ou seja, é um valor que tenta controlar a convergência da rede delimitando passos (variações) a serem incluídos pelo algoritmo de *backpropagation*.

2.2.3 Redes Neurais Recorrentes

Assim como as MLPs, as redes neurais recorrentes (do inglês, *recurrent neural networks* - RNN) são um tipo de arquitetura de redes em camadas, porém com a adição da opção de existirem conexões recorrentes entre as camadas, conforme ilustrado na Figura 2.6.

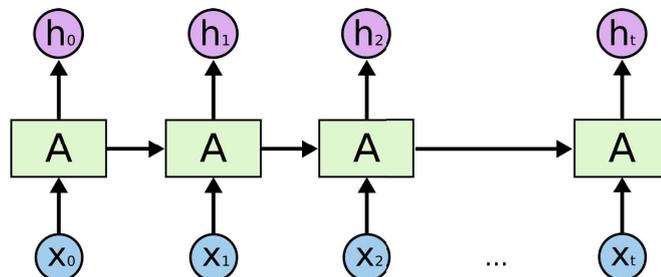
A adição da recorrência cria um efeito temporal na rede, permitindo com que certas conexões da rede (que podem ser chamados de parâmetros) sejam utilizadas em instantes de tempos diferentes. A ideia de compartilhar os parâmetros de uma rede neural

Figura 2.6: Rede neural recorrente

Fonte: Olah (2015)

entre diferentes instantes de tempos surgiu com Lang, Waibel e Hinton (1990). Isso permite as RNNs capturarem relações temporais de uma sequência de dados, pois o compartilhamento de parâmetros permite o uso de informações passadas para ser gerada uma previsão.

A Figura 2.6 ilustra uma RNN em um instante de tempo t e uma entrada x_t , produzindo uma saída h_t . Na qual a camada A é recorrente, portanto ela tem como entrada x_t e h_{t-1} . Um exemplo de uma entrada temporal (ou sequencial) é uma sentença, na qual cada palavra corresponde a uma entrada em um instante de tempo. Essa característica de utilizar sequências de dados como entrada para realizar previsões é um dos principais motivos das RNNs serem utilizadas em diversas tarefas de PLN, como por exemplo, a detecção de paráfrase (SOCHER et al., 2011), análise de sentimento (SOCHER et al., 2013), extração de opinião (IRSOY; CARDIE, 2014), previsão de interesse (GAO et al., 2014) e outros.

Figura 2.7: Rede neural recorrente com passos expandidos

Fonte: Olah (2015)

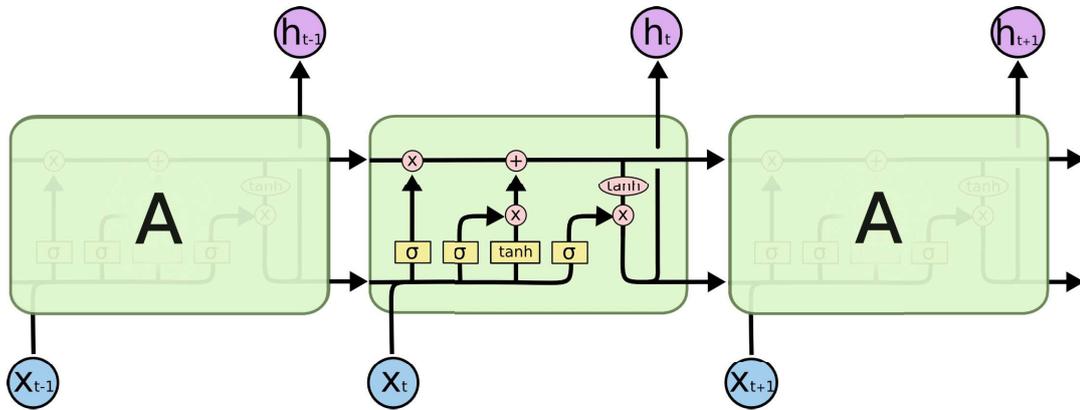
Outra forma de visualizar redes neurais recorrentes é expandindo as suas recorrências. A Figura 2.7 mostra a expansão das recorrências da rede ilustrada na Figura 2.6. Nessa forma é possível visualizar o fluxo dos dados de uma RNN. Nesse trabalhos, chamamos de unidade oculta a quantidade de neurônios que pertencem a recorrência. Por exemplo, caso consideramos que na Figura 2.6 a RNN possui uma unidade de tamanho 10, sua saída será um vetor de tamanho 10 para cada instante de tempo.

Para realizar o treinamento de RNNs, ou seja, minimizar a função custo, os pesos da rede são ajustados utilizando uma adaptação do algoritmo *backpropagation* chamado de *backpropagation through time* (BTT) (WERBOS, 1990). Resumidamente, o BTT ajusta os pesos da rede ao longo de todo intervalo de tempo somando os gradientes da função custo para cada instante t . Um dos problemas encontrados quando utilizado o BTT em RNNs é a explosão de gradiente, em que o gradiente atinge um valor muito alto e o algoritmo de descida de gradiente não consegue ajustar os valores, e o desvanecimento de gradientes que torna os valores do gradiente muito pequenos ao decorrer dos passos de tempo de modo que os gradientes acabam perdendo relevância, dificultando a convergência da rede (GOODFELLOW; BENGIO; COURVILLE, 2016). Esse tipo de problema ocorre principalmente em recorrências muito longas. Para resolver esse problema, Hochreiter e Schmidhuber (1997) propuseram um tipo de neurônio chamado de *Long-Short Term Memory* (LSTM) que utiliza células de memória especiais como unidade de ativação e controle.

2.2.4 Long short-term memory

Como explicado na seção anterior, a *Long-Short Term Memory* (LSTM) (HOCHREITER; SCHMIDHUBER, 1997) é um tipo especial de neurônio recorrente que permite solucionar os problemas de explosão e desvanecimento de gradientes em RNNs. Nesse trabalho, o termo LSTM é utilizado para arquiteturas de redes neurais que utilizam vários ou apenas um neurônio LSTM.

Figura 2.8: Rede neural recorrente *LSTM*



Fonte: Olah (2015)

A principal característica da LSTM é a utilização informações que foram armazenadas internamente em sua célula de estado por meio de *gates*. Os *gates* são responsáveis por gerenciar a memória e a entrada e saída de dados. São utilizadas funções sigmoide (σ) em *gates* que regulam as informações a serem passadas adiante ou a serem armazenadas na memória interna. Os *gates* que utilizam funções *tanh*, são análogos a funções de recompensa.

A Figura 2.8 ilustra um neurônio LSTM e suas operações internas e com as recorrências expandidas. A primeira operação realizada dentro da LSTM é quais informações serão removidas da memória (também chamada de estado da célula), essa etapa é realizada pelo *gate* de esquecimento, descrito na Equação 2.7.

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (2.7)$$

Para selecionar quais informações serão armazenadas no estado da célula, são realizadas duas operações que utilizam a entrada atual x_t e a saída da recorrência anterior h_{t-1} , essas operações são descritas na Equação 2.8 e Equação 2.9.

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (2.8)$$

$$\tilde{\mathbf{C}}_t = \sigma(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C) \quad (2.9)$$

Por fim, a saída do estado da célula em um determinado instante t é atualizada conforme descrito na Equação 2.10.

$$\mathbf{C}_t = \mathbf{f}_t * \mathbf{C}_{t-1} + \mathbf{i}_t * \tilde{\mathbf{C}}_t \quad (2.10)$$

O *gate* de saída é responsável por gerenciar a saída da LSTM. Para isso, é aplicado a função sigmoide na entrada h_{t-1} e x_t como descrito na Equação 2.11.

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t]) + \mathbf{b}_o \quad (2.11)$$

E por fim, para gerar a saída da LSTM em um instante de tempo t , é realizada uma multiplicação entre o *gate* de saída e o estado da célula, como descrito na Equação 2.12.

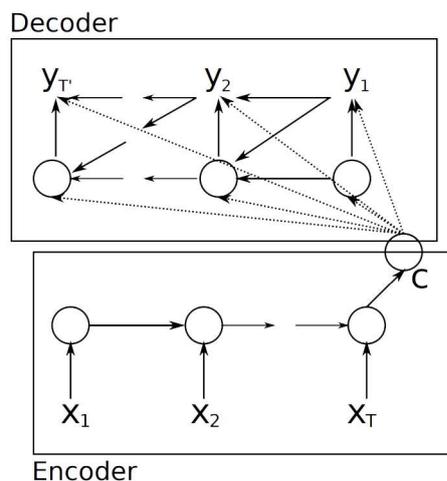
$$\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{C}_t) \quad (2.12)$$

Onde W são os pesos das conexões entre o *gate* e as entradas h_{t-1} e x_t , e o b é o *bias* para esse *gate*.

2.2.5 Encoder-Decoder

A arquitetura de um *encoder-decoder* foi proposta por Cho et al. (2014). Trata-se de utilizar RNNs para transcodificar uma sequência de tamanho variável em um vetor de tamanho fixo (*encoder*), e decodificar um vetor de tamanho fixo em uma sequência de tamanho variável (*decoder*). A Figura 2.9 mostra a arquitetura da rede de um *encoder-decoder*.

Figura 2.9: Arquitetura de um *encoder-decoder*



Fonte: Cho et al. (2014)

A principal característica dessa arquitetura é que força a rede a aprender uma representação intermediária da sequência de entrada. É possível implementar essa arquitetura de modo que o *encoder* seja uma RNN na qual é utilizada somente a última saída da recorrência, conforme ilustrado na Figura 2.9. Na qual C representa essa última saída.

O *decoder* também pode ser implementado como uma RNN que recebe como entrada C todos os intervalos de tempos, ou seja, repetindo C t vezes como entrada. Sua saída é correspondente a saída da RNN.

Tanto para o *encoder* quanto para o *decoder* também é possível utilizar LSTM ou outros tipos de camadas recorrentes no lugar da RNN sem nenhuma alteração necessária.

O *encoder* utiliza a recorrência para armazenar todas as informações da entrada em apenas uma saída, e portanto, criando uma representação intermediária. Já o *decoder* apesar de repetir a mesma entrada, apresenta uma recorrência na qual sua própria saída é utilizada como entrada e isso possibilita o aprendizado de sequências de saídas.

Cho et al. (2014) utilizaram essa arquitetura para o problema de tradução, no qual tinha como entrada uma sentença em um idioma, e como saída uma sentença em outro idioma. Isso fez com que a rede aprendesse uma representação intermediária que possibilitava uma correspondência entre os dois idiomas.

2.2.6 Word Embeddings

Boa parte dos resultados obtidos por redes neurais no processamento de linguagem natural ocorreram devido ao uso de representações distribuídas de palavras ou *words embeddings* (LEVY; GOLDBERG; RAMAT-GAN, 2014).

A representação distribuída é feita transformando palavras em um vetor de valores, por exemplo, por meio de uma função W que mapeia palavras de uma língua em um espaço vetorial de dimensão n , conforme mostrado na definição 2.13. Tendo em vista que a dimensão dos vetores corresponde a quantidade de características a serem extraídas, normalmente são utilizados grandes valores de dimensões.

$$W : \text{palavras} \rightarrow \mathbb{R}^n \quad (2.13)$$

No início, a função W é criada de forma que todas as palavras de um vocabulário possuam um vetor com um valor único (Figura 2.10). Por exemplo, considerando a palavra "olá" como a primeira palavra de um vocabulário que está sendo gerado, e o *embedding* e a matriz W exibida na Figura 2.10, e como cada linha da matriz é um vetor que representa

uma palavra do vocabulário, nesse exemplo a palavra "olá" terá o vetor $[0, 20, 30, 1...0, 5]$, pois o índice da linha na matriz corresponde ao índice da palavra no vocabulário.

