7
Experimentos com Tarefas de PLN

Com o objetivo de mostrar a qualidade dos classificadores gerados através das abordagens BAS apresentadas, conduzimos experimentos com duas classes de problemas de Processamento de Linguagem Natural comumente empregadas em tarefas de recuperação de informação: anotação morfossintática e anotação de sintagmas.

Neste capítulo, mostramos os resultados do algoritmo Comitê BAS para tais tarefas. Também são reportados os resultados de outros algoritmos de estado-da-arte para os problemas.

7.1
Medidas de Qualidade

Uma medida de qualidade comumente utilizada em tarefas de PLN é a acurácia por palavra. A acurácia é definida formalmente como

$$\text{acurácia} = \frac{|\{\text{palavras corretamente classificadas}\}|}{|\{\text{palavras}\}|}$$

Outras medidas importantes em tarefas de recuperação de informação são a precisão e a abrangência.

A precisão pode ser encarada como uma medida de exatidão ou fidelidade e é definida como a quantidade de verdadeiros positivos (VP), itens corretamente classificados, dividida pela quantidade total de itens classificados como positivos, que é a soma dos verdadeiros positivos e dos falsos positivos (FP), itens incorretamente classificados.

Logo, a precisão é definida como

$$\text{precisão} = \frac{VP}{VP + FP}$$

Por outro lado, a abrangência pode ser encarada como uma medida de plenitude e é definida como a quantidade de verdadeiros positivos (VP) dividida pela quantidade total de itens que deveriam ser classificados como positivos, que é a soma dos verdadeiros positivos e dos falsos negativos (FN), itens não classificados.

Logo, a abrangência é definida como

$$\text{abrangência} = \frac{VP}{VP + FN}$$

Frequentemente, existe uma relação entre a precisão e a abrangência, de tal forma que é possível melhorar uma medida em detrimento da outra. Em consequência disso, tais medidas não são discutidas isoladamente. Elas são combinadas em uma medida única, a medida F_β, que é a média harmônica ponderada da precisão e da abrangência e é definida como

$$F_\beta = (1 + \beta^2) \times \frac{\text{precisão} \times \text{abrangência}}{\beta^2 \times \text{precisão} + \text{abrangência}}$$

O parâmetro β pondera a importância relativa que pode ser dada para a precisão em relação a abrangência. Comumente, o parâmetro β é ajustado para 1, o que informa que a precisão possui a mesma importância que a abrangência, definindo assim a medida-F (F_1) como

$$F_1 = 2 \times \frac{\text{precisão} \times \text{abrangência}}{\text{precisão} + \text{abrangência}}$$

7.2 Modelagem BAS para PLN

Boosting é uma abordagem de aprendizado de máquina muito utilizada em tarefas de classificação de dados, como as descritas no Capítulo 6. Entretanto, o Boosting não é largamente utilizado em tarefas de PLN.

Em todos os experimentos de PLN mostrados neste Capítulo, utilizamos a mesma configuração de parâmetros de Boosting. Tais parâmetros foram obtidos empiricamente através de uma validação cruzada nos menores conjuntos. Já os parâmetros para o algoritmo-base ETL, utilizado pelo Boosting, são os mesmos indicados em (San09) para cada tarefa.

A seguir, os diversos parâmetros utilizados são listados.

Classificador Base: foram realizados experimentos com dois novos algoritmos de aprendizado de máquina baseados em TBL, o algoritmo TBL Genético, descrito na Seção 5.3 e o algoritmo ETL, descrito na Seção 5.4.

Número de Classificadores-Base: foram treinados comitês com um total de cem classificadores-base, a mesma quantidade aplicada no algoritmo Comitê ETL (San09), e também com um total de quatrocentos classificadores-base. Este último tem por objetivo mostrar que pode-se

melhorar um pouco mais o desempenho aumentando-se a quantidade de membros do comitê.

Tamanho do Conjunto de Validação: devido à grande quantidade de exemplos e de forma a poupar tempo de treinamento optamos por não utilizar um conjunto de validação.

Algoritmo de agrupamento: devido à sua simplicidade e bons resultados, foi aplicado o algoritmo K-Médias com a distância TD-IDF. Como uma grande quantidade de classificadores BAS são treinados nestes experimentos, é necessária uma grande diversidade para os pesos utilizados. Por isso, optamos por gerar dez grupos de exemplos e utilizar o método de ponderação híbrido.

Esquema de Votação: como não foi empregado um conjunto de validação, o esquema de votação adotado é uma simples votação pela maioria entre todos os classificadores BAS treinados.

