

Campos Aleatórios Condicionais Aplicados à Detecção de Estrutura Retórica em Resumos de Textos Acadêmicos em Português

Alexandre C. Andreani¹, Valéria D. Feltrim¹

¹Departamento de Informática – Universidade Estadual de Maringá (UEM)
Av. Colombo, 5790 – Bloco C56 – CEP 87020-900 – Maringá – PR – Brasil

alexandre.c.andreani@gmail.com, valeria.feltrim@gmail.com

Abstract. *This paper presents CRF-AZPort, a rhetorical structure predictor for abstracts written in Portuguese. Unlike the AZPort classifier, which predicts the category of each sentence independently, the proposed predictor uses Conditional Random Fields for predicting a sequence of rhetorical categories for a given abstract. CRF-AZPort uses three new attributes in addition to the original AZPort attributes. Results show that rhetorical classification can be modeled as a sequence labeling problem and that considering rhetorical structure contributes to the classification.*

Resumo. *Este trabalho apresenta o CRF-AZPort, um preditor de estrutura retórica para resumos científicos escritos em português. Diferente do classificador AZPort, que prediz a categoria retórica de cada sentença de forma independente, o preditor proposto utiliza Conditional Random Fields para prever uma sequência de categorias retóricas para um resumo. Além dos atributos originais do AZPort, o CRF-AZPort utiliza três novos atributos. Os resultados obtidos mostram que a classificação retórica pode ser modelada como um problema de rotulação sequencial e que considerar a estrutura como um todo contribui para a classificação.*

1. Introdução

Cada gênero literário desperta no leitor uma expectativa do que encontrar no texto, seja um artigo científico, um livro técnico ou um romance. Para o gênero científico é importante que as informações, contidas no texto, sejam transmitidas da maneira mais direta possível. Para isso, esses trabalhos possuem uma estrutura bem estabelecida e reconhecida pelos demais leitores como adequada. Um elemento textual muito importante para artigos científicos é o resumo. O leitor usa o resumo para decidir se o trabalho é de seu interesse ou não. Além disso, o resumo é parte essencial em revisões sistemáticas e indexação em serviços eletrônicos de busca. Portanto, para a divulgação do trabalho, é importante escrever um resumo que tenha estrutura e elementos que o público-alvo espera encontrar.

Em geral, sistemas que fazem a detecção automática de estrutura retórica são construídos aplicando-se algoritmos de aprendizado de máquina para a construção de classificadores. Exemplos de classificadores retóricos são o AZ [Teufel and Moens 2002], o *E-Rater* [Burstein et al. 2003], o *Mover* [Anthony and Lashkia 2003], o AZPort

[Feltrim et al. 2006] e o AZEA [Genoves Junior 2007]. Cada um dos classificadores citados foi projetado com objetivos e domínios específicos, mas todos fazem a classificação de cada sentença do texto em categorias previstas em um modelo de estrutura retórica pré-definido, como os definidos por [Swales 1990] e [Weissberg and Buker 1990]. Assim, a abordagem mais comum é tratar a detecção de estrutura retórica como um problema de classificação.

Um problema resultante dessa abordagem é que os classificadores têm pouca ou nenhuma informação sobre a ordem das categorias previstas e da dependência entre elas no momento da previsão. É sabido que a ordem dos componentes da estrutura retórica de um texto não é aleatória. Por exemplo, no contexto de resumos científicos, sequências como Propósito-Metodologia-Resultado são mais prováveis de acontecer do que sequências como Resultado-Lacuna-Contexto. Assim, um sistema capaz de prever uma sequência de categorias retóricas, considerando assim a ordem de ocorrência das categorias, pode obter melhores resultados do que um sistema que prevê cada categoria isoladamente. De fato, trabalhos como [Liakata et al. 2012], [Merity et al. 2009], [Hirohata et al. 2008] utilizaram abordagens baseadas na predição de sequências de categorias nesse contexto com bons resultados.