A próxima etapa é fazer com que esses vetores apresentem características semelhantes. Essa etapa foi realizada pelos trabalhos de Bengio et al. (2003) e Mikolov et al. (2011) utilizando redes neurais para ajustar os valores de forma que os *words embeddings* sejam construídos para captar as similaridades entre palavras de um vocabulário, fazendo com que palavras que aparecem em um contexto similar possuam vetores com características similares entre os vocabulários. Isso é feito de forma não supervisionada e tem como objetivo somente extrair a similaridade.

Figura 2.10: Exemplo de uma representação distribuída.

$$W = \begin{bmatrix} \boxed{0,2} & \boxed{0,3} & \boxed{0,1} & \dots & \boxed{0,5} \\ 0,0 & 0,6 & 0,7 & \dots & 0,8 \\ 0,9 & 0,2 & 0,3 & \dots & 0,4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -0,1 & 0,6 & 0,7 & \dots & 0,6 \end{bmatrix}$$

Fonte: Elaborado pelo autor.

Cabe destacar que é possível generalizar essa etapa como uma camada de uma rede neural de forma que os vetores de características das palavras sejam aprendidos especificamente para o problema para o qual a rede está sendo treinada. Isso é possível criando uma "camada *embedding*" que é responsável por mapear uma sequência de índices às suas respectivas representações distribuídas (vetores). Por fim, o valor desses vetores são utilizados como os pesos das conexões entre as seguintes camadas, permitindo com que eles sejam otimizados de acordo a função custo da rede neural.

Trabalhos Relacionados

Nesse capítulo são apresentados os principais trabalhos encontrados na literatura sobre a classificação de estruturas retóricas. O trabalho de Teufel e Moens (1999) recebe um foco maior, pois trata-se do primeiro trabalho que utiliza a classificação de estruturas retóricas, o que o torna base para todos os outros trabalhos encontrados na literatura. Os trabalhos serão apresentados em ordem cronológica. A Seção 3.1 traz um resumo da revisão da literatura, apresentando as principais características observadas nos trabalhos abordados.

Teufel e Moens (1999) elaboraram um *corpus* de artigos científicos para ser utilizado em sumarização. Em vez de serem anotadas somente as sentenças que poderiam aparecer em um sumário, foram anotados os papéis retóricos de cada sentença. O *corpus* foi criado a partir de 80 artigos da *Computation and Language E-Print Archive*, contendo aproximadamente 12.000 sentenças. Baseado na estrutura de Swales (1990), foram considerados os seguintes papéis retóricos:

- **Contexto:** Sentenças que descrevem algum conhecimento prévio.
- **Outro:** Sentenças que descrevem aspectos de algum outro trabalho, de modo neutro.
- **Alvo:** Sentenças que descrevem o principal objetivo da pesquisa realizada no artigo.
- **Textual:** Sentenças que descrevem a estrutura do texto.
- **Contraste:** Sentenças que contrastam o próprio trabalho em relação a trabalhos realizados por outros; sentenças que apontam limitações em outros trabalhos;

sentenças que afirmam que a pesquisa realizada nunca tinha sido realizada até o presente momento; comparações diretas com outros trabalhos.

- **Base:** Afirmações de que o trabalho realizado usa outros trabalhos como base.

Após a anotação, foram realizados experimentos para verificar se era possível a classificação automática dos papéis retóricos. Para isso, foram extraídas características do texto representando a estruturação em seções, posição relativa das sentenças, citação, características sintáticas e semânticas do texto. Utilizando um classificador *Naive Bayes*, a acurácia obtida foi de 71,2% e, quando comparado com anotadores humanos, a medida Kappa foi de 0,37. Cabe destacar que a medida Kappa pode variar de -1 a 1, sendo que quanto mais próximo o valor está de 1, maior é a concordância entre os anotadores. Esses resultados também foram comparados com duas *baselines*. Na primeira, todas as sentenças assumiam como classe o papel retórico mais frequente no *corpus*, com isto, foi obtida uma acurácia de 69% e Kappa de -0,12. Na segunda *baseline* as sentenças assumiam como classe papéis retóricos escolhidos de forma aleatória, e levando em conta a distribuição de papéis retóricos do *corpus*, com isso, o Kappa era 0.

Seguindo a ideia de Teufel e Moens (1999), Feltrim et al. (2004) propuseram a detecção de estrutura retórica em resumos de textos científico em português. Para isto, foram anotados 52 resumos de teses em Ciência da Computação, com um total de 366 sentenças. Os papéis retóricos considerados são similares aos utilizados no trabalho de Teufel e Moens (1999). Eles são: contexto, lacuna, propósito, metodologia, resultados, conclusão e estrutura. As características consideradas foram derivadas das utilizadas por Teufel e Moens (1999). Foi utilizado o classificador *Naive Bayes* e obtidos os resultados de uma acurácia de 74% e Kappa de 0,65, que é um valor mais alto de concordância quando comparado com Teufel e Moens (1999). Feltrim et al. (2004) dizem que este resultado maior pode ser devido ao fato do *corpus* utilizado ser somente de resumos, enquanto Teufel e Moens (1999) usaram artigos completos.

Para melhorar a leitura e escrita de resumos de artigos, Anthony e Lashkia (2003) propuseram o classificador MOVER. Para o auxílio da leitura, é mostrado ao leitor qual o papel retórico da sentença, facilitando o entendimento de algumas sentenças. O mesmo vale para a escrita. Apresentando o papel retórico das sentenças, o usuário pode identificar quais componentes estão faltando. Anthony e Lashkia (2003) criaram um *corpus* a partir de 100 resumos de artigos científicos extraídos do *IEEE Transactions on Parallel and Distributed Systems*, na anotação foram considerados seis papéis retóricos. Para representar as sentenças em características foi utilizada uma variação do *bag-of-words*

chamado de *bag-of-clusters*. Por fim, foi avaliado o desempenho do classificador *Naive Bayes*, que obteve acurácia de 68%.

Diferente dos trabalhos anteriores, Hirohata et al. (2008) propôs o uso do seccionamento de resumos científicos disponíveis em alguns artigos científicos para não depender de uma anotação manual. Os autores propõem a categorização de sentenças de artigos científicos em quatro categorias de papéis retóricos: objetivos, metodologia, resultados e conclusão. Como características a serem extraídas foram consideradas *n-gram*, posição relativa da sentença e características das sentenças vizinhas. Foi criado um *corpus* com 51.000 resumos extraídos da base de artigos científicos Medline com suas sentenças rotuladas a partir do seccionamento feito pelos próprios autores. Como experimento foi utilizado o modelo *Conditional Random Fields* (CRF) para a classificação das sentenças em suas possíveis seções, obtendo acurácia de 95,5% por sentença.

Yepes, Mork e Aronson (2013) utilizaram das informações disponíveis no PubMed, base na qual a partir de 16 de agosto de 2010 começou a se formatar os resumos de artigos científicos em seções. As seções são semelhantes aos papéis retóricos de Hirohata et al. (2008), porém com a adição da categoria de contexto. A partir de 100 mil resumos extraídos do PubMed com as sentenças rotuladas conforme suas seções, foram avaliados diversos modelos de classificadores. O classificador que obteve melhor resultado foi o CRF, com Macro F1-score de 79%.

Em contrapartida aos métodos de aprendizagem supervisionada utilizados nos trabalhos anteriores, Guo, Korhonen e Poibeau (2011) investigaram o uso de métodos de aprendizagem fracamente supervisionados. Esse tipo de método utiliza os dados que estão disponíveis com anotações para realizar um treinamento supervisionado, porém em seguida classifica dados que não estão anotados e os utilizam para melhorar o treino do modelo. Uma das motivações foi a dificuldade de conseguir grandes quantidades de dados anotados, que normalmente é necessária para uma aprendizagem supervisionada. Para a experimentação foi utilizado um *corpus* com 1.000 resumos da área de biomedicina com um total de 7.895 sentenças, anotadas em sete classes de papéis retóricos: contexto, objetivo, metodologia, resultados, conclusão, trabalhos relacionados e trabalhos futuros. Foram extraídas características mais profundas, por meio de ferramentas de anotação automática de classes gramaticais. De quatro algoritmos de aprendizagem fracamente supervisionados, o *Active SVM with self-training* (ASSVM) obteve os melhores resultados. Com apenas 10% dos dados de treinamento, ele obteve 81% de F1-score e algoritmos supervisionados como SVM e CRF obtiveram resultados de 77% e 70%, respectivamente.

Uma das dificuldades encontradas por Feltrim et al. (2004) na anotação de um *corpus* para a detecção de estrutura retórica foi que algumas sentenças podiam se encaixar em

duas ou mais classes de papéis retóricos. Nesse sentido, Dayrell et al. (2012) anotou dois *corpora* com multirrótulos. O primeiro *corpus* é composto por 645 resumos de textos científicos das áreas de física e engenharia. O segundo *corpus* é composto por 690 resumos da área da saúde. Os papéis retóricos considerados foram: contexto, lacuna, propósito, metodologia, resultado e conclusão. Foi verificado que grande parte das sentenças anotadas acabaram possuindo um único papel retórico. Para avaliar o desempenho de um classificador multirrótulo foram realizados experimentos com os dois *corpora* separadamente e com ambos combinados. Com os *corpus* combinados, foram avaliados dois métodos. O primeiro utilizando encadeamento de classificadores junto ao SMO, obteve acurácia de 65%. O segundo método utilizou Rakel combinado ao algoritmo de classificação J48 e obteve acurácia de 65%.

Em um trabalho mais recente, Fisas, Ronzano e Saggion (2015) apresentaram um *corpus* anotado a partir de 40 artigos da área de computação gráfica, contendo um total de 8.777 sentenças. Os papéis retóricos considerados foram: desafio, contexto, metodologia, contribuições e trabalhos futuros. As características extraídas foram baseadas no trabalho de Teufel e Moens (1999), com adição de *unigrams*, *bigrams*, *trigrams* e uma característica chamada de árvore de dependência. Os autores usaram os algoritmos de classificação SVM e Regressão Logística e obtiveram Macro F1-score de 76,4% e 80,1%, respectivamente.

3.1 Síntese dos Trabalhos Relacionados

A Tabela 3.1 sumariza os trabalhos discutidos na seção anterior, apresentando quais classes de papéis retóricos foram consideradas, o principal classificador e, por fim, o melhor resultado obtido. A maioria dos trabalhos utilizou um *corpus* em inglês, exceto o trabalho de Feltrim et al. (2004) que utilizou um em português.

Tabela 3.1: Resumo das características e dos resultados principais dos trabalhos abordados nesta seção

Autor	Classificador	Papéis Retóricos	Resultados
Teufel e Moens (1999)	<i>Naive Bayes</i>	<i>Aim, Contrast, Textual, Own, Background, Basis, Other</i>	Acurácia de 71,2%
Anthony e Lashkia (2003)	<i>Naive Bayes</i>	<i>Claim centrality, Generalize topic, Gap, Announce research, Findings, Evaluate</i>	Acurácia de 68%
Feltrim et al. (2004)	<i>Naive Bayes</i>	<i>Background, Gap, Purpose, Methodology, Results, Conclusion, Outline</i>	Acurácia de 74%
Hirohata et al. (2008)	CRF	<i>Objective, Method, Results, Conclusions</i>	Acurácia de 95,5%
Yepes, Mork e Aronson (2013)	CRF	<i>Background, Conclusions, Methods, Objective, Results</i>	Macro F1-score de 79,33%
Guo, Korhonen e Poibeau (2011)	ASSVM	<i>Background, Objective, Method, Result, Conclusion</i>	Macro F1-score de 76%
Dayrell et al. (2012)	J48	<i>Background, Gap, Purpose, Method, Result, Conclusion</i>	Acurácia de 65%
Fisas, Ronzano e Saggion (2015)	SVM	<i>Approach, Background, Challenge, Future Work, Outcome</i>	Macro F1-score de 80,1%

Fonte: Elaborada pelo autor.