Agora, apresentamos os parâmetros empregados no treinamento dos algoritmos-base utilizados.

Tamanho da Janela de Contexto: utilizamos uma janela de tamanho sete. Assim, permitimos gerar gabaritos com termos atômicos que fazem referência à palavra corrente e às suas seis palavras vizinhas, sendo três delas para a esquerda e três para a direita.

Tamanho máximo do gabarito: empregamos gabaritos cujo tamanho máximo é de seis termos atômicos.

Atributos com alta dimensionalidade: no caso do atributo palavra, somente os 200 valores mais significativos são utilizados.

Limiar de pontuação para as regras: os algoritmos aprendem regras cuja pontuação é pelo menos dois.

Limiar de pontuação para regras redundantes: os algoritmos aprendem regras redundantes com qualquer pontuação.

Quantidade de gabaritos: nas abordagens que utilizam comitê, cada classificador TBL emprega uma quantidade máxima de cinquenta gabaritos em seu treinamento. Logo, caso um número maior do que esse seja encontrado, um subconjunto aleatório desse tamanho é selecionado.

Parâmetros Genéticos: No caso especial do algoritmo TBL Genético, os valores enumerados a seguir são os adotados para os parâmetros.

Tamanho do conjunto de treinamento: 150.

Tamanho da população: 20.
Número de Gerações: 10.
Percentual de mutação: 10%.
Percentual de cruzamento: 80%.

7.3 Corpora de Anotação Morfossintática

Anotação morfossintática é a tarefa de PLN que consiste na determinação da classe gramatical de uma palavra e da função que ela exerce em determinado contexto. Na tabela 7.1 mostramos um exemplo de anotação morfossintática de uma sentença em Português.

Tabela 7.1: Exemplo de anotação morfossintática do Português.

Palavra	**Anotação**	**Função**
Mantinha	V	Verbo
a	ART	Artigo
tabela	N	Substantivo
progressiva	ADJ	Adjetivo
,	,	Vírgula
mas	KC	Conjunção Coordenativa
reduzia	V	Verbo
suas	PROADJ	Pronome Possessivo
escalas	N	Substantivo
pela	PREP+ART	Preposição e Artigo
metade	N	Substantivo
.	.	Ponto

O algoritmo Comitê BAS foi aplicado em quatro corpora com anotações morfossintáticas. Dois desses corpora são do idioma Português, o Mac-Morpho (Alu03), um corpus da língua Portuguesa do Brasil atual com textos extraídos de jornais; e o Tycho Brahe (Tyc08), um corpus histórico do Português com textos escritos por autores nascidos entre 1380 e 1845. Os outros dois corpora, possuem anotações morfossintáticas para os idiomas Inglês, o corpus Brown (Fra82), e Alemão, o corpus Tiger (Bra02).

A Tabela 7.2 mostra as principais características dos corpora utilizados nesta tarefa.

7.4 Corpora de Anotação de Sintagmas

Anotação de Sintagmas é a tarefa de PLN que separa e segmenta sentenças em seus subconstituintes, tais como sintagmas nominais, verbais e preposicionais. Assim como na anotação morfossintática, os sintagmas são

Tabela 7.2: Corpora de anotação morfossintática.

Corpus	**Idioma**	**Número de Classes**	**Treinamento**		**Teste**	
			Sentenças	**Palavras**	**Sentenças**	**Palavras**
Mac-Morpho	Português	22	44.233	1.007.671	9.141	213.794
Tycho Brahe	Português	383	30.698	775.601	10.234	259.991
Tiger	Alemão	54	41.954	742.189	8.520	146.389
Brown	Inglês	182	47.027	950.975	10.313	210.217

atributos chaves em tarefas mais complexas de recuperação de informação, como reconhecimento de entidades e extração de relações. Na tabela 7.3 mostramos um exemplo de anotação de sintagmas nominais para uma sentença em Português.

Tabela 7.3: Exemplo de anotação de sintagmas do Português.

Palavra	**Anotação**	**Marcação de Sintagma**
Somente	O	
em	O	
algumas	I	Início
localidades	I	
do	I	
Vale	I	
do	I	
Paranapanema	I	Final
,	O	
sudoeste	I	Início e Final
e	O	
litoral	I	Início e Final
a	B	Início
reserva	I	
hídrica	I	Final
é	O	
razoável	O	
.	O	

De forma a transformar o problema de classificação de sintagmas em um problema de anotação de palavras, adotamos um esquema de marcação denominado IOB1 (Ram95). No padrão IOB1, a marcação *O* significa que a palavra não pertence a um sintagma, a marcação *I* significa que a palavra pertence a um sintagma nominal e marcação *B* é utilizada sempre que dois sintagmas forem adjacentes.