Dessa forma, este trabalho apresenta um preditor de estrutura retórica para resumos científicos escritos em português baseado em predição estruturada. O preditor, chamado de CRF-AZPort, é uma nova versão do classificador AZPort [Feltrim et al. 2006] que utiliza *Conditional Random Fields* (CRF) [Lafferty 2001] para prever uma sequência de categorias. Além dos atributos do AZPort, o CRF-AZPort utiliza três novos atributos. Em comparação com o AZPort original, os resultados obtidos com o CRF-AZPort foram superiores, confirmando que existe uma relação condicional na ocorrência das categorias retóricas e que essa informação contribui para a classificação.

O restante deste trabalho está organizado da seguinte forma: trabalhos relacionados ao problema de detecção de estrutura retórica são apresentados na Seção 2. Na Seção 3 é apresentado o preditor proposto. A avaliação do preditor é apresentada na Seção 4. Por fim, as conclusões e trabalhos futuros são apresentados na Seção 5.

2. Trabalhos relacionados

Os trabalhos sobre detecção de estrutura retórica em textos científicos podem ser divididos em dois grupos: os que classificam cada sentença de forma independente e os que empregam uma abordagem estruturada, prevendo uma sequência de categorias retóricas para o texto como um todo. Entre os trabalhos que se enquadram no primeiro grupo estão os de [Kupiec et al. 1995], [Teufel and Moens 2002], [Burstein et al. 2003], [Anthony and Lashkia 2003], [Mullen et al. 2005], [Feltrim et al. 2006], [Genoves Junior 2007], [Pendar and Cotos 2008] e [Guo et al. 2013]. Entre os trabalhos que se enquadram no segundo grupo estão os de [Hirohata et al. 2008], [Merity et al. 2009] e [Liakata et al. 2012]. Visto que este trabalho trata a detecção de estrutura retórica como um problema de rotulação sequencial, apenas os trabalhos do segundo grupo são descritos nesta seção.

[Hirohata et al. 2008] trataram a detecção de estrutura retórica usando CRF para detectar quatro categorias retóricas em resumos científicos extraídos da literatura médica, a saber: Propósito, Método, Resultado e Conclusão. Os atributos utilizados pelos auto-

res podem ser divididos em três grupos: (i) conteúdo (n-gramas), que buscam identificar expressões que caracterizam cada categoria; (ii) Localização relativa da sentença, composto por atributos binários, que dividem o resumo em cinco regiões; e (iii) atributos das n ($n = 0, 1, 2$) sentenças anteriores/posteriores. Como *corpus* foram utilizados 50.000 resumos extraídos da base *Medline* já com a anotação de categorias, uma vez que a divisão de seções usada para os resumos da base foi utilizada como anotação. Um classificador SVM foi usado como *baseline*. Os resultados foram apresentados em termos da acurácia (% de acerto) por sentença e por resumos. A acurácia por sentença foi de 93,3% para o SVM e de 94,4% para o CRF. A acurácia por resumo foi de 55,5% para o SVM e 60,4% para o CRF, deixando mais evidente a vantagem do CRF e o impacto que a informação sobre a estrutura tem na classificação.

[Merity et al. 2009] propuseram uma nova versão do classificador AZ [Teufel and Moens 2002] usando um modelo de máxima entropia. Os autores usaram dois *corpora* para treinar e avaliar o modelo: um com 7.840 sentenças provenientes de artigos de astronomia (ASTRO) e outro com 12.000 sentenças provenientes de artigos de linguística computacional (CMP-LG). O esquema de anotação usado no *corpus* CMP-LG foi o proposto por [Teufel 1999], composto pelas categorias: Propósito, Estrutura, Próprio, Contexto, Contraste, Base e Outros; o *corpus* ASTRO foi anotado com uma adaptação desse esquema. Como atributos foram utilizados n-gramas, número de seções, localização da sentença, posição da sentença dentro de um parágrafo, histórico, além de um subconjunto dos atributos propostos por [Teufel 1999]. Para evitar um super ajuste do modelo, os autores aplicaram um limiar de corte para os atributos que ocorrem de maneira esparsa, como é o caso dos n-gramas. O modelo proposto obteve 96,88% de F-score para o *corpus* CMP-LG, o que corresponde a uma melhora de pelo menos 20% sobre os resultados obtidos por [Teufel 1999] com um classificador *Naïve Bayes*. Os resultados obtidos para o *corpus* ASTRO foram similares aos do *corpus* CMP-LG, evidenciando a vantagem da abordagem estruturada sobre a abordagem tradicional de classificação.