Metodologia

Neste capítulo, são apresentadas as etapas necessárias e o detalhamento das decisões tomadas para o desenvolvimento do trabalho.

A seção 4.1 apresenta o *corpus* utilizado no trabalho, detalhando como foi realizada a coleta do material, quais resumos foram selecionados, a estruturação utilizada, e por fim uma análise dos dados que foram coletados e selecionados. A seção 4.2 detalha a escolha de RNNs para a predição de estrutura retórica e apresenta uma arquitetura de rede recorrente proposta para este problema de classificação. Por fim, na seção 4.3 os algoritmos de aprendizagem de máquina que foram utilizados como *baselines* são apresentados e quais características foram extraídas do *corpus* para serem utilizadas com tais algoritmos.

4.1 Criação do corpus

Apesar de existirem *corpora* de resumos científicos anotados com papéis retóricos para a língua portuguesa (FELTRIM et al., 2004), a quantidade de amostra é pequena. Para a língua inglesa, os *corpora* disponíveis não são de resumos científicos e também possuem uma quantidade de amostra pequena. Visando a criação de *corpora* grandes, trabalhos como o de Hirohata et al. (2008) e Yepes, Mork e Aronson (2013) utilizaram uma base pública de artigos científicos para a criação de um *corpus* de forma automática, resultando em um *corpus* com uma grande quantidade de resumos de artigos científicos com seus papéis retóricos anotados. Portanto, de modo a evitar que os resultados sejam afetados devido a um *corpus* com uma quantidade baixa de amostras, seguimos uma metodologia similar a proposta no trabalho de Hirohata et al. (2008) para a criação de um corpus para a língua inglesa.

4.1.1 Coleta

Assim como em Hirohata et al. (2008), os resumos foram extraídos da MEDLINE/PubMed[®]¹. A MEDLINE/PubMed[®] é uma base de dados da U.S. National Library of Medicine (2011) que contém mais de 27 milhões de citações da literatura biomédica de diversos livros e revistas. Essas citações incluem resumos e links para as páginas das publicadoras.

É comum que resumos de artigos científicos na literatura biomédica sejam divididos em seções que indicam um papel retórico. Mostramos essa característica na Figura 4.1 na qual é apresentado um exemplo de um resumo dividido nas seções: Contexto (Background), Objetivos (Objectives), Métodos de Pesquisa (Search Methods), Critério de Seleção (Selection Criteria), Coleta de dados e Análise (Data Collection and Analysis), Resultados Principais (Main Results) e Conclusões dos Autores (Authors Conclusions). Devido a essa característica, a MEDLINE/PubMed[®] consegue transformar e armazenar em seu banco de dados cada seção dos resumos de artigos científicos disponíveis na sua base.

A MEDLINE/PubMed[®] disponibiliza² gratuitamente para download todos esses artigos e suas informações em formato XML (eXtensible Markup Language). Como é possível observar na Figura 4.2, que representa um artigo que possui o resumo segmentado, cada seção do texto é definida por um *AbstractText* que contém os atributos *Label* e *NlmCategory*. O valor do atributo *Label* é dado pelo o que foi especificado pelo próprio autor no texto do resumo. Já o valor do atributo *NlmCategory* é o resultado de um mapeamento³ realizado pela U.S. National Library of Medicine (2011) entre os diversos tipos de categorias de papéis retóricos, dados pelos autores de todos os resumos da base, em cinco categorias definidas pela NLM. São elas: Contexto (*Background*), Métodos (*Methods*), Conclusões (*Conclusions*), Resultados (*Results*) e Objetivo (*Objective*).

A criação do *corpus* criado nesse trabalho se deu em duas etapas. Na primeira etapa foram coletados da MEDLINE/PubMed[®] 3.326.605 resumos, que possuem 11.751 categorias de papéis retóricos distintos especificadas pelos autores. Na Tabela 4.1 mostramos que existem classes de papéis retóricos que, apesar de terem nomes distintos, são semelhantes entre si e possuem uma mesma função retórica no resumo.

Na segunda etapa, foram selecionados somente os resumos que possuem em sua estrutura todas as cinco categorias NLM. Após essa filtragem, nosso *corpus* foi reduzido para 387.705 resumos. Esses resumos por sua vez foram segmentados em sentenças utilizando *Punkt*, uma ferramenta disponibilizada na biblioteca NLTK (LOPER; BIRD,

¹<https://www.ncbi.nlm.nih.gov/pubmed/>

²Disponível em: <ftp://ftp.ncbi.nlm.nih.gov/pubmed/baseline>

³Disponível em: <https://structuredabstracts.nlm.nih.gov/downloads.shtml>

Figura 4.1: Exemplo de um resumo de artigo da literatura biomédica dividido em seções.

BACKGROUND: Malignant pleural mesothelioma is an almost always fatal tumour, for which palliative platinum-based chemotherapy is currently the standard treatment. Multimodal therapeutic strategies incorporating surgery, radiation therapy or photodynamic therapy and chemotherapy have been recommended for selected patients but there is no consensus about their effectiveness.

OBJECTIVES: To assess the benefits and harms of radical multimodal treatment options (including radical surgery +/- radical radiotherapy +/- photodynamic therapy +/- systemic therapy) compared to each other or to palliative treatments, for people with malignant pleural mesothelioma.

SEARCH METHODS: We reviewed data from the Cochrane Lung Cancer group's Specialised Register, Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE and Embase. We also checked reference lists of primary original studies, review articles and relevant conference proceedings manually for further related articles up to 21 March 2017.

SELECTION CRITERIA: We included parallel-group randomised controlled trials of multimodal therapy for people with malignant pleural mesothelioma (stages I, II or III) that measured at least one of the following endpoints: overall survival, health-related quality of life, adverse events or progression-free survival. We considered studies regardless of language or publication status.

DATA COLLECTION AND ANALYSIS: Two review authors independently extracted relevant information on participant characteristics, interventions, study outcomes, and data on the outcomes for this review, as well as information on the design and methodology of the studies. Two review authors assessed the risk of bias in the included trials using pre-defined 'Risk of bias' domains. We assessed the methodological quality using GRADE.

MAIN RESULTS: We conducted this review in accordance with the published Cochrane protocol. Two randomised clinical trials with 104 participants fulfilled our inclusion criteria. Both trials were at high risk of bias (for outcomes other than overall survival), and we rated the evidence as moderate quality for overall survival and low quality for all other outcomes. One trial compared combined extrapleural pneumonectomy (EPP) plus neoadjuvant platinum-based chemotherapy plus postoperative high-dose hemithoracic radiotherapy with combined EPP plus platinum-based chemotherapy. The other trial compared EPP plus postoperative hemithoracic radiotherapy with standard (non-radical) therapy alone following platinum-based chemotherapy (patients in the standard therapy arm received continued oncological management according to local policy, which could include further chemotherapy or palliative radiotherapy). For the first trial, median overall survival calculated from registration was 20.8 months (95% confidence interval (CI) 14.4 to 27.8) in the no-radiotherapy group and 19.3 months (95% CI 11.5 to 21.8) in the radiotherapy group. For the second trial, median overall survival was 14.4 months (95% CI 5.3 to 18.7) for patients allocated to EPP and 19.5 months (95% CI 13.4 to time not yet reached) for patients randomised to standard non-radical therapy. In the second trial, 12 serious adverse events were reported during the study period: ten in the EPP group and two in the non-radical therapy group. Overall health-related quality of life scores were not different between the two arms in either study. We could not perform a meta-analysis of the two included trials due to clinical heterogeneity. We also identified three ongoing trials evaluating the topic of our review.

AUTHORS CONCLUSIONS: The overall strength of the evidence gathered in this review is low and there is a lack of available evidence to support the use of radical multimodality therapy in routine clinical practice (particularly as one trial suggests greater harm). Given the added cost of multimodality treatment and the possible increase in risk of adverse effects, the lack of evidence of their effectiveness probably means that these interventions should currently be limited to clinical trials alone.

Fonte: Elaborada pelo autor.

2002). Cada sentença obtida pela segmentação foi anotada com o mesmo papel retórico da seção a qual essa sentença pertence. Por exemplo, todas as sentenças dentro de uma seção com a categoria Métodos foram anotadas como sendo da categoria Métodos. A Figura 4.3 mostra como exemplo um resumo do *corpus* após a etapa de filtragem e segmentação de sentenças.

4.1.2 Estruturação

Devido a grande quantidade de resumos, eles foram armazenados em um formato binário, resultando em um arquivo com aproximadamente 720MB. Foi utilizada uma estrutura simples de listas, na qual o *corpus* é uma lista de resumos. Cada resumo, por sua vez, também é uma lista formada por sentenças. As sentenças são formadas por uma lista de tuplas, nas quais o primeiro valor da tupla armazena o texto da sentença e o segundo valor armazena o papel retórico da sentença.

Figura 4.2: Exemplo do XML da MEDLINE/PubMed®

```

<PubmedArticle>
  <MedlineCitation Status="Publisher" Owner="NLM">
    ...
    <Article PubModel="Print-Electronic">
      ...
      <Abstract>
        <AbstractText Label="BACKGROUND" NlmCategory="BACKGROUND">
          Malignant pleural mesothelioma is an almost always fatal tumour...
        </AbstractText>
        <AbstractText Label="OBJECTIVES" NlmCategory="OBJECTIVE">
          To assess the benefits and harms of radical multimodal treatment options...
        </AbstractText>
        <AbstractText Label="SELECTION CRITERIA" NlmCategory="METHODS">
          We included parallel-group randomised controlled trials of multimodal therapy...
        </AbstractText>
        <AbstractText Label="DATA COLLECTION AND ANALYSIS" NlmCategory="METHODS">
          Two review authors independently extracted relevant information on...
        </AbstractText>
        <AbstractText Label="MAIN RESULTS" NlmCategory="RESULTS">
          We conducted this review in accordance with the published Cochrane protocol...
        </AbstractText>
        <AbstractText Label="AUTHORS CONCLUSIONS" NlmCategory="CONCLUSIONS">
          The overall strength of the evidence gathered...
        </AbstractText>
      </Abstract>
    </Article>
  </MedlineCitation>
  ...
</PubmedArticle>

```

Fonte: Abdel-Rahman et al. (2017)

Figura 4.3: Exemplo de um resumo do *corpus*

Background: Irritable bowel syndrome (IBS) is a very common functional gastrointestinal (GI).
Background: Diagnosis of IBS is based on the fulfilment of the Rome III criteria.
Background: Common GI symptoms are lower abdominal pain, bloating and disturbed defecation, such as urgent diarrhoea and/or episodes of chronic constipation.
Background: Many agents have been employed in the management of IBS, although only few have been demonstrated to show a relevant efficacy.
Objective: To evaluate the effectiveness of the administration of a mixture of beta-glucan, inositol and digestive enzymes (Biointo) in improving GI symptoms in patients affected by IBS.
Methods: 50 IBS patients (20 males, 30 females; mean age 51 +/- 19) were treated with Biointo (group A) while another group consisting of 40 IBS patients (15 males, 25 females; mean age 50 +/- 18) did not receive any therapy (group B).
Results: Biointo administration improved significantly bloating, flatulence and abdominal pain, with a slight increasing of urgency for bowel movements.
Results: On the contrary, Biointo did not show any significant effect on the other IBS symptoms.
Conclusions: Currently, only few agents used in the management of IBS have been proven to be effective.
Conclusions: Biointo administration has shown to improve some IBS symptoms, such as bloating, flatulence and abdominal pain, all connected to the presence of gas inside the intestinal lumen.