De forma a aplicar o Comitê BAS nesta tarefa, selecionamos quatro corpora com diferentes sintagmas de diferentes idiomas. O primeiro corpus é o SNR-CLIC (Fre05), que possui marcações de sintagmas nominais do Português do Brasil. O segundo corpus é o Ramshaw & Marcus (Ram95), que também possui anotações de sintagmas nominais, porém, do Inglês. Outro corpus com

sintagmas do Inglês utilizado, derivado do corpus anterior, foi o aplicado na competição CoNLL-2000 (Tjo00). Tal corpus possui anotações para todos os tipos de sintagmas e não somente para os sintagmas nominais. Por último, também foi utilizado o corpus SPSAL-2007 (Bha07), contendo todos os tipos de sintagmas do Hindi.

A Tabela 7.4 mostra as principais características dos corpora utilizados nesta tarefa.

Tabela 7.4: Corpora de anotação de sintagmas.

Corpus	**Idioma**	**Treinamento**		**Teste**	
		Sentenças	**Palavras**	**Sentenças**	**Palavras**
SNR-CLIC	Português	3.514	83.346	878	20.798
Ramshaw & Marcus	Inglês	8.936	211.727	2.012	47.377
CoNLL-2000	Inglês	8.936	211.727	2.012	47.377
SPSAL-2007	Hindi	924	20.000	210	5.000

7.5 Software e Hardware

Todo o código-fonte envolvido na experimentação foi desenvolvido no Laboratório de Engenharia de Algoritmos e Redes Neurais (LEARN) da PUC-Rio, com exceção do algoritmo C4.5 utilizado como árvore de decisão para o algoritmo ETL. Tal implementação foi adaptada, a partir de uma implementação de domínio público (Qui86), com o objetivo de permitir a utilização de pesos. A linguagem utilizada para codificação dos algoritmos é o Python 2.4.

Os experimentos com tarefas de PLN foram executados em um *cluster* de 15 máquinas, cada uma com um processador Intel® de 3.4 GHz e 2 GB de memória RAM.

7.6 Resultados

Nesta seção, apresentamos os resultados dos experimentos conduzidos utilizando tarefas de processamento de linguagem natural.

7.6.1 Experimentos com Anotação Morfossintática

Realizamos experimentos com quatro corpora de anotação morfossintática, Mac-Morpho e Tycho Brahe (Português), Brown (Inglês) e Tiger (Alemão). Para cada corpus, os algoritmos enumerados a seguir são comparados.

Classificador Inicial: um algoritmo simples, utilizado como algoritmo-base para as abordagens baseadas em TBL, que consiste em duas etapas. Na primeira etapa o algoritmo assinala para cada palavra a etiqueta morfossintática mais frequente no conjunto de treinamento. Na fase de classificação, caso uma palavra não tenha sido treinada, uma etiqueta padrão é assinalada. Na segunda etapa, é aplicado um classificador fixo TBL proposto por Brill (Bri95). Este algoritmo assinala etiquetas baseado em regras derivadas de sufixos ou prefixos e de palavras antecessoras ou sucessoras. Por exemplo, se a palavra termina em "mente" então atribua a etiqueta *advérbio*,, Ou ainda, se a palavra anterior for "que" ou "se" então atribua a etiqueta "verbo".

TBL GEN: o algoritmo TBL Genético, que utiliza os atributos palavra e anotação morfossintática.

ETL: o algoritmo ETL, que utiliza os atributos palavra e anotação morfossintática, com a configuração descrita na Seção 7.2.

Comitê ETL: um comitê Bagging de cem algoritmos ETL.

Comitê BAS TBL GEN: um comitê de cem algoritmos TBL Genético, utilizando a abordagem Comitê BAS.

Comitê BAS ETL 100: um comitê de cem algoritmos ETL, utilizando a abordagem Comitê BAS.

Comitê BAS ETL 400: o mesmo comitê porém formado por quatrocentos algoritmos ETL.

Estado-da-Arte: para cada corpus, apresentamos o algoritmo com melhor desempenho até o momento.

Na Tabela 7.5, apresentamos os resultados para o Corpus Mac-Morpho. O sistema com melhor desempenho até o momento, para tal tarefa, é baseado na abordagem Comitê ETL (San09).