[Liakata et al. 2012] avaliaram o desempenho de classificadores SVM e CRF na detecção de estrutura retórica de artigos científicos completos. Como *corpus* foram utilizados 265 artigos na áreas de química e bioquímica, totalizando 39.915 sentenças manualmente anotadas. O esquema usado na anotação era composto por 11 categorias, a saber: Hipótese, Motivação, Propósito, Objeto, Contexto, Método, Experimento, Modelo, Observação, Resultado e Conclusão. Como atributos foram utilizados 16 atributos binários que buscam capturar informações a respeito do tamanho e localização das sentenças, da seções do texto, das citações, do histórico, do verbo principal, da estrutura sintática, da voz e do conteúdo (n-gramas). Diferentes classificadores foram induzidos usando SVM e CRF. O desempenho da classificação para os artigos completos foi em torno de 50%, não tendo sido observada diferença significativa entre os classificadores SVM e CRF quando todos os atributos foram utilizados. Vale destacar que o esquema de anotação usado por [Liakata et al. 2012] é mais refinado do que o usado por [Teufel 1999] e [Merity et al. 2009]. Em geral, esquemas de anotação menores simplificam a tarefa de anotação manual, o que contribui para um melhor desempenho na classificação.

3. Preditor de Estrutura Retórica para Resumos

Assim como os trabalhos descritos na seção anterior, o preditor de estrutura retórica proposto neste trabalho trata a detecção de estrutura de estrutura retórica como um problema de rotulação sequencial. A motivação para o uso de tal abordagem vem do fato da estrutura retórica ser composta por uma sequência de movimentos articulados de modo a se obter o efeito esperado no leitor, que tende a apresentar padrões específicos do gênero textual, e não de um conjunto de movimentos aleatórios. No caso do gênero científico, esses padrões são mais evidentes, o que motivou propostas de diferentes modelos de estrutura retórica, tanto para seções específicas do texto, como resumos e introduções [Swales 1990], [Weissberg and Buker 1990], como para artigos completos [Teufel and Moens 2002], [Liakata et al. 2012].

O preditor proposto, chamado de CRF-AZPort, foi criado no mesmo contexto do classificador AZPort [Feltrim 2004]. O AZPort é um classificador *Naïve Bayes* que estima a probabilidade de uma sentença S ter a categoria C , dados os valores dos atributos extraídos de S . Os oito atributos utilizados pelo AZPort foram adaptados do conjunto de atributos propostos para o AZ e são determinados automaticamente a partir do texto de entrada. O treinamento e teste do AZPort foi feito com 52 resumos do CorpusDT [Feltrim et al. 2003], totalizando 366 sentenças. Cada resumo foi manualmente anotado segundo um esquema pré-definido de sete categorias, a saber: Contexto (B); Lacuna (L); Propósito (P); Metodologia (M); Resultado (R); Conclusão (C); e Estrutura (E).

Enquanto o AZPort prevê a categoria de cada sentença de forma independente, o CRF-AZPort prevê a melhor sequência de categorias dadas as sentenças de um resumo. Isso é feito por meio de um classificador CRF, que faz a previsão de cada categoria de maneira condicional às categorias da sequência completa.

3.1. *Conditional Random Fields*

Conditional Random Fields (CRF) é um método probabilístico que tem sido amplamente aplicado no Processamento de Linguagem Natural. Proposto por [Lafferty 2001], o método é usado em predição estruturada por permitir considerar amostras vizinhas e fazer predições interdependentes. CRF pode ser pensado como um modelo de estados finitos com transições não normalizadas [McCallum et al. 2000].