Fonte: Elaborada pelo autor.

4.1.3 Análise

Esta seção apresenta uma análise do *corpus* relativa às seguintes características: tamanho das sentenças, tamanho dos resumos, distribuição dos papéis retóricos, distribuição das transições entre papéis retóricos, e sequências de transições mais frequentes.

Tabela 4.1: As 20 categorias de papéis retóricos mais frequentes das seções obtidas.

Nome	Quantidade
RESULTS	2.879.431
METHODS	1.940.656
CONCLUSIONS	1.729.860
CONCLUSION	1.314.384
BACKGROUND	1.265.519
OBJECTIVE	935.306
PURPOSE	392.008
OBJECTIVES	337.356
DESIGN	236.286
MATERIALS AND METHODS	223.725
INTRODUCTION	180.014
METHOD	165.372
AIM	157.016
SETTING	145.346
STUDY DESIGN	97.625
PATIENTS AND METHODS	96.592
AIMS	88.628
METHODS AND RESULTS	75.999
DISCUSSION	72.654
MATERIAL AND METHODS	72.446

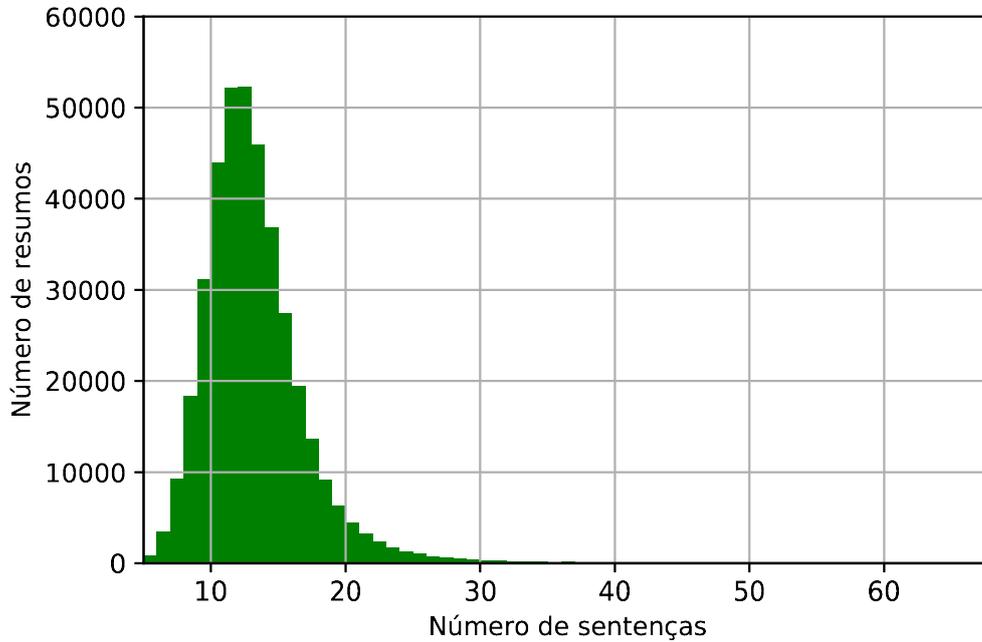
Fonte: Elaborada pelo autor.

A Figura 4.4 mostra a distribuição de resumos em relação ao tamanho dos resumos. É possível observar que grande parte dos resumos do *corpus* possuem menos que 30 sentenças, correspondendo a 99% desses resumos. O menor resumo possui 5 sentenças, e como explicado anteriormente, isso ocorre devido ao fato de todos os resumos possuírem pelo menos uma sentença para cada papel retórico (Contexto, Objetivo, Métodos, Resultados, Conclusões). Em contrapartida, o maior resumo do *corpus* possui 68 sentenças. A quantidade média de sentenças por resumo é de 13.

A Figura 4.5 mostra a quantidade de sentenças em relação ao tamanho das sentenças. O tamanho da sentença é dado pela quantidade de caracteres que ela contém. É possível observar que grande parte das sentenças dos resumos possuem menos que 300 caracteres, correspondendo a 97,24% de todas as sentenças. Em média, uma sentença contém 140 caracteres.

A Tabela 4.2 mostra a distribuição de papéis retóricos no *corpus*. Como é possível observar, essa distribuição não é balanceada. O papel retórico mais frequente é o de Resultados, sendo 32,8% do total de sentenças, e o papel retórico menos frequente é o

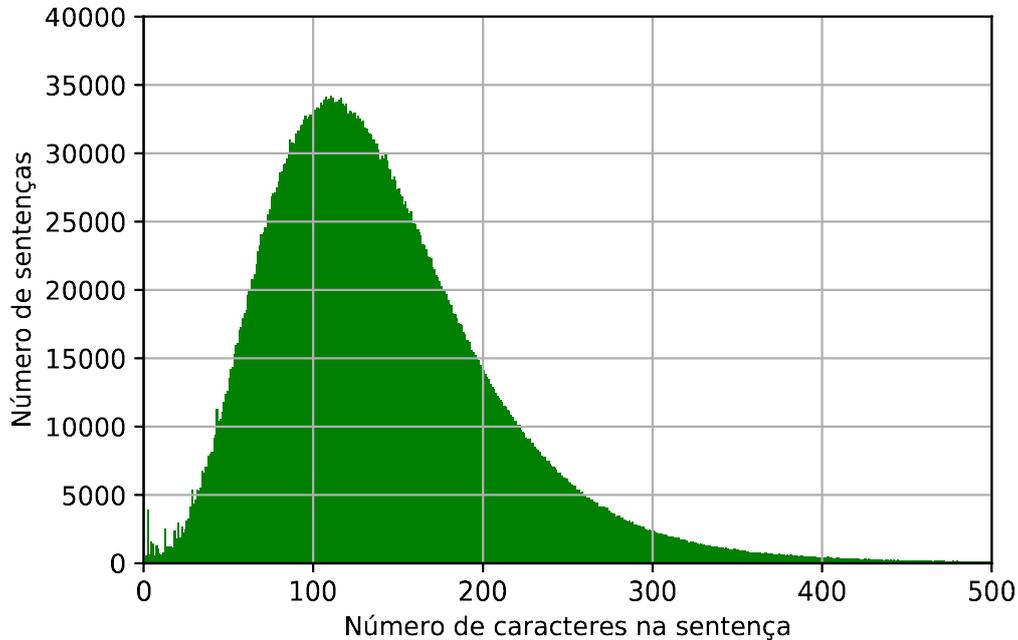
Figura 4.4: Histograma do tamanho dos resumos no *corpus*



Fonte: Elaborada pelo autor.

Objetivo, correspondendo a 8,7% do total de sentenças. Essa distribuição era esperada, já que em resumos de artigos científicos os autores buscam enfatizar os resultados encontrados em seus trabalhos, o que causa essa proporção maior nessa categoria. Já a baixa ocorrência de sentenças da categoria Objetivos pode ser atribuída ao fato dos autores buscarem enunciar os objetivos de seus trabalhos de forma direta, muitas vezes em uma única sentença. Também é possível observar que a categoria Métodos possui a segunda maior porcentagem de ocorrências, correspondendo a 26,5% do total de sentenças. O motivo para essa grande porcentagem também é devido a importância dessa categoria para os autores, que precisam explicar a metodologia empregada nos seus trabalhos de forma a atrair leitores interessados. Ambos papéis retóricos (Resultados e Métodos) correspondem a 59,3% do total de sentenças do *corpus*, o que demonstra a importância dessas categorias para resumos de artigos científicos.

A Tabela 4.3 mostra a quantidade de transições entre papéis retóricos, início do resumo (#s) e fim do resumo (#e), seguindo a ordem de *linha da tabela* → *coluna da tabela*. A partir dela podemos observar que a maior parte dos papéis retóricos que iniciam um

Figura 4.5: Histograma do tamanho das sentenças no *corpus*

Fonte: Elaborada pelo autor.

Tabela 4.2: Distribuição das sentenças no *corpus*

Papel Retórico	Número de ocorrências	Porcentagem de ocorrências
Resultados	1.612.851	32,8%
Métodos	1.306.519	26,5%
Conclusões	806.730	16,4%
Contexto	770.009	15,6%
Objetivo	427.928	8,7%
Total	4.924.037	100%

Fonte: Elaborada pelo autor.

resumo são da categoria Contexto, com 84,04% (325.861) do total (387.705) de sentenças que iniciam os resumos (#s). Também é possível observar que em sentenças que terminam os resumos (#e), 95% (369.965) possuem o papel retórico Conclusões. Grande parte das transições do *corpus* ocorrem entre a mesma categoria, por exemplo, dado uma sentença de Contexto de um resumo, a categoria mais provável para a próxima sentença também

será Contexto. O único papel retórico que não segue essa tendência é Objetivo, para o qual a próxima sentença tem maior probabilidade de ser da categoria Métodos.

Tabela 4.3: Distribuição de transições de papéis retóricos no *corpus*

Papel Retórico	Contexto	Objetivo	Métodos	Resultados	Conclusões	#e	Total
#s	325.861	47.729	14.094	1	20	0	387.705
Contexto	377.641	325.565	53.249	362	250	12.942	770.009
Objetivo	50.882	40.146	335.374	235	1.253	38	427.928
Métodos	1.147	14.251	899.199	386.858	766	4.298	1.306.519
Resultados	140	15	375	1.224.793	387.066	462	1.612.851
Conclusões	14.338	222	4.228	602	417.375	369.965	806.730

Fonte: Elaborada pelo autor.

A Tabela 4.4 mostra as 10 sequências de papéis retóricos mais frequentes no *corpus*. É possível visualizar que a maior parte dos resumos possui a sequência "Contexto → Objetivos → Métodos → Resultados → Conclusões", correspondendo a 82,17% do total de resumos. Essa característica era esperada, pois é sabido que resumos de artigos científicos possuem uma grande padronização, o que faz com que exista uma prevalência de um tipo de sequência. Também é possível observar que grande parte dos resumos do *corpus* segue uma estrutura similar a proposta por Weissberg e Buker (1990), indicando que de fato essa estrutura é característica de resumos de artigos científicos.

Tabela 4.4: As 10 mais frequentes sequências de transições no *corpus*

Sequência de transições	Total de ocorrências (%)
Contexto → Objetivos → Métodos → Resultados → Conclusões	318570 (82.17)
Objetivos → Contexto → Métodos → Resultados → Conclusões	36998 (9.54)
Métodos → Objetivos → Contexto → Métodos → Resultados → Conclusões	11748 (3.03)
Objetivos → Métodos → Resultados → Conclusões → Contexto	8095 (2.09)
Contexto → Objetivos → Métodos → Resultados → Conclusões → Contexto	3443 (0.89)
Contexto → Objetivos → Métodos → Resultados → Conclusões → Métodos	2102 (0.54)
Métodos → Objetivos → Contexto → Métodos → Resultados → Conclusões → Métodos	1782 (0.46)
Objetivos → Conclusões → Contexto → Métodos → Resultados → Conclusões → Contexto	962 (0.25)
Objetivos → Métodos → Contexto → Métodos → Resultados → Conclusões	599 (0.15)
Contexto → Objetivos → Métodos → Conclusões → Resultados	306 (0.08)

Fonte: Elaborada pelo autor.