Tabela 7.5: Desempenho dos algoritmos na tarefa de anotação morfossintática do Português.

Algoritmo	**Acurácia**
Comitê BAS ETL 400	96.98
Comitê BAS ETL 100	96.96
Comitê ETL	96.94
ETL	96.75
Comitê BAS TBL GEN	96.56
TBL GEN	96.45
Classificador Inicial	91.66

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 59%, de 8.34 para 3.44 e 64% (3.04) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 64% (3.02). Já o algoritmo Comitê ETL reduz o erro em 63% (3.06), sendo a redução dos erros entre os dois comitês de 1%. Consequentemente, para esta tarefa, o algoritmo Comitê BAS apresenta o melhor desempenho.

Na Tabela 7.6, apresentamos os resultados para o Corpus Tycho Brahe. O sistema com melhor desempenho até o momento, para tal tarefa, também é baseado na abordagem Comitê ETL (San09).

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 43%, de 7.00 para 3.97 e 53% (3.31) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro cai para 52% (3.34). Por outro lado, o algoritmo Comitê ETL reduz o erro em 53% (3.28). Para essa tarefa, apesar da redução do erro ter sido maior quando comparado com o algoritmo ETL (52%), houve uma queda de qualidade quando comparado o Comitê ETL.

Tabela 7.6: Desempenho dos algoritmos na tarefa de anotação morfossintática do Português Histórico.

Algoritmo	**Acurácia**
Comitê ETL	96.72
Comitê BAS ETL 100	96.69
Comitê BAS ETL 400	96.66
ETL	96.64
Comitê BAS TBL GEN	96.03
TBL GEN	95.45
Classificador Inicial	93.00

Na Tabela 7.7, apresentamos os resultados para o Corpus Brown. O sistema com melhor desempenho até o momento, para tal tarefa, também é baseado na abordagem Comitê ETL (San09).

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 55%, de 7.57 para 3.40 e 59% (3.12) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 59% (3.11). Já o algoritmo Comitê ETL reduz o erro em 58% (3.17), sendo a redução dos erros entre os dois comitês de 2%. Consequentemente, para esta tarefa, o algoritmo Comitê BAS apresenta o melhor desempenho.

Na Tabela 7.8, apresentamos os resultados para o Corpus Tiger. O sistema com melhor desempenho até o momento, para tal tarefa, também é

Tabela 7.7: Desempenho dos algoritmos na tarefa de anotação morfossintática do Inglês.

Algoritmo	**Acurácia**
Comitê BAS ETL 400	96.89
Comitê BAS ETL 100	96.88
Comitê ETL	96.83
ETL	96.69
Comitê BAS TBL GEN	96.60
TBL GEN	96.30
Classificador Inicial	92.43

baseado na abordagem Comitê ETL (San09).

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 34%, de 6.69 para 4.41 e 51% (3.31) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 52% (3.24). O algoritmo Comitê ETL reduz o erro em 50% (3.32), sendo a redução dos erros entre os dois comitês de 2%. Consequentemente, para esta tarefa, o algoritmo Comitê BAS apresenta o melhor desempenho.

Tabela 7.8: Desempenho dos algoritmos na tarefa de anotação morfossintática do Alemão.

Algoritmo	**Acurácia**
Comitê BAS ETL 400	96.76
Comitê BAS ETL 100	96.69
Comitê ETL	96.68
ETL	96.57
Comitê BAS TBL GEN	96.15
TBL GEN	95.59
Classificador Inicial	93.31

7.6.2 Experimentos com Anotação de Sintagmas

Realizamos experimentos com quatro corpora de anotação de sintagmas, SNR-CLIC (Português), Ramshaw & Marcus e CONLL-2000 (Inglês) e SPSAL-2007 (Hindi). Para cada corpus, os seguintes algoritmos são comparados.

Classificador Inicial: um algoritmo simples utilizado como algoritmo-base para as abordagens baseadas em TBL. Este algoritmo consiste em assinalar para cada etiqueta morfossintática a etiqueta de anotação de

sintagmas mais frequente no conjunto de treinamento. Para o caso de anotação de sintagmas do Português, o algoritmo funciona de uma maneira ligeiramente diferente, considerando individualmente cada palavra que esteja marcada como preposição.

TBL GEN: o algoritmo TBL Genético, que utiliza os atributos palavra, anotação morfossintática e anotação de sintagmas, com a configuração descrita na Seção 7.2.

