

9 Resultados

9.1 Melhores Resultados

Modelo*	Acurácia (%)
Comitê	87,14
1,2,3-gramas de palavras	86,38
1,2,3-gramas de palavras + 1,2-gramas de Morf.	86,19
1,2,3-gramas de palavras + Morf. + Chunk	86,00
Unigramas	82,76

* Com filtro, atributos estruturais, seleção por PD e FS

Tabela 9.1: Melhores Resultados

Na tabela acima mostramos na primeira linha o resultado de um comitê formado pelos 3 melhores modelos. Em seguida mostramos quais são esses modelos e suas acurácias e na última linha o modelo básico de sacos de palavras. A ideia de formar um comitê é antiga, o objetivo é combinar "opiniões" de diferentes modelos que são formados por atributos distintos. A combinação de opiniões é mais robusta e tolerante a falhas já que não utiliza a informação de apenas um modelo.

Sharkey defende a utilização de comitê na seguinte frase: "A ideia de aproveitar, ao invés de perder, a informação contida em estimadores imperfeitos, é central para a utilização de máquinas de comitê".

Modelos combinados normalmente generalizam mais, pois os erros que um membro pode ter são consertados pelos outros integrantes do comitê. É importante porém, que os membros do comitê tenham boa eficiência e que quando os erros ocorram, que eles não ocorram em todos os modelos do comitê ao mesmo tempo.

O melhor modelo é formado por filtro por tópico, 1,2,3-gramas de palavras, atributos estruturais e seleção de atributos PD e FS combinados. Os dois seguintes melhores modelos ainda recebem bigramas de atributos morfosintáticos e a combinação de palavra_etiqueta PLN além das características do primeiro modelo. O comitê diminui o erro em 5,58% com relação ao melhor

modelo e 25,40% quando comparado ao modelo clássico de saco de palavras sem o filtro por tópico ou qualquer seleção de atributos.

9.2

Resultados por Categoria

	Ações PETR	Mercado	Institucional
Positivas	4, 15	7, 28	16, 11
Negativas	12, 16	22, 02	40, 40

Tabela 9.2: Erro por categoria de notícias

Nessa seção mostramos a distribuição de erro por categoria de notícias descritas no capítulo 3. Note que notícias positivas são mais fáceis de classificar. A explicação pode ser pelo fato de que existem mais notícias positivas do que negativas. A quantidade de notícias é sempre um fator importante, pois quanto mais notícias melhor o aprendizado do classificador, e por isso menor o erro. A Categoria "Ações PETR" é a mais fácil de classificar pelos anotadores humanos e a com maior número de exemplos, por isso era razoável esperar que o erro seria menor nessa categoria. A Categoria "Institucional" é a mais difícil tanto para os anotadores humanos quanto para a máquina, isso se comprova ao observar que mesmo tendo mais exemplos positivos do que a Categoria "Mercado", sua porcentagem de erro foi maior.