4.2 Redes neurais recorrentes para predição de estrutura retórica

As RNNs estão sendo utilizadas com sucesso em diversos problemas de PLN, como por exemplo, a detecção de paráfrase (SOCHER et al., 2011), análise de sentimento (SOCHER et al., 2013), extração de opinião (IRSOY; CARDIE, 2014) e predição de interesse (GAO et al., 2014) e atingiram resultados superiores aos considerados estado da arte.

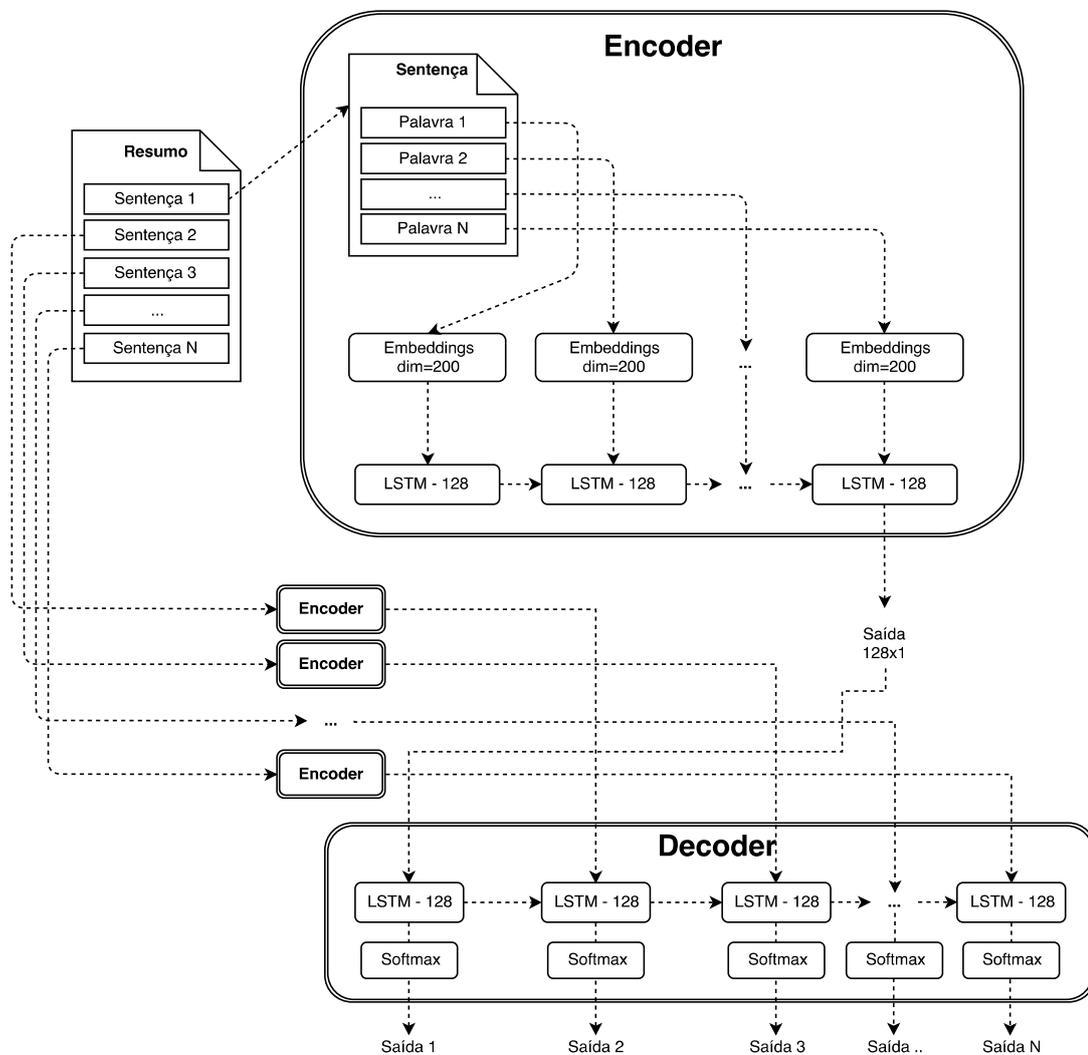
4.2.1 Arquitetura

De modo a tirar o melhor proveito das redes neurais recorrentes, este trabalho tem interesse na predição da estrutura retórica como um todo em vez de classificar cada sentença individualmente. Assim, a tarefa de classificação de papéis retóricos em resumos de artigos científicos foi considerada como um problema de rotulação sequencial. Como

discutido anteriormente no capítulo de trabalhos relacionados, estudos anteriores usaram métodos similares, tendo como destaque o uso do algoritmo de classificação *CRF*.

Para evitar problemas de explosão ou de desvanecimento de gradiente na rede, foram utilizadas camadas *LSTM* para a criação da nossa arquitetura de rede recorrente. Cho et al. (2014) utilizou a arquitetura *encoder-decoder* para obter representações intermediárias entre dois idiomas distintos. Neste trabalho, o *encoder-decoder* foi utilizado para obter representações intermediárias das sentenças dos resumos. Desse modo, foi possível transformar a sequência de palavras na sentença do resumo em um vetor que representa essa sequência.

Figura 4.6: Arquitetura de rede neural recorrente proposta



Fonte: Elaborada pelo autor.

A Figura 4.6 ilustra a arquitetura de rede neural recorrente proposta com as recorrências expandidas, na qual cada sentença de um resumo é transcodificada pelo *encoder* em um vetor de dimensão fixa (no caso, 128). Em seguida cada vetor de saída do *encoder* é utilizado como entrada para o *decoder*. A saída do *decoder* utiliza uma camada *softmax* para retornar a probabilidade do papel retórico de cada sentença do resumo.

A entrada do *encoder* é uma sentença, ou seja, um vetor com palavras, entretanto, como redes neurais não trabalham diretamente com texto, a entrada primeiramente deve ser transformada. Essa transformação é feita de modo simples, cada palavra da sentença é substituída pelo seu índice em um dicionário. No caso deste trabalho, o dicionário é criado a partir de todas as palavras que pertencem ao conjunto de treinamento da rede, mas vale notar que também é possível utilizar um dicionário criado externamente. Caso uma palavra não exista no dicionário, ela é removida de sua sentença. A partir do índice da palavra a camada de *embeddings* seleciona um vetor correspondente, ou seja, cada palavra têm um vetor próprio. Para este trabalho foi utilizada uma camada de *embeddings* que armazena vetores de 200 dimensões. Esse valor foi selecionado devido às limitações de *hardware* no qual a rede foi avaliada. A saída da camada de *embeddings* é a entrada para uma camada *LSTM*, portanto, cada índice de palavra na entrada do *encoder* passa pela camada de *embeddings* e em seguida é utilizado como entrada para a *LSTM*. A camada *LSTM* do *encoder* tem uma unidade oculta de tamanho 128, valor que foi escolhido também devido a limitações de *hardware*. O *encoder* retorna somente a saída da última recorrência da camada *LSTM*.

O *decoder* têm como entrada a saída do *encoder* para cada sentença de um resumo, ou seja, a entrada é uma sequência de vetores de 128 de dimensão.

A arquitetura foi implementada utilizando *keras* (CHOLLET et al., 2015), uma *framework* para o desenvolvimento de redes neurais profundas e recorrentes. Essa *framework* foi escolhida pois as camadas (*embeddings*, *LSTM*, *softmax*) necessárias para a arquitetura da rede proposta já estão implementada, outro motivo para a escolha foi a facilidade de poder utilizar GPUs para a realização do treino e teste da rede proposta. Para realizar o treino, foi selecionado o otimizador *adam* (KINGMA; BA, 2014) com seus parâmetros padrões.

4.2.2 Dados de entrada

Foi utilizado como dados de entrada o *corpus* criado neste trabalho. Porém, para isso, foi preciso fazer uma transformação nesses dados de forma a que seja possível utiliza-los

como entrada a uma rede neural. Essa transformação foi feita em três etapas: tokenização, fixação do tamanho das sentenças e *bucketing*.

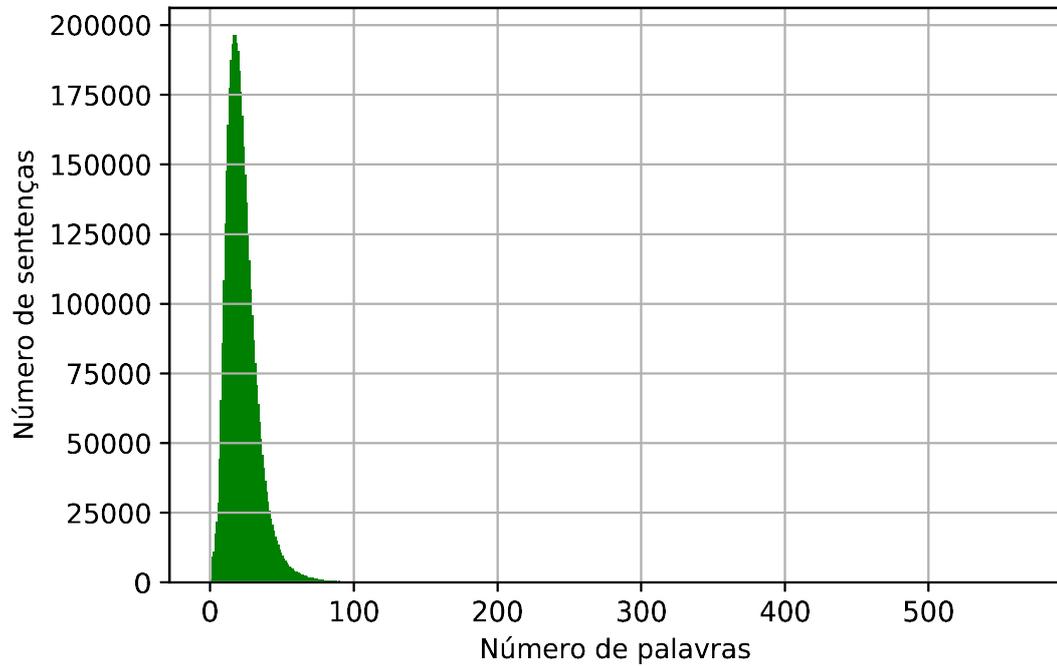
A tokenização é a etapa na qual o texto das sentenças dos resumos são separados em palavras. A separação em palavras na língua inglesa em um *corpus* com uma linguagem formal é relativamente fácil, pois cada palavra é separada por espaços. Para evitar qualquer ruído junto as palavras, foi aplicado um filtro dos seguintes caracteres: `!"#$%&'()*+,-./:;<=>?@[_`{|}~\t\n`

Após a tokenização, foi obtido um conjunto de resumos com suas sentenças separadas em palavras. Como redes neurais recorrentes precisam de uma quantidade fixa de sequências para sua execução, nesse caso, a quantidade de palavras em uma sentença, foi realizada uma pequena análise nesse conjunto para considerar qual seria o tamanho ideal a ser considerado. A quantidade média de palavras nas sentenças do conjunto é de 22, com 11 de desvio padrão. De modo a poder utilizar o máximo de palavras das sentenças, foi escolhida a quantidade máxima de 50 palavras por sentença. Como é possível observar na Figura 4.7, esse valor corresponde a grande parte das sentenças do *corpus*, mais especificamente, corresponde a 97% de todo o conjunto de sentenças. Com isso, sentenças que possuíam mais que 50 palavras foram reduzidas para apenas as 50 primeiras palavras, para as sentenças que possuíam menos de 50 palavras foi adicionado uma palavra reservada que corresponde a ausência de uma palavra.

