

8 Experimentos

Para todos os experimentos, utilizamos validação cruzada *10-fold*. Separamos 10 conjuntos, cada um contendo todas as notícias, sendo 90% utilizadas no treinamento do algoritmo e 10% para testes, onde se verifica a acurácia. No PETRONEWS o grupo de teste é composto de 62 notícias favoráveis e 43 desfavoráveis, para o grupo de treino a proporção se mantém. Os grupos não possuem o mesmo número de notícias positivas e negativas pois o corpus não está balanceado, dessa maneira precisamos manter a mesma proporção de notícias positivas e negativas tanto nos grupos de treino como nos grupos de teste.

Para cada grupo de treino é feita ainda outra validação cruzada para calibrar os parâmetros do SVM. Essa calibração também utiliza *10-fold*. Nós fizemos uma série de experimentos para verificar a eficácia de cada abordagem descrita nesse trabalho, a seguir detalhamos o impacto de cada uma delas.

8.1 Impacto do uso de Filtro por tópico

Modelo	Acurácia (%)
Saco-de-Palavras	82,76
Saco-de-Palavras filtrada por tópico	85,81

Tabela 8.1: Impacto do uso de Filtro por tópico

Primeiro investigamos o impacto de filtrar as frases sobre a Petrobras na performance do classificador. Nossa hipótese é de que frases não relacionadas com a Petrobras podem causar ruído para o classificador SVM. Nós comparamos o modelo clássico de saco-de-palavras dos textos sem o filtro e com o filtro. A tabela 8.1 mostra que nossa hipótese se confirma diminuindo o erro em 17,7% quando filtramos apenas as frases que falam da companhia.

Como o filtro do tópico melhora a performance do modelo, além de tornar os experimentos mais rápidos, fixamos esse parâmetro para os experimentos seguintes, ou seja os próximos modelos serão sempre com o filtro por tópico ligado.

8.2

Impacto do uso de atributos estruturais

Modelo (com filtro por tópico)	Acurácia (%)
Saco-de-Palavras com atributos estruturais	85,90
Saco-de-Palavras	85,81
Apenas atributos estruturais	69,71

Tabela 8.2: Impacto do uso de atributos estruturais

O uso de atributos estruturais somados ao dicionário de palavras reduz o erro em apenas 0,6% quando comparado ao modelo que considera apenas o dicionário de palavras. Por outro lado, a utilização apenas dos atributos estruturais decai muito a acurácia. Concluímos com isso que o uso apenas dos atributos estruturais não é suficiente para representar o sentimento do texto, porém combinado com outros atributos, como por exemplo o dicionário de palavras, ele ajuda o classificador a decidir se o texto é favorável ou não a Petrobras.

Abbasi obtém uma melhora de 5% em seus experimentos com o uso de saco de palavras mais atributos estruturais em críticas de cinema quando comparado a apenas ao saco de palavras. Acreditamos que para textos jornalísticos esses atributos não são de grande importância pois a estrutura dos textos é sempre muito parecida independente do sentimento. Abbasi também mostra que o uso de apenas atributos estruturais não é muito informativo, mesmo para críticas de cinema, a acurácia cai em 12% quando comparado com o experimento com apenas o léxico.

8.3

Impacto do uso de n-gramas de palavras

Modelo*	Acurácia (%)
Unigramas (1-gramas)	85,81
Unigramas + Bigramas (1,2-gramas)	85,90
Unigramas + Bigramas + Trigramas (1,2,3-gramas)	86,38

*Modelos com filtro, atributos estruturais, seletores: FS e PD

Tabela 8.3: Impacto do uso de n-gramas de palavras

A utilização de unigramas, bigramas e trigramas reduz o erro em 4,01% quando comparado ao uso apenas de unigramas do léxico. Nos experimentos mostrados na tabela 8.3 utilizamos os modelos com filtro de tópico da

Petrobras, atributos estruturais além dos n-gramas do léxico e os seletores de atributos PD combinado com F-Score.

É importante ressaltar que o uso de 1,2,3-gramas do lexico sem a seleção de atributos não resulta em uma acurácia boa, pois acaba-se criando muitos atributos que causam ruído para o classificador. No caso desse experimento, ao mesmo tempo que estamos criando muitos atributos, estamos selecionando apenas os melhores, por isso há uma melhora dos resultados.