ETL: o algoritmo ETL, que utiliza os atributos palavra, anotação morfossintática, anotação de sintagmas. Também são utilizados outros dois atributos, *verbo a esquerda*, que indica o verbo predecessor mais próximo e *estrutura*, que possui informação sobre as letras da palavra, primeira maiúscula, todas maiúsculas, nenhuma maiúscula, número, número com "" ou "-", pontuação e outros.

Comitê ETL: um comitê Bagging de cem algoritmos ETL.

Comitê BAS TBL GEN: um comitê de cem algoritmos TBL Genético, utilizando a abordagem Comitê BAS.

Comitê BAS ETL 100: um comitê de cem algoritmos ETL, utilizando a abordagem Comitê BAS.

Comitê BAS ETL 400: o mesmo comitê porém formado por quatrocentos algoritmos ETL.

Estado-da-Arte: para cada corpus, é apresentado o algoritmo com melhor desempenho até o momento.

Na Tabela 7.9, são apresentados os resultados para o Corpus SNR-CLIC. O sistema com melhor desempenho até o momento, para tal tarefa, também é baseado na abordagem Comitê ETL (San09).

Tabela 7.9: Desempenho dos algoritmos na tarefa de extração de sintagmas nominais do Português.

Algoritmo	**Acurácia**	**Precisão**	**Abrangência**	**Medida-F**
Comitê BAS ETL 400	98.12	89.91	89.96	89.93
Comitê BAS ETL 100	98.01	89.44	90.01	89.72
Comitê ETL	98.09	89.66	89.51	89.58
ETL	97.97	88.77	88.93	88.85
Comitê BAS TBL GEN	97.42	84.71	87.56	86.11
TBL GEN	97.18	80.38	85.84	83.02
Classificador Inicial	96.57	62.69	74.45	68.06

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 57%, de 31.94 para 13.89 e 68% (10.28) utilizando respectivamente os algoritmos

TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 68% (10.07). Por outro lado, o algoritmo Comitê ETL reduz o erro em 67% (10.42), sendo a redução dos erros entre os dois comitês de 3.5%. Consequentemente, para esta tarefa, o algoritmo Comitê BAS apresenta o melhor desempenho, quase quebrando a barreira dos 90% de medida-F.

Na Tabela 7.10, apresentamos os resultados para o Corpus Ramshaw & Marcus. O sistema com melhor desempenho até o momento, para tal tarefa, é baseado em uma abordagem SVM (Kud01).

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 60%, de 20.01 para 7.93 e 67% (6.69) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 67% (6.63). Por outro lado, o algoritmo Comitê ETL reduz o erro em 66% (6.71). Embora o desempenho do Comitê BAS seja superior ao Comitê ETL, ele não supera o desempenho da abordagem SVM apesar de se manter competitivo.

Tabela 7.10: Desempenho dos algoritmos na tarefa de extração de sintagmas nominais do Inglês.

Algoritmo	**Acurácia**	**Precisão**	**Abrangência**	**Medida-F**
SVM	-	94.15	94.29	94.22
Comitê BAS ETL 400	95.81	93.23	93.51	93.37
Comitê BAS ETL 100	97.89	93.14	93.48	93.31
Comitê ETL	97.89	93.09	93.49	93.29
ETL	97.57	91.88	92.36	92.12
Comitê BAS TBL GEN	97.53	91.92	92.21	92.07
TBL GEN	96.79	89.54	89.73	89.63
Classificador Inicial	94.48	78.20	81.87	79.99

Na Tabela 7.11, apresentamos os resultados para o Corpus CoNLL-2000. O sistema com melhor desempenho até o momento, para tal tarefa, também é baseado em uma abordagem SVM (Wu06).

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 60%, de 22.93 para 7.93 e 71% (6.72) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 71% (6.67). Por outro lado, o algoritmo Comitê ETL reduz o erro em 71% (6.73). Embora o desempenho do Comitê BAS seja superior ao Comitê ETL, ele também não supera o desempenho da abordagem SVM apesar de se manter competitivo.

Na Tabela 7.12, são apresentados os resultados para o Corpus SPSAL-2007. O sistema com melhor desempenho até o momento, para tal tarefa, é

Tabela 7.11: Desempenho dos algoritmos na tarefa de extração de sintagmas do Inglês.