De maneira similar as sentenças, os resumos do *corpus* formam uma estrutura sequencial, na qual cada resumo possui uma sequência de sentenças. Como dito anteriormente, para execução de tal estrutura com redes neurais recorrentes, é preciso fixar uma quantidade fixa de sequências. Porém, diferentemente de palavras em uma sentenças, não é possível delimitar uma quantidade de modo a não afetar a estrutura retórica do resumo. Uma forma simples de solucionar esse problema seria utilizar o valor da quantidade de sentenças do maior resumo. Entretanto, como vimos anteriormente, 99% dos resumos do *corpus* possuem menos que 30 sentenças, portanto a maior parte dos resumos não utilizariam essas sequências, mas para a execução da rede ainda seria utilizado esse valor, o que acarretaria em um desperdício de memória e faria com que o tempo de execução para o treino e teste da rede aumentassem significativamente.

Para resolver esse problema, foi utilizado um método chamado de *bucketing*. Esse método consegue lidar com dados que têm diferenças nos tamanhos de sequências criando *buckets* (baldes, em português) que têm o mesmo tamanho. Para exemplificar o método, considere quatro amostras com tamanhos de sequência de 2, 3, 5, 7, é possível criar dois *buckets* que juntem as amostras que possuem sequências com tamanhos nos intervalos de 2-4 e 4-7, nos quais as amostras com sequências de tamanho maior que 2 e menor ou igual

Figura 4.7: Histograma do tamanho das sentenças em palavras do *corpus*



Fonte: Elaborada pelo autor.

a 4 estão no primeiro *bucket* e as amostras com sequências de tamanho maior que 4 e menor ou igual a 7 estão no segundo *bucket*.

Nesse trabalho foram utilizados 3 *buckets* para agrupar os resumos com tamanhos nos intervalos de 0-23, 23-46 e 46-68.

4.3 Baselines

Foram utilizados como *baselines* para a classificação de papéis retóricos dois algoritmos de aprendizagem de máquina supervisionada considerados estado-da-arte. A primeira *baseline* utiliza o algoritmo *SVM*. Sua implementação foi feita utilizando a biblioteca *scikit-learn* (PEDREGOSA et al., 2011), e foi empregado um *kernel* linear com os parâmetros padrão. A segunda *baseline* utiliza o algoritmo *CRF* (Conditional Random Fields). Para sua implementação foi utilizada a biblioteca *CRFSuite* (OKAZAKI, 2007), que têm um alto desempenho em sua execução. Os parâmetros utilizados foram os padrões da biblioteca.

Ambos algoritmos já foram utilizados com sucesso em trabalhos anteriores para a classificação de estruturas retóricas (HIROHATA et al., 2008; YEPES; MORRIS; ARONSON, 2013; FISAS; RONZANO; SAGGION, 2015). Esses trabalhos possuem diferenças em seus experimentos, como a quantidade de papéis retóricos a serem classificados e o tipo de estrutura retórica, porém em seus respectivos trabalhos, ambos *SVM* e *CRF* foram considerados os algoritmos com o melhor desempenho, o que motivou a escolha desses algoritmos.

A escolha de boas características é essencial para algoritmos de aprendizagem de máquina supervisionada, por este motivo foram utilizadas as características propostas por Hirohata et al. (2008). A escolha do trabalho de Hirohata et al. (2008) se deu por se tratar de um trabalho que também classificou papéis retóricos em resumos de textos científicos extraídos da MEDLINE/PubMed[®], que também utilizou essas características em ambos algoritmos *SVM* e *CRF*. Desse modo, as características que foram utilizadas nas *baselines* são:

1. Características extraídas do conteúdo da sentença. Essas características são representadas por palavras (unigramas) e palavras duplas (bigramas) da sentença. Diferentemente de Hirohata et al. (2008), não foi feita uma normalização nas palavras de entrada do texto. Para reduzir o grande número de características dessa categoria, foram selecionadas apenas as 20.000 características com o maior valor no teste de χ^2 (Qui-quadrado). As características dessa categoria são essenciais, pois tentam representar certos tipos de expressões que são diretamente associadas a um tipo de papel retórico. Um exemplo desse tipo de expressão utilizada é "Nós concluímos que..." que normalmente é utilizada somente em sentenças com o papel retórico de Conclusões.
2. Cinco características binárias que representam a posição relativa da sentença em um resumo, ou seja, blocos que representam um espaço na qual a sentença está inserida sendo o primeiro bloco representa estar mais próximo do início de um resumo e o último bloco represente que está mais próximo ao fim de um resumo. Essas características são importantes por darem ao classificador uma boa representação de alguns tipos de papéis retóricos que acontecem com mais probabilidade em certas posições do texto. Um exemplo é o papel retórico de Conclusões, que normalmente aparece nas últimas sentenças dos resumos. Foram utilizadas cinco características devido ao fato dos resumos do *corpus* possuírem no mínimo cinco sentenças, e com isso cada sentença consegue ser posicionada dentro dessas características.

3. Características extraídas do conteúdo da sentença anterior e próxima. Que corresponde a utilizar para a sentença atual as características descritas no primeiro item das sentenças anterior e da próxima, ou seja, inclui as características obtidas no primeiro item das sentenças vizinhas. Isso é importante pois cria características que são relativas as sentenças vizinhas e que, de certa forma, expressa a sequencialidade entre os papéis retóricos.

4.3.1 Comparação com as baselines

Para realizar a comparação entre as *baselines*, foram utilizadas as métricas de F1-macro, acurácia, e acurácia na classificação por resumo. A medida F1-macro e acurácia levam em consideração somente a classe individual das sentenças do *corpus*.

A acurácia é descrita na Equação 4.1, na qual tp é a quantidade de elementos classificados corretamente como pertencentes a uma classe e tn é a quantidade de elementos classificados corretamente como não pertencentes a uma classe, fp a quantidade de elementos classificados incorretamente como pertencentes a uma classe e fn elementos classificados incorretamente como não pertencentes a uma classe. Para calcular a medida F1-macro é necessário primeiramente calcular a medida F_1 (Equação 4.4) para cada papel retórico. Portanto, é preciso calcular a precisão (Equação 4.2) e revocação (Equação 4.3). Após isso, para calcular a F1-macro basta calcular a média entre os valores obtidos pela F_1 em cada papel retórico.

Para avaliar o desempenho dos algoritmos na classificação de todas as sentenças de um resumo, foi utilizada a métrica de acurácia por resumo. Para calcular essa métrica para resumos, só foram considerados corretamente classificados os resumos que tiveram todas as suas sentenças classificadas com o papel retórico correto. Os resultados dos experimentos das *baselines* e da arquitetura de rede proposta foram calculados a partir de 10 execuções de validação cruzada de 10-*folds*.

$$\text{Acurácia} = \frac{tp + tn}{tp + tn + fp + fn} \quad (4.1)$$

$$\text{precisão} = \frac{tp}{tp + fp} \quad (4.2)$$

$$\text{revocação} = \frac{tp}{tp + fn} \quad (4.3)$$

$$F_1 = 2 \cdot \frac{\text{precisão} \cdot \text{revocação}}{\text{precisão} + \text{revocação}} \quad (4.4)$$

Experimentos e Resultados

Este capítulo apresenta os experimentos realizados com a arquitetura de rede neural recorrente proposta, bem como os resultados obtidos. Também é feita uma discussão sobre os resultados de cada experimento realizado.

5.1 Comparação com as baselines

A *baseline SVM* teve nos resultados o pior desempenho, em relação à medida F1-macro seu melhor resultado nas execuções foi de 91,88% e o pior de 91,86%. Em acurácia por sentença o seu melhor resultado foi de 91,85% e o pior 91,84%. Por fim, na acurácia da classificação por resumos o melhor resultado foi 46,28% e o pior 46,24%. Apesar do desempenho do *SVM* ser menor que os outros algoritmos avaliados, seu resultado foi similar ao encontrado por Hirohata et al. (2008), que obteve em seu *corpus* expandido, no qual diversas classes de papéis retóricas foram agrupadas de modo similar a este trabalho, uma acurácia de 91,70% na classificação por sentenças e 44% na classificação por resumo.

Como esperado, os resultados da *baseline CRF* foram melhores que os da *baseline SVM*, devido ao algoritmo *CRF* utilizar informações a respeito das sequências de sentenças para a classificação. Na medida F1-macro o *CRF* teve o valor de 93,75% como seu maior resultado nas 10 execuções, e seu menor valor foi de 93,70%. A maior acurácia por sentença foi de 93,98%, e a menor foi de 93,93%. Na classificação dos papéis retóricos de resumos inteiros sua maior acurácia foi de 59,93% e seu menor valor foi de 59,77%. Esses resultados também são similares aos encontrados no trabalho de Hirohata et al. (2008), que obteve em seu *corpus* expandido uma acurácia de 93% na classificação de papéis retóricos por sentença e 55% na acurácia por resumo.

Apesar de ambos resultados das *baselines* serem próximas ao encontrado por Hirohata et al. (2008), vale destacar que existem diferenças significativas na metodologia. A principal delas é a diferença de quantidades de papéis retóricos: neste trabalho foram utilizadas 5 categorias (Contexto, Objetivo, Métodos, Resultados e Conclusões) a serem classificadas, enquanto Hirohata et al. (2008) utilizou 4 categorias (Objetivos, Métodos, Resultados e Conclusão).

Os resultados apresentados pela arquitetura de rede proposta foi superior as *baselines* em todas as métricas empregadas. Nas execuções, a rede teve o melhor F1-macro de 94,77% e o pior de 94,47%. Na medida de acurácia por sentença, o melhor resultado foi de 94,86% e o pior de 94,52%. Na acurácia da classificação de resumos, o melhor resultado foi de 63,11% e o pior de 61,77%.

Na Tabela 5.1 é apresentado os valores de precisão, revocação e F1-score por categoria na execução de todas as validações cruzadas. Em todas as classes de papéis retóricos, a rede obteve um F1-score maior que as *baselines*. Na medida de revocação e precisão, os resultados da rede só foram inferiores ao da *baseline CRF* para os papéis retóricos Resultados e Conclusões, respectivamente.

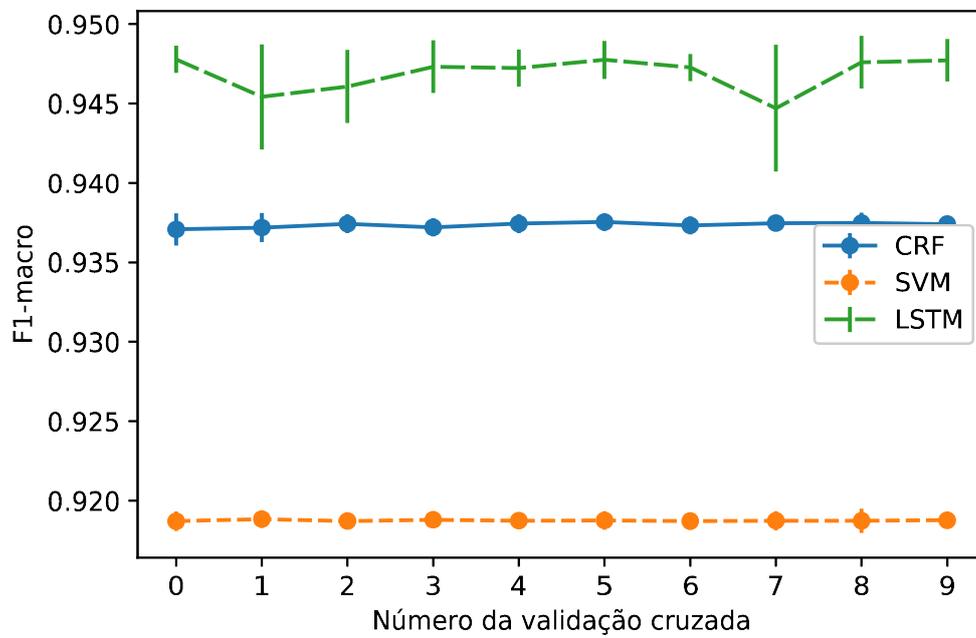
A Figura 5.1, Figura 5.2 e Figura 5.3 mostram gráficos com os resultados das 10 execuções da validação cruzada com 10-*folds* para as métricas de F1-macro, acurácia, e acurácia por resumo, respectivamente. A partir desses gráficos fica claro que a rede proposta se saiu melhor que as *baselines* em todas as métricas. Também é possível visualizar que os resultados dentro da validação cruzada da rede oscila mais (tem um desvio padrão maior) que as *baselines*.