8.4

Impacto do uso de NLP

Modelo*	Acurácia (%)
Unigramas de palavras	85,81
Unigramas de palavras + Morf.	85,24
Unigramas de palavras + Morf. + Chunk	85,24
Unigramas de palavras + Chunk	85,14

*Modelos com filtro, atributos estruturais, seletores: FS e PD

Tabela 8.4: Impacto do uso de NLP

Esses experimentos medem a eficiência dos atributos formados pelo par palavra_EtiquMorf e palavra_EtiquChk e do trio palavra_EtiquMorf_EtiquChk. Pang verifica em seus experimentos que os atributos formados pela etiqueta morfossintática não são muito informativos e acabam criando ruído para o classificador, por outro lado, Abbasi mostra que esses atributos podem ser importantes para classificar críticas de cinema e textos de foruns de internet. Pesquisas com textos jornalísticos sobre o mercado financeiro não reportam o ganho ou perda que obtiveram utilizando esses atributos. Nossos experimentos mostram uma pequena redução da acurácia quando utilizamos esse tipo de atributo, mas isso não significa necessariamente que ele não são informativos, pois ao acrescentar esse tipo de atributo estamos aumentando exponencialmente nosso número de atributos, o que dificulta a classificação do SVM. A seguir iremos mostrar que como membro de um comitê o modelo que utiliza o par palavra_EtiquMorf, e palavra_EtiquChk e o trio palavra_EtiquMorf_EtiquChk ajuda a melhorar a acurácia, portanto esse atributo é informativo.

8.5

Impacto do uso de n-gramas de etiquetas Morfossintáticas

A tabela acima avalia a eficácia do uso de bigramas e trigramas de etiquetas morfossintáticas. Como pode ser observado o uso de bigramas dessa

Modelo*	Acurácia (%)
1,2,3-gramas de palavras	86,38
1,2,3-gramas de palavras + 2-gramas de Morf.	86,19
1,2,3-gramas de palavras + 2,3-gramas de Morf.	84,95

*Modelos com filtro, atributos estruturais, seletores: FS e PD

Tabela 8.5: Impacto do uso de n-gramas de etiquetas morfossintáticas

etiqueta diminui muito pouco a acurácia quando comparada ao modelo que não a utiliza. No entanto o uso de bigrama e trigrama dessa etiqueta diminui 1.66% da acurácia quando comparado ao modelo que não utiliza esses atributos. Essa diminuição é significativa e por isso acreditamos que os trigramas causam grande ruído para o classificador.

8.6

Impacto do uso do caminho da árvore de decisão como atributo

Modelo*	Acurácia (%)
1,2-gramas de palavras + 2-gramas de Morf.	84.38
1,2-gramas de palavras + 2-gramas de Morf. + árvore	80.29

*Modelos com filtro, atributos estruturais, seletores: FS

Tabela 8.6: Impacto do uso do caminho na árvore de decisão como atributo

Nesses experimentos estamos medindo a eficiência dos atributos gerados pelo caminhamento na árvore de decisão. Como pode ser observado, esses atributos não são informativos e causam muito ruído para o classificador diminuindo em quase 5% a acurácia quando comparada com o modelo que não utiliza esse atributo.

8.7

Impacto do uso de Seleção de Atributos

Para avaliar a eficiência de cada um dos seletores de atributos, executamos 5 experimentos utilizando o mesmo modelo. Esse modelo contempla 1,2,3-gramas de palavras com filtro de tópico, os atributos estruturais e bigramas de anotação morfossintática. O resultado de cada experimento é mostrado na tabela acima.

Com exceção do seletor por árvore de decisão, todos os outros se mostram mais eficientes do que não fazer seleção nenhuma. A seleção por árvore de decisão diminui muito o número original de atributos e conseqüentemente acaba eliminando atributos informativos e importantes para o classificador.

Modelo*	Acurácia (%)
PD e Fisher Score	86,19
Fisher Score	86,10
PD	85,43
EWGA	84,95
Árvore de Decisão	79,81
Sem Seleção	84,76

* Modelos com filtro; atributos estruturais; 1,2,3-grams palavra; 2-grams Morf.

Tabela 8.7: Impacto do uso de Seleção de Atributos

A seleção por EWGA é muito custosa computacionalmente e por isso não vale a pena utilizá-lo como seletor apesar de apresentar resultados positivos. Fisher Score e PD são rápidos de rodar e eficientes na tarefa de selecionar os melhores atributos. O melhor resultado foi obtido utilizando a combinação desses dois últimos seletores. Primeiramente o PD se encarrega de eliminar os atributos categóricos com PD abaixo de 0,125, em seguida o FScore recebe os exemplos já com a redução prévia do PD e elimina outros atributos, categóricos ou não, menos informativos. Essa combinação de seletores de atributos resulta em uma diminuição do erro de 9,38% quando comparado com o experimento que não utiliza seleção de atributos.