Uma forma especial de *Conditional Random Fields* é a cadeia linear que modela as variáveis de saída, neste caso, as categorias, como uma sequência. Considerando que para um resumo com sentenças $x = (x_1, \dots, x_n)$ é desejado determinar uma sequência ótima de categorias $y = (y_1, \dots, y_n)$ de todas as possíveis sequências, *Conditional Random Fields* usa a probabilidade condicional conforme a Equação (1).

$$p(y|x) = \frac{1}{Z_\lambda(x)} \exp(\lambda \cdot F(y, x)) \quad (1)$$

A função $F(y, x)$ representa um vetor global de atributos para a sequência de entrada x e uma sequência de saída y , como mostra a Equação (2).

$$F(y, x) = \sum_i f(y, x, i) \quad (2)$$

A variável i varia sobre a sequência de entradas, ou seja, a função $f(y, x, i)$ é um vetor de atributos para a sequência de entradas x e a sequência de saída y na posição i . Na Equação (2), λ é um vetor no qual um elemento λ_k representa o peso do atributo $F_k(y, x)$ e $Z_\lambda(x)$ é o fator de normalização, que é calculado pela Equação (3). A sequência de maior probabilidade, dada pela Equação (4), é obtida aplicando-se o algoritmo de Viterbi [Forney 1973].

$$Z_\lambda(x) = \sum_y \exp(\lambda \cdot F(y, x)) \quad (3)$$

$$\hat{y} = \arg \max_y p(y|x) \quad (4)$$

3.2. Atributos

Além dos atributos usados pelo AZPort, o CRF-AZPort utiliza três novos atributos. Os oito atributos provenientes do AZPort são apresentados na Tabela 1 e os novos atributos propostos para o CRF-AZPort são mostrados na Tabela 2. Assim, ao todo, foram implementados 11 atributos para o CRF-AZPort.

Tabela 1. Atributos usados pelo AZPort [Feltrim 2004]

Nome	Descrição	Valores possíveis
Tamanho	Tamanho da sentença em comparação aos dois limiares 20 e 40 palavras	curta, média ou longa
Localização	Posição da sentença	primeira, segunda, mediana, penúltima ou última
Citação	A sentença contém citações?	sim ou não
Expressão	Tipo de expressão padrão observado na sentença	C, L, P, M, R, Co(conclusão) ou noexpr(sem ocorrência)
Tempo	Tempo do primeiro verbo finito da sentença	IMP, PRES, PAST, FUT, PRES-CPO, PAST-CPO, FUT-CPO, PRES-CT, PAST-CT, FUT-CT, PRES-CPO-CT, PAST-CPO-CT, FUT-CPO-CT ou <i>noverb</i> (sem verbo)
Voz	Voz do primeiro verbo finito da sentença	passiva, ativa ou <i>noverb</i> (sem verbo)
Modal	O primeiro verbo finito da sentença é modal?	sim, não ou <i>noverb</i> (sem verbo)
Histórico	Categoria da sentença anterior	..., C, L, P, M, R, Co ou E

Os atributos da Tabela 2 foram inspirados no trabalho de [Hirohata et al. 2008]. O atributo *Janela deslizante* foi implementado usando os mesmos valores previstos por [Hirohata et al. 2008]. Já para o atributo *Segmentação* foi proposto um conjunto próprio de valores possíveis que ajudam a mapear as sequências de categorias com base em uma marcação de início, meio e fim. Os valores possíveis são:

- I(Início): ocorre quando a sentença anterior possui um movimento retórico diferente e a sentença posterior tem o mesmo movimento retórico da atual;
- IF(Início-Fim): ocorre quando as sentenças anterior e posterior possuem movimentos retóricos diferentes;

- M(Meio): ocorre quando as sentenças anterior e posterior possuem o mesmo movimento retórico que a sentença atual;
- MF(Meio-Fim): ocorre quando a sentença anterior possui o mesmo movimento retórico da sentença atual, mas a sentença posterior tem um movimento retórico diferente.

O atributo *Classe por frequência de n-gramas* utiliza 2-gramas, 3-gramas e 4-gramas e medidas de relevância (TF-IDF, χ^2 e k-vizinhos mais próximos) para estimar uma categoria para sentença entre as sete categorias retóricas possíveis.

Experimentos preliminares mostraram que alguns dos atributos utilizados prejudicam o desempenho do CRF com *corpora* pequenos. Esse foi o caso do atributo *Citação*, que diminuiu o desempenho do preditor treinado com o corpus de 366 sentenças usado pelo AZPort. Com base nessas observações, foi definido um limiar $\delta = 400$ para selecionar o conjunto de atributos de acordo com o tamanho do *corpus* utilizado na etapa de treinamento. Dessa maneira, os atributos *Citação*, *Classe por frequência de n-gramas* e *Segmentação* não são utilizados quando o número de sentenças do conjunto de treinamento é menor que δ . Se o número de sentenças disponível para o treinamento é maior que δ , então todos os atributos são utilizados.