Tabela 5.1: Precisão (P), Revocação (R) e F1-score.

Papel retórico	CRF			SVM			Rede Neural Proposta		
	P	R	F1	P	R	F1	P	R	F1
Contexto	95,42	96,04	95,73	94,24	95,57	94,90	96,40	96,97	96,68
Conclusões	92,47	91,25	91,86	89,64	88,05	88,84	92,03	93,28	92,65
Métodos	95,03	95,16	95,09	92,95	93,26	93,10	96,06	95,82	95,94
Objetivo	93,87	90,85	92,33	94,28	89,07	91,60	95,20	92,89	94,03
Resultados	93,16	94,18	93,67	90,30	91,58	90,93	94,19	94,08	94,14

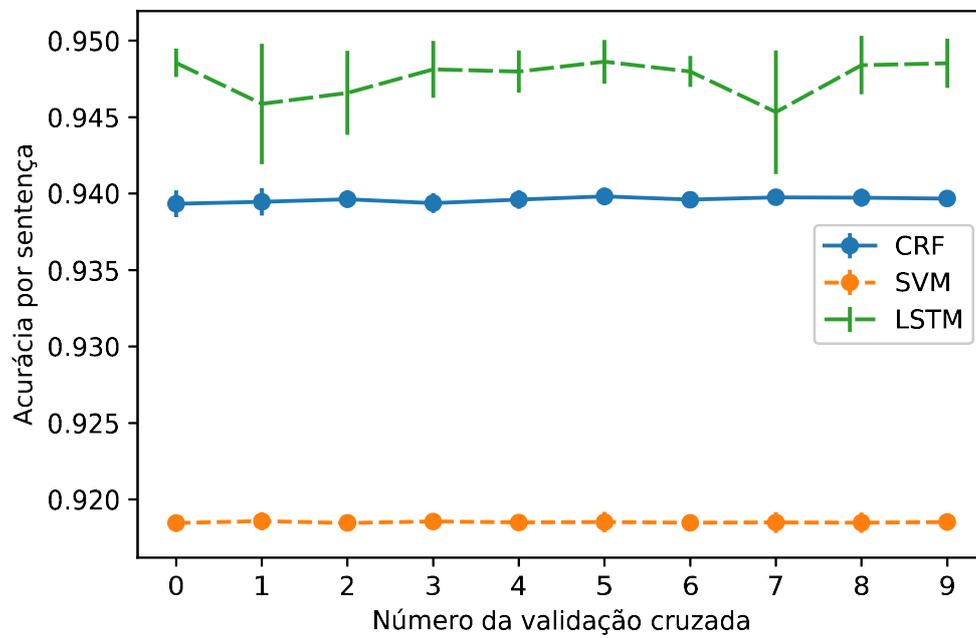
Fonte: Elaborada pelo autor.

Figura 5.1: F1-macro das execuções da validação cruzada



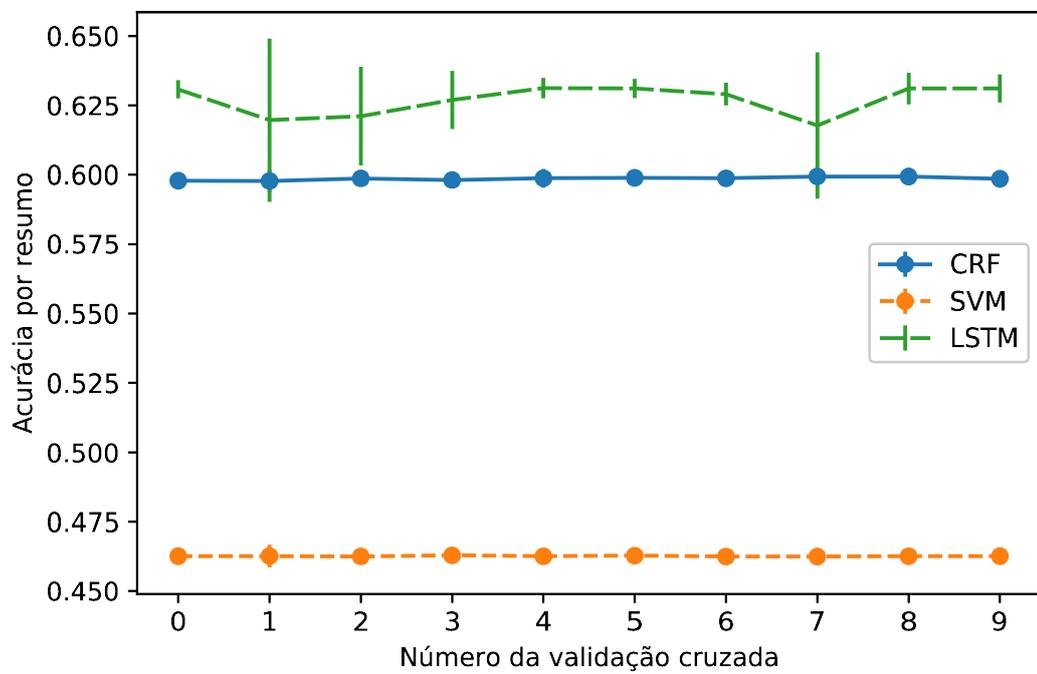
Fonte: Elaborada pelo autor.

Figura 5.2: Acurácia das execuções da validação cruzada



Fonte: Elaborada pelo autor.

Figura 5.3: Acurácia por resumo das execuções da validação cruzada



Fonte: Elaborada pelo autor.

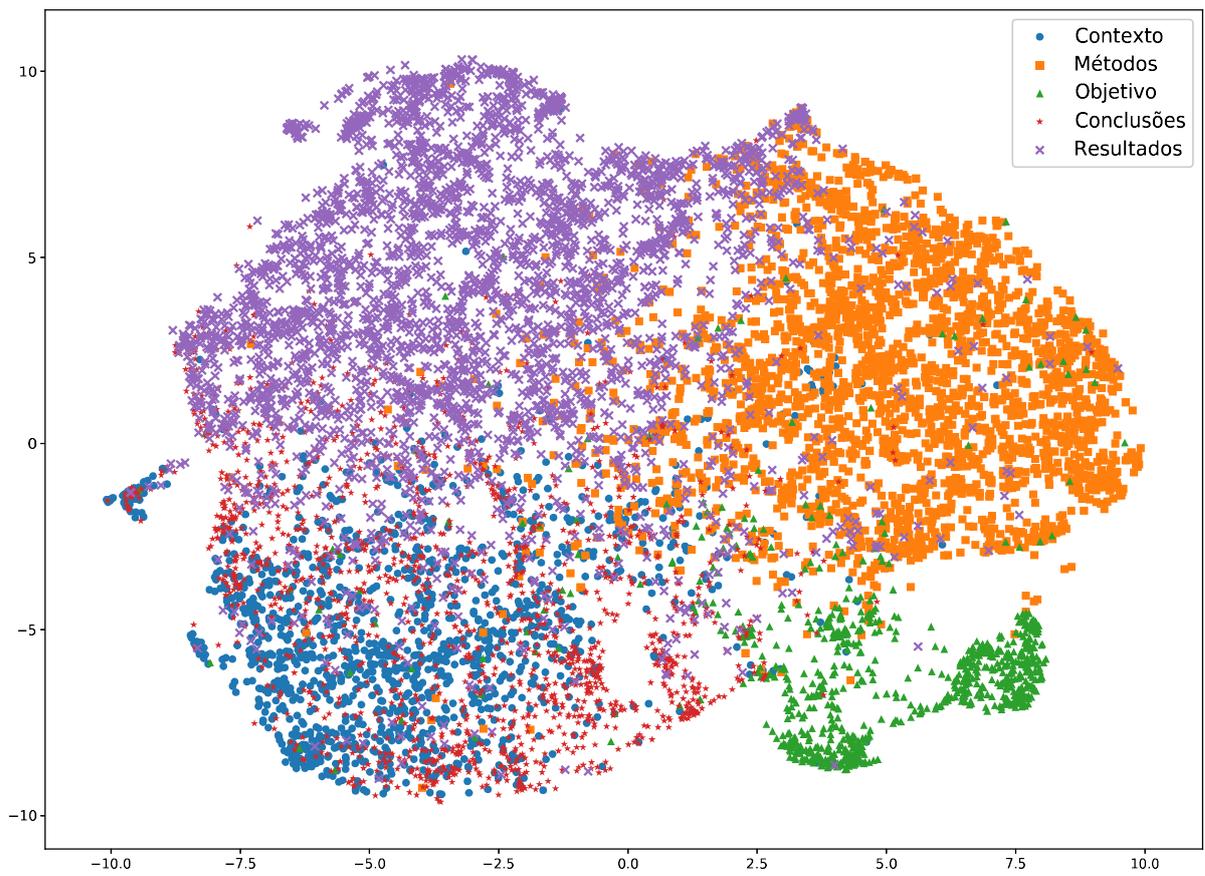
5.2 Análise do encoder

Para obter mais detalhes sobre de que forma a rede está classificando os papéis retóricos, foi realizado um experimento para visualizar a saída do *encoder* da rede proposta. Com essa visualização é possível analisar se as características extraídas pelo *encoder* são expressivas.

Esse experimento foi realizado utilizando o primeira execução da validação cruzada com 10 *folds*. O treino da rede foi realizado da mesma forma que os outros experimentos, somente utilizando o conjunto de treinamento. Em seguida foram selecionadas aleatoriamente 10.000 sentenças do conjunto de teste. Cada sentença selecionada foi então utilizada como entrada para a camada de *encoder* e as ativações da camada de saída do *encoder* foram armazenadas. Os vetores armazenados têm 128 dimensões, portanto, para a visualização em um gráfico foi necessário utilizar um algoritmo de redução de dimensionalidade. Para isso, foi utilizado o algoritmo t-SNE (*t-distributed Stochastic Neighbor Embedding*) (MAATEN; HINTON, 2008). A Figura 5.4 mostra o resultado desse experimento.

É possível visualizar na Figura 5.4 que a maior parte da sentenças com o mesmo papel retórico permanecem próximas. Somente os papéis retóricos de Contexto e Conclusões aparecem mesclados no gráfico, indicando uma possível dificuldade em classificar sentenças pertencentes a essas categorias. Os papéis retóricos de Métodos, Resultados e Objetivo criaram espaços bem definidos, na qual é possível dividir esses espaços com retas. Isso indica que o espaço dimensional criado pelo *encoder* consegue separar os papéis retóricos que são diferentes, ou seja, os vetores gerados pelo *encoder* extraem características importantes de sua entrada (no caso, as entradas são vetores de índices de palavras).

Figura 5.4: Gráfico da saída do *encoder*



Fonte: Elaborada pelo autor.

Conclusões

Este capítulo apresenta uma discussão sobre os principais pontos observados em relação aos experimentos realizados, bem como uma visão geral sobre o trabalho desenvolvido e possíveis trabalhos futuros.