Algoritmo	**Acurácia**	**Precisão**	**Abrangência**	**Medida-F**
SVM	-	94.12	94.13	94.12
Comitê BAS ETL 400	95.97	93.20	93.46	93.33
Comitê BAS ETL 100	95.86	93.13	93.43	93.28
Comitê ETL	95.85	93.11	93.42	93.27
ETL	95.20	92.24	92.53	92.28
Comitê BAS TBL GEN	94.74	90.58	92.03	91.30
TBL GEN	93.68	88.78	90.23	89.50
Classificador Inicial	77.29	72.58	82.14	77.07

baseado em uma abordagem que combina Cadeias de Markov Escondidas e Campos Aleatórios Condicionais (HMM+CRF) (PVS07).

Tabela 7.12: Desempenho dos algoritmos na tarefa de extração de sintagmas do Hindi.

Algoritmo	**Acurácia**
HMM+CRF	80.97
Comitê BAS ETL 400	80.58
Comitê BAS ETL 100	80.54
Comitê ETL	80.44
Comitê BAS TBL GEN	79.00
ETL	78.53
TBL GEN	78.07
Classificador Inicial	70.04

O algoritmo Comitê BAS reduz o erro do Classificador Inicial em 30%, de 29.96 para 21.00 e 35% (19.46) utilizando respectivamente os algoritmos TBL Genético e ETL. Aumentado o número de classificadores ETL gerados, a redução do erro aumenta para 35% (19.42). Por outro lado, o algoritmo Comitê ETL reduz o erro em 35% (19.56). Embora o desempenho do Comitê BAS seja superior ao Comitê ETL, ele também não supera o desempenho da abordagem SVM se mostrando apenas competitivo.

7.7 Sumário

Este capítulo apresenta resultados empíricos para a aplicação de uma abordagem BAS genérica, em duas tarefas de Processamento de Linguagem Natural largamente empregadas, anotação morfossintática e anotação de sintagmas. Para cada tarefa, quatro corpora de diversos idiomas foram utilizados.

Na maioria dos casos, a abordagem BAS genérica é superior aos algoritmos-base utilizados, TBL Genético e ETL. Além disso, nesses casos, tal abordagem é superior a abordagem Bagging representada pelo algoritmo Comitê ETL.

Nos demais casos, o BAS se mostrou competitivo em relação a classificadores derivados de outros algoritmos de aprendizado de máquina, como SVM, HMM e CRF, atingindo desempenhos próximos do estado da arte para cada tarefa, utilizando uma abordagem bem genérica. Cabe ressaltar que tais classificadores apresentados como estado da arte não são soluções genéricas como a utilizada pelo Comitê BAS e foram construídos especificamente para a resolução de cada tarefa. Caso seja aplicado uma estratégia específica derivada do BAS para cada tarefa, os resultados podem se tornar ainda melhores, inclusive ultrapassando o desempenho encontrado por tais classificadores.

Um exemplo da qualidade dos classificadores gerados utilizando tal abordagem é o extrator de sintagmas nominais do Português utilizando o Corpus SNR-CLIC. Tal abordagem possui o melhor desempenho, 89.93%, entre os classificadores conhecidos para esta tarefa.

Em termos de desempenho computacional, as abordagens também são eficientes. A abordagem Comitê BAS processa, em média para cada tarefa, 9.000 palavras por segundo utilizando o ETL como algoritmo-base na tarefa de Anotação de Sintagmas do Inglês. Essa velocidade, entretanto, cai para um terço quando da utilização do algoritmo TBL Genético.

Um dos gargalos em termos de desempenho é a fase de agrupamento dos exemplos. Entretanto, tal fase também pode ser facilmente paralelizável, ou executada *offline*, armazenando-se os grupos encontrados, tornando o aprendizado ainda mais rápido.

Também em termos de desempenho computacional, os classificadores gerados são bastante eficientes, visto que são conjuntos de conjuntos de regras simples que facilmente podem ser paralelizáveis, gerando um classificador final robusto e rápido.

Apesar de possuir um pior desempenho que as abordagens Comitê BAS ETL, aproximadamente 1.2% pior em termos absolutos, a abordagem Comitê BAS TBL GEN gera classificadores mais rápidos devido ao pequeno conjunto de gabaritos utilizado no treinamento dos classificadores.

Os resultados aqui encontrados apontam que a utilização de abordagens BAS permite gerar classificadores com um desempenho competitivo, por vezes até melhor para tarefas de PLN independentemente do idioma utilizado. Tais abordagens empregam um mínimo esforço de modelagem e permitem a geração de classificadores rápidos e facilmente paralelizáveis, demonstrando o alto grau

de robustez do método.