Tabela 2. Atributos exclusivos do CRF-AZPort

Nome	Descrição	Valores possíveis
Classe por frequência de <i>n</i> -gramas	Classe de acordo com TF-IDF, χ^2 e k-vizinhos mais próximos	Um dos valores possíveis de classe
Segmentação	A sentença atual continua o movimento retórico da sentença anterior?	I(Início), IF(Início-Fim), M=(Meio), MF(Meio-Fim)
Janela deslizante	Atributos dos <i>k</i> elementos vizinhos	$k = \{0, 1, 2\}$

3.3. Corpora

Para o treinamento e teste do CRF-AZPort foram utilizados dois *corpora* de resumos¹ escritos em português extraídos a partir de dissertação e teses em Computação. Ambos tiveram as sentenças anotadas manualmente por três anotadores treinados e com experiência em escrita científica. O primeiro, que chamaremos de Corpus366, é o mesmo *corpus* utilizado no treinamento e teste do AZPort original, sendo composto por 52 resumos e totalizando 366 sentenças. O segundo, que chamaremos de Corpus466, também é composto por 52 resumos, que totalizam 466 sentenças. O valor da medida *kappa* calculado para os três anotadores que participaram da anotação manual sobre um conjunto de 320 sentenças do Corpus366 e 455 sentenças do Corpus466 foi, em ambos os casos, de $K = 0,695$. Todas as sentenças dos *corpora* foram utilizadas no treinamento e teste.

3.4. Avaliação

Para permitir a comparação de resultados foram realizados experimentos com o CRF-AZPort e com o AZPort. Nas avaliações do AZPort foi utilizada a implementação de [Feltrim 2004]. As avaliações do CRF-AZPort foram feitas com a ferramenta CRFSuite² [Okazaki 2007].

¹Os corpora e demais programas utilizados neste trabalho podem ser obtidos no endereço acandreami.info/rmd

²<http://www.chokkan.org/software/crfsuite/>

Os resultados experimentais foram calculados a partir de 30 execuções de validação cruzada de *13-fold*. Em cada execução, os resumos foram aleatoriamente distribuídos em 13 *folds*, sendo 12 *folds* usados no treinamento e 1 no teste.

4. Resultados Experimentais

Os resultados das avaliações do AZPort com o Corpus366 e com o Corpus466, em termos das métricas de Precisão, Cobertura e *F1-score* são apresentados, respectivamente, na Tabela 3 e na Tabela 4. Os valores médios foram obtidos por meio de média ponderada dos valores observados para cada categoria. A coluna Suporte corresponde ao total de sentenças da categoria considerando 30 avaliações.

Em ambos os *corpora*, os melhores resultados foram observados para a categoria Propósito (P) e o pior resultado para a categoria Estrutura (E). Com exceção da categoria Contexto (B), os resultados obtidos com o Corpus366 foram superiores aos obtidos com o Corpus466. Isso pode ser atribuído ao fato do Corpus366 ter sido utilizado como *corpus* de desenvolvimento para o AZPort (por exemplo, na construção das expressões regulares utilizadas pelo atributo Expressão). Outro fator a ser observado é que as distribuições de categorias observadas nos dois *corpora* são diferentes.

Tabela 3. AZPort com o Corpus366

Categoria	Precisão	Cobertura	F1-score	Suporte
C	47,02%	34,17%	39,58%	600
B	68,20%	74,55%	71,23%	2310
L	77,48%	63,70%	69,92%	1080
M	80,63%	56,74%	66,61%	1350
E	0,00%	0,00%	0,00%	180
P	86,05%	69,59%	76,95%	1950
R	62,57%	81,57%	70,81%	3510
Média	69,73%	69,22%	68,51%	-

Tabela 4. AZPort com o Corpus466

Categoria	Precisão	Cobertura	F1-score	Suporte
C	37,58%	22,96%	28,51%	540
B	69,51%	84,93%	76,45%	5340
L	60,47%	45,46%	51,90%	1080
M	38,52%	27,70%	32,23%	1350
E	0,00%	0,00%	0,00%	120
P	86,29%	63,53%	73,18%	2070
R	49,18%	52,84%	50,95%	3480
Média	61,41%	62,07%	60,94%	-

Os resultados das avaliações do CRF-AZPort com o Corpus366 e com o Corpus466 são apresentados, respectivamente, na Tabela 5 e na Tabela 6. Considerando os valores médios observados, o CRF-AZPort teve desempenho superior ao AZPort para ambos os *corpora*. A maior diferença de desempenho foi obtida com o Corpus466 (5,84%), sugerindo que a superioridade do CRF-AZPort fica mais evidente quando os resumos são maiores, conseqüentemente correspondendo à sequências maiores de categorias. De fato, o número médio de sentenças por resumo no Corpus466 é de 8,96 (desvio padrão 4,89), enquanto no Corpus366 é de 7,04 (desvio padrão de 2,96).

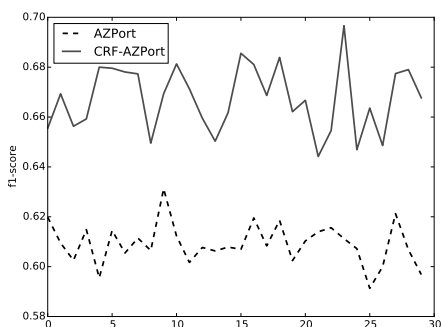


Figura 1. F1-score para as execuções do AZPort e CRF-AZPort com o Corpus466

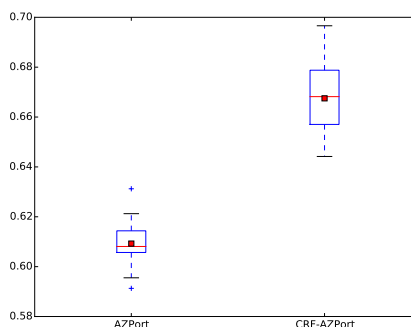


Figura 2. Diagrama de caixas do F1-score do AZPort e CRF-AZPort com o Corpus466

A Figura 1 mostra a comparação das medidas F1-score obtidas para as 30 execuções do AZPort e do CRF-AZPort com o Corpus466. A Figura 2 mostra o diagrama de caixas correspondente, que evidencia a relação entre as médias, medianas e desvio padrão. Conforme pode ser observado, o CRF-AZPort foi superior em todas as execuções.

A significância dos resultados foi avaliada utilizando o T-test. Considerando um nível de significância (α -value) de 1%, é possível afirmar que o CRF-AZPort teve um desempenho superior ao AZPort para as medidas F1-score com os dois corpora usados nas avaliações.

Tabela 5. CRF-AZPort com o Corpus366

Categoria	Precisão	Cobertura	F1-score	Suporte
C	48,07%	18,67%	26,89%	600
B	81,99%	78,05%	79,97%	2310
L	75,89%	69,07%	72,32%	1080
M	81,18%	55,93%	66,23%	1350
E	0,00%	0,00%	0,00%	180
P	84,53%	78,77%	81,55%	1950
R	63,14%	86,67%	73,05%	3510
Média	72,51%	72,80%	71,38%	-

Tabela 6. CRF-AZPort com o Corpus466

Rótulos	Precisão	Cobertura	F1-score	Suporte
C	37,58%	22,96%	28,51%	540
B	69,51%	84,93%	76,45%	5340
L	60,47%	45,46%	51,90%	1080
M	38,52%	27,70%	32,23%	1350
E	0,00%	0,00%	0,00%	120
P	86,29%	63,53%	73,18%	2070
R	49,18%	52,84%	50,95%	3480
Média	66,46%	68,74%	66,78%	-

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou o CRF-AZPort, um preditor de estrutura retórica para resumos científicos escritos em português baseado em CRF, proposto para uso no mesmo contexto do classificador AZPort.

