

Analizando Manchetes e Posts de Redes Sociais para inferência de Aspectos e Sentimentos no Contexto Financeiro

Dayan de F. Costa¹, Nádia Félix Felipe da Silva¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

{dayancosta,nadia}@inf.ufg.br

Abstract. *Investors have always been interested in forecasting the stock prices. Hundreds of financial news are released on different media every day. In this article we present a supervised learning approach based on Support Vector Machines in which the main objective is to detect sentiments and aspects in messages from microblogs, news and headlines from financial sector. The system presented was a participant in the Open Challenge - Financial Opinion Mining in FiQA (2018)¹ - Task 1, reaching a mean squared error of 0.0958436, the best of the competition, in one of the datasets provided by organizers.*

Resumo. *Investidores sempre se interessaram pela previsão do preço das ações. Centenas de notícias financeiras são divulgadas em diferentes mídias todos os dias. Neste artigo apresentamos uma abordagem de aprendizado supervisionado baseada em Support Vector Machines em que o principal objetivo é detectar sentimentos e aspectos em mensagens de microblogs, manchetes e notícias do setor financeiro. O sistema apresentado foi participante do Open Challenge – Financial Opinion Mining in FiQA (2018)¹ - Tarefa 1, atingindo um erro médio quadrado de 0.0958436, o melhor da competição, em um dos datasets fornecidos pelos organizadores.*

1. Introdução

O uso de veículos de comunicação como Twitter² e StockTwits³, chamados de *microblogs*, tem aumentado nos últimos anos, fazendo com que o mercado dê mais atenção no que é dito sobre as empresas nestes ambientes. Este aumento é dado por conta da facilidade de uso, o formato das mensagens e a acessibilidade nestas plataformas. As pessoas estão utilizando estes veículos de comunicação para expressar seus sentimentos sobre a vida, negócios, trabalho, esportes entre outros, fazendo com que as marcas fiquem mais preocupadas com o que as pessoas dizem sobre elas a fim de identificar qual o sentimento das pessoas sobre determinada marca, atrair mais clientes, etc. No domínio financeiro, estes textos cheios de opiniões são capazes de mudar o setor, elevando ou decaindo o valor do mercado de ações [Smailovic et al. 2014].

Dado o crescente uso das plataformas de *microblogging* na última década, as aplicações de tecnologia utilizando os dados gerados por elas também têm crescido, fazendo com que estas começassem a migrar da academia para o mercado. Nos

¹<https://sites.google.com/view/fiqa/home>

²<https://twitter.com>

³<https://stocktwits.com>

últimos anos, a análise de sentimentos, assim como outras metodologias, começaram a se tornar um interesse do mercado e um interesse de pesquisa no setor financeiro [Bollen et al. 2010, Smailovic et al. 2014, Takala et al. 2014, Cortis et al. 2017]. Desde então os pesquisadores vêm mostrando que as opiniões expressadas nos *microblogs* e em redes sociais podem ter um grande impacto no mercado [Goonatilake et al. 2007].

A análise de sentimentos é uma técnica de *machine learning* que consiste em extrair, em vários níveis de granularidade, informações de textos em linguagem natural com o objetivo de obter a polaridade do texto ou sentença em questão. Os níveis de granularidade são três:

1. Nível de documento que observa o sentimento global expresso em determinado texto [Pang and Lee 2004];
2. Nível de sentença cujo o objetivo é classificar a polaridade de cada sentença em determinado texto [Kim and Hovy 2004];
3. Nível de característica ou aspectos cujo objetivo é analisar a polaridade das opiniões sobre características e/ou atributos do objeto [Wilson et al. 2009].

A aplicabilidade da análise de sentimentos é vasta[Liu 2012], podendo ser utilizada para monitorar a reputação de determinada marca, como determinada notícia afeta determinados grupos de pessoas, e ultimamente também tem sido utilizada por candidatos políticos e administradores para monitorar opiniões gerais sobre mudanