

## 3 Petronews

### 3.1 Geração

Coletar notícias do Mercado financeiro é uma tarefa importante. Existem diversos Web sites que fornecem notícias em tempo real. Esses Web sites são normalmente classificados por língua. Para o inglês, por exemplo, temos o *PRNewsWire*, que fornece notícias gratuitas assim como informa as notícias mais lidas. Outro Web site em inglês é o *Yahoo! Finance*. Esse último agrega diversas fontes de notícias e redistribui através de um serviço gratuito de RSS. Para o português, os dois principais Web sites de notícias do mercado são: *Gazeta Mercantil* e *ValorOnline*. Para criar o PETRONEWS, um corpus anotado de notícias financeiras em português sobre a Petrobras, desenvolvemos um sistema de coleta de dados. Esse sistema coleta notícias financeiras relacionadas com a Petrobras a partir das duas principais fontes: *Gazeta Mercantil* e *ValorOnline*.

O *ValorOnline* oferece serviço RSS gratuito, o que facilita a coleta de dados, porém o RSS não fornece notícias antigas, apenas as atuais. Para ter acesso a base histórica de notícias é necessário fazer uma assinatura. A assinatura também dá acesso às notícias do jornal impresso que foram digitalizadas. As notícias online são liberadas em tempo real, por isso podemos usá-las futuramente para fazer uma análise dentro de um mesmo dia com o mercado, as notícias impressas, porém normalmente reportam notícias escritas no dia anterior, e por isso sua análise seria interessante apenas para impacto entre um dia e outro. Para nosso trabalho de análise de sentimento, qualquer notícia é válida, independente do momento em que é lançada na Web, por isso utilizamos tanto notícias online como impressas. O site da *Gazeta Mercantil* libera sua base histórica sem a necessidade de uma assinatura, porém o número de notícias liberadas por dia é menor do que o *ValorOnline*. Juntas essas duas fontes fornecem em média 10 notícias por dia.

Ocasionalmente, ambos os *feeds* liberam notícias com a mesma informação, quando isso ocorre deixamos apenas uma das cópias, para não alterar

o peso dessa notícia.

O PETRONEWS contém 1050 notícias sobre a Petrobras de 02/06/2006 a 29/01/2010. Cada notícia está manualmente classificada como positiva ou negativa com relação à Petrobras.

### 3.2 Características

Classificar notícias jornalísticas, diferentemente de outros tipos de texto, apresenta grandes dificuldades. A maioria dos trabalhos de análise de sentimento classifica textos que contém explicitamente opiniões ou críticas (13, 01). Geralmente, notícias jornalísticas não expressam qualquer tipo de sentimento, procurando ser imparciais e objetivas. A tarefa de classificar o sentimento de notícias financeiras está muito mais relacionada ao fato da notícia trazer informações que revelem impactos positivos ou negativos no desempenho da Petrobras do que em encontrar opinião ou subjetividade nos textos. Devitt e Ahmad (04) mostram que a opinião pessoal e profissional sobre uma empresa influencia no momento da classificação do texto jornalístico, assim como o fato da pessoa trabalhar ou não no mercado de compra e venda de ações. Porém essa classificação é usualmente consistente num mesmo corpus.

No PETRONEWS, a classificação das notícias foi feita por 3 pessoas: 2 homens e 1 mulher, que não são do mercado financeiro. Além disso, podemos dividir o corpus em 3 categorias de notícias, dependendo do conteúdo das mesmas, a saber:

- ações PETR - Notícias de preço das ações da Petrobras no mercado financeiro;
- Mercado - Notícias sobre todas as ações operadas no mercado financeiro;
- institucional - Notícias sobre a Petrobras, sem mencionar o preço das ações.

A primeira categoria descreve a variação de preços das ações da Petrobras no mercado. A segunda categoria mostra um panorama geral das principais bolsas de valores do mundo, incluindo a Bovespa juntamente com informações da Petrobras. Essa segunda categoria é muito importante, porque não mostra apenas a situação da Petrobras, mas sim uma visão geral do mercado. A terceira categoria contém informações sobre a Petrobras que não estão diretamente relacionadas com o mercado financeiro. Ela descreve investimentos da empresa, rumores, escândalos e outras informações que podem ajudar a prever o comportamento dos preços das ações no mercado.

As duas primeiras categorias são mais fáceis de classificar: se o preço das ações subiu temos uma notícia positiva; e se caiu temos uma notícia negativa. A princípio não estamos utilizando as informações sobre as outras

ações do mercado financeiro, ou seja, estamos descartando a visão geral do mercado. Fazemos isso por dois motivos, primeiro porque a Petrobras não possui uma concorrente ou aliada direta no mercado. Dessa maneira a queda ou valorização de suas ações não está diretamente relacionada com nenhuma outra companhia. Por exemplo, a valorização da Vale, não significa nada para a Petrobras. Segundo, porque para uma classificação binária (positiva ou negativa) a visão do mercado não importa muito. Porém, a partir do momento em que a classificação for feita por uma graduação isso terá que mudar. Se a Petrobras desvalorizou junto com todas as outras ações do mercado é menos negativo do que o mercado ter valorizado e a Petrobras caído.

A categoria "Institucional" é a mais difícil de classificar e é onde a opinião do classificador pode influenciar um pouco a classificação. Por exemplo:

"José Sergio Gabrielli, presidente da Petrobras, diz que não teme a instalação de uma CPI para investigar irregularidades (...)".

Esse trecho poderia ser considerado positivo já que o presidente da empresa não teme uma investigação, porém o fato de que irão fazer uma investigação é algo negativo.

Em casos como esse a classificação é consistente com outras notícias sobre o assunto. As notícias em que houve muita dúvida na classificação, ou que não ficou claro o sentimento, foram descartadas.

Na tabela 3.2, pode ser vista a distribuição do número de notícias classificadas em cada uma das categorias como positivas e negativas. As notícias que foram descartadas não estão sendo consideradas nessa tabela. O PETRONEWS contém 41,18% de notícias negativas e 58,82% de notícias positivas.

Sentimento	Ações PETR	Mercado	Institucional	TOTAL
Positivas	27,36%	14,16%	17,30%	58,82%
Negativas	21,34%	10,40%	9,44%	41,18%
TOTAL	48,70%	24,56%	26,74%	100%

Tabela 3.1: Distribuição de Notícias do PETRONEWS por Sentimento.