Os resultados obtidos na avaliação do CRF-AZPort com dois corpora de resumos científicos foram superiores aos obtidos com o AZPort, especialmente quando o corpus de treinamento possui resumos com um número maior de sentenças. Isso mostra que existe uma relação condicional entre as categorias retóricas e que a contribuição dessa informação para a classificação retórica fica mais evidente quando as sequências utilizadas no treinamento são maiores.

Entre os trabalhos futuros está prevista a coleta e a anotação de um novo *corpus* de resumos científicos com objetivo de aumentar o tamanho do conjunto de treinamento disponível. Também está prevista a investigação de novos atributos, bem como experimentos para a seleção de atributos, visando melhorar o desempenho do preditor, mesmo para *corpora* compostos por resumos menores.

Agradecimentos

A Capes pelo apoio financeiro.

Referências

- Anthony, L. and Lashkia, G. (2003). Mover: A machine learning tool to assist in the reading and writing of technical papers. *IEEE Transactions on Professional Communication*, 46(3):185–193.
- Burstein, J., Chodorow, M., and Leacock, C. (2003). Criterion online essay evaluation: An application for automated evaluation of student essays. In *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*.
- Feltrim, V. D. (2004). *Uma abordagem baseada em córpus e em sistemas de crítica para a construção de ambientes Web de auxílio à escrita acadêmica em português*. Tese de doutorado, Universidade de São Paulo.
- Feltrim, V. D., Aluísio, S. M., and Nunes, M. d. G. V. (2003). Analysis of the rhetorical structure of computer science abstracts in portuguese. In *Proceedings of Corpus Linguistics*, volume 16, pages 212–218.
- Feltrim, V. D., Teufel, S., Nunes, M. G. V. d., and Aluísio, S. M. (2006). Argumentative zoning applied to critiquing novices’ scientific abstracts. In Shanahan, J. G., Qu, Y., and Wiebe, J., editors, *Computing Attitude and Affect in Text: Theory and Applications*, number 20 in The Information Retrieval Series, pages 233–246. Springer Netherlands.
- Forney, Jr., G. (1973). The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278.
- Genoves Junior, L. C. (2007). *Avaliação automática da qualidade de escrita de resumos científicos em inglês*. Dissertação de mestrado, Universidade de São Paulo.
- Guo, Y., Silins, I., Stenius, U., and Korhonen, A. (2013). Active learning-based information structure analysis of full scientific articles and two applications for biomedical literature review. *Bioinformatics*, 29(11):1440–1447.
- Hirohata, K., Okazaki, N., Ananiadou, S., Ishizuka, M., and Biocentre, M. I. (2008). Identifying sections in scientific abstracts using conditional random fields. In *IJCNLP*, pages 381–388.
- Kupiec, J., Pedersen, J., and Chen, F. (1995). A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’95*, pages 68–73, New York, NY, USA. ACM.
- Lafferty, J. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann.

- Liakata, M., Saha, S., Dobnik, S., Batchelor, C., and Rebholz-Schuhmann, D. (2012). Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, 28(7):991–1000.
- McCallum, A., Freitag, D., and Pereira, F. C. N. (2000). Maximum entropy markov models for information extraction and segmentation. In *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, pages 591–598, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Merity, S., Murphy, T., and Curran, J. R. (2009). Accurate argumentative zoning with maximum entropy models. In *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, pages 19–26. Association for Computational Linguistics.
- Mullen, T., Mizuta, Y., and Collier, N. (2005). A baseline feature set for learning rhetorical zones using full articles in the biomedical domain. *SIGKDD Explor. Newsl.*, 7(1):52–58.
- Okazaki, N. (2007). Crfsuite: a fast implementation of conditional random fields (crfs).
- Pendar, N. and Cotos, E. (2008). Automatic identification of discourse moves in scientific article introductions. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 62–70. Association for Computational Linguistics.
- Swales, J. (1990). *General analysis: english in academic and research settings*. Cambridge University Press, Cambridge [etc.].
- Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text*. PhD thesis, University of Edinburgh.
- Teufel, S. and Moens, M. (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Comput. Linguist.*, 28(4):409–445.
- Weissberg, R. and Buker, S. (1990). *Writing up research: experimental research report writing for students of English*. Prentice Hall Regents, Englewood Cliffs, NJ.