Como não existia um *corpus* publicamente disponível para a classificação de papéis retóricos com uma grande quantidade de amostras, para avaliar o desempenho dos classificadores propostos foi necessário primeiramente criá-lo. A coleta das informações da base de dados MEDLINE/PubMed® possibilitou a criação desse *corpus* com resumos de artigos científicos e seus respectivos papéis retóricos anotados para a língua inglesa. A partir disso, foi possível fazer uma análise das estruturas retóricas presentes no *corpus*. Foi notado que a maior parte das sentenças nos resumos coletados possuíam o papel retórico de resultados, com 32,8% do total de ocorrências. Um resultado esperado, principalmente tratando-se de resumos de artigos científicos, no qual os autores tentam enfatizar seus resultados obtidos.

Para a rede neural recorrente proposta, foi escolhida uma arquitetura baseada no trabalho de Cho et al. (2014), e foram utilizadas camadas *LSTM* em seu *encoder* e *decoder*. A implementação das características das *baselines* seguiram a proposta do trabalho de Hirohata et al. (2008) com algumas alterações. Os algoritmos de classificação escolhidos para as *baselines* foram o *SVM* e *CRF*. Ambos já foram utilizado em trabalhos anteriores para a classificação de papéis retóricos e obtiveram resultados satisfatórios.

Os resultados obtidos nos experimentos pela rede proposta foram encorajadores. Quando comparados os melhores resultados da rede com os melhores resultados das *baselines*, foi possível observar 3,18% de ganho na acurácia da classificação de estruturas

retóricas em resumos como um todo, na classificação por sentenças um ganho de 1,02% na acurácia e 0,88% na medida de F1-macro.

Apesar da similaridades com o trabalho de Hirohata et al. (2008), não foi possível realizar uma comparação direta entre os resultados dos dois trabalhos. Ambos utilizaram um *corpus* criado a partir da base de artigos científicos MEDLINE/PubMed[®], porém não é possível afirmar que ambos utilizaram de fato os mesmos resumos. Outra diferença foram os papéis retóricos utilizados. Neste trabalho foram utilizadas 5 categorias (Contexto, Objetivos, Métodos, Resultados e Conclusões) e no trabalho de Hirohata et al. (2008) foram utilizadas 4 categorias (Objetivos, Métodos, Resultados e Conclusões).

A anotação do *corpus* e a implementação das *baselines* permitiram a investigação do uso de uma rede neural recorrente para a classificação de estruturas retóricas. Os resultados obtidos no trabalho mostraram que de fato é possível usar redes neurais recorrentes para a classificação de papéis retóricos, e que, além disso, existe um potencial de que esses resultados sejam melhores que outras técnicas utilizadas anteriormente para a classificação. Com isso, pode-se concluir que os objetivos do trabalho foram alcançados.

Uma das maiores dificuldades encontradas na realização do trabalho ocorreu devido a grande quantidade de dados utilizados, o que exigiu um grande poder computacional. Apesar de ter sido utilizada uma placa de vídeo NVIDIA GeForce 980 GTX, o treinamento e avaliação da rede neural demorou cerca de 1 hora e 20 minutos por fold, portanto para rodar o primeiro experimento foi necessário cerca de 100 horas de execução. Isso acabou limitando a possibilidade de investigar a rede com outros parâmetros. As *baselines* também sofreram devido ao grande poder computacional necessário. Em um primeiro momento, foi utilizada a mesma quantidade de características do trabalho de Hirohata et al. (2008), porém o tempo de execução necessário era impraticável, o que fez com que a quantidade de características nas *baselines* fossem reduzidas para em torno de 20.000. Tanto o SVM quanto o CRF demoraram cerca de 1 hora para o treino e avaliação por fold.

6.1 Trabalhos futuros

Apesar de todos os pontos positivos, ainda existem possibilidades para aprimoramentos. Como a rede neural proposta no trabalho consegue extrair papéis retóricos de resumos de artigos científicos, uma possibilidade de trabalho futuro seria a criação de uma ferramenta para auxílio a escrita desses resumos. A arquitetura de rede também possui potencial para ser aprimorada. Uma possibilidade é realizar experimentos aumentando a quantidade de unidades presente nas camadas da LSTMs e avaliar o impacto dessa mudança na

rede. Outra proposta seria alterar a arquitetura de modo a considerar camadas LSTM bidirecionais. Embora o estudo tenha sido realizado utilizando um *corpus* em inglês, ele também pode ser utilizado para o treino e avaliação em outras línguas, pois não existe efetivamente uma dependência de língua. Portanto, seria interessante avaliar o desempenho da rede na classificação de estruturas retóricas em outros idiomas, mas para isto primeiramente seria necessário a criação de um *corpus* com uma grande quantidade de amostras.

REFERÊNCIAS

ABDEL-RAHMAN, O. et al. Radical multimodality therapy for malignant pleural mesothelioma. *The Cochrane Library*, Wiley Online Library, 2017.

ANTHONY, L.; LASHKIA, G. Mover: a machine learning tool to assist in the reading and writing of technical papers. *Professional Communication, IEEE Transactions on*, v. 46, n. 3, p. 185–193, Sept 2003. ISSN 0361-1434.

BENGIO, Y. et al. A neural probabilistic language model. *The Journal of Machine Learning Research*, JMLR. org, v. 3, p. 1137–1155, 2003.

BURSTEIN, J.; MARCU, D.; KNIGHT, K. Finding the write stuff: automatic identification of discourse structure in student essays. *Intelligent Systems, IEEE*, v. 18, n. 1, p. 32–39, Jan 2003. ISSN 1541-1672.

CAUCHY, A. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, v. 25, n. 1847, p. 536–538, 1847.

CHO, K. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

CHOLLET, F. et al. *Keras*. [S.l.]: GitHub, 2015. (<https://github.com/fchollet/keras>).

DAYRELL, C. et al. Rhetorical move detection in english abstracts: Multi-label sentence classifiers and their annotated corpora. In: CALZOLARI, N. et al. (Ed.). *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. ISBN 978-2-9517408-7-7.

DIJK, T. A. V.; KINTSCH, W. *Strategies of discourse comprehension*. [S.l.]: Academic Press New York, 1983.

FELTRIM, V. et al. Applying argumentative zoning in an automatic critiquer of academic writing. In: BAZZAN, A.; LABIDI, S. (Ed.). *Advances in Artificial Intelligence – SBIA 2004*. Springer Berlin Heidelberg, 2004, (Lecture Notes in Computer Science, v. 3171). p. 214–223. ISBN 978-3-540-23237-7. Disponível em: (http://dx.doi.org/10.1007/978-3-540-28645-5_22).

FISAS, B.; RONZANO, F.; SAGGION, H. On the discursive structure of computer graphics research papers. In: *The 9th Linguistic Annotation Workshop held in conjunction with NAACL 2015*. [S.l.: s.n.], 2015. p. 42.

- GAO, J. et al. Modeling interestingness with deep neural networks. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. [S.l.: s.n.], 2014.
- GARDNER, M. W.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, Elsevier, v. 32, n. 14, p. 2627–2636, 1998.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. [⟨http://www.deeplearningbook.org⟩](http://www.deeplearningbook.org).
- GROVER, C.; HACHEY, B.; KORYCINSKI, C. Summarising legal texts: Sentential tense and argumentative roles. In: *Proceedings of the HLT-NAACL 03 on Text Summarization Workshop - Volume 5*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (HLT-NAACL-DUC '03), p. 33–40. Disponível em: [⟨http://dx.doi.org/10.3115/1119467.1119472⟩](http://dx.doi.org/10.3115/1119467.1119472).
- GUO, Y.; KORHONEN, A.; POIBEAU, T. A weakly-supervised approach to argumentative zoning of scientific documents. In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011. p. 273–283. Disponível em: [⟨http://aclweb.org/anthology/D11-1025⟩](http://aclweb.org/anthology/D11-1025).
- HIROHATA, K. et al. Identifying sections in scientific abstracts using conditional random fields. In: *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. [s.n.], 2008. Disponível em: [⟨http://aclweb.org/anthology/I08-1050⟩](http://aclweb.org/anthology/I08-1050).
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- IRSOY, O.; CARDIE, C. Opinion mining with deep recurrent neural networks. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [S.l.: s.n.], 2014. p. 720–728.
- ISHII, R. P. *Otimização de operações de entrada e saída visando reduzir o tempo de resposta de aplicações distribuídas que manipulam grandes volumes de dados*. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2010.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. Disponível em: [⟨http://arxiv.org/abs/1412.6980⟩](http://arxiv.org/abs/1412.6980).
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- LANG, K. J.; WAIBEL, A. H.; HINTON, G. E. A time-delay neural network architecture for isolated word recognition. *Neural networks*, Elsevier, v. 3, n. 1, p. 23–43, 1990.

- LEVY, O.; GOLDBERG, Y.; RAMAT-GAN, I. Linguistic regularities in sparse and explicit word representations. In: *Conference on Natural Language Learning*. [S.l.: s.n.], 2014. p. 171–180.
- LI, Y.; GORMAN, S. L.; ELHADAD, N. Section classification in clinical notes using supervised hidden markov model. In: ACM. *Proceedings of the 1st ACM International Health Informatics Symposium*. [S.l.], 2010. p. 744–750.
- LOPER, E.; BIRD, S. Nltk: The natural language toolkit. In: *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. (ETMTNLP '02), p. 63–70. Disponível em: <https://doi.org/10.3115/1118108.1118117>.
- MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. *Journal of Machine Learning Research*, v. 9, n. Nov, p. 2579–2605, 2008.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. In: . [S.l.]: Springer, 1943. v. 5, n. 4, p. 115–133.
- MIKOLOV, T. et al. Extensions of recurrent neural network language model. In: IEEE. *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. [S.l.], 2011. p. 5528–5531.
- OKAZAKI, N. *CRFsuite: a fast implementation of Conditional Random Fields (CRFs)*. 2007. Disponível em: <http://www.chokkan.org/software/crfsuite/>.
- OLAH, C. Deep learning, nlp, and representations. *GitHub blog, posted on August*, v. 27, 2015. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. Pearson Education, Limited, 2016. (Always learning). ISBN 9781292153964. Disponível em: <https://books.google.com.br/books?id=XS9CjwEACAAJ>.
- SOCHER, R. et al. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2011. p. 801–809.
- SOCHER, R. et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: CITESEER. *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. [S.l.], 2013. v. 1631, p. 1642.
- SWALES, J. *General analysis: english in academic and research settings*. Cambridge [etc.]: Cambridge University Press, 1990. ISBN 0521338131 9780521338134.

SWALES, J. M. *Academic writing for graduate students: essential tasks and skills: a course for nonnative speakers of English*. Ann Arbor: University of Michigan Press, 1994. (English for specific purposes). ISBN 0472082639.

TEUFEL, S.; MOENS, M. Discourse-level argumentation in scientific articles: human and automatic annotation. In: CITESEER. *In Proceedings of the ACL-1999 Workshop Towards Standards and Tools for Discourse Tagging*. [S.l.], 1999.

U.S. National Library of Medicine. *Structured Abstracts in MEDLINE*. 2011.

WEISSBERG, R.; BUKER, S. *Writing up research: experimental research report writing for students of English*. Englewood Cliffs, NJ: Prentice Hall Regents, 1990. ISBN 0139708316 9780139708312 0139708499 9780139708497.

WERBOS, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, IEEE, v. 78, n. 10, p. 1550–1560, 1990.

YEPES, A. J.; MORK, J.; ARONSON, A. Using the argumentative structure of scientific literature to improve information access. In: *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, 2013. p. 102–110. Disponível em: <http://www.aclweb.org/anthology/W13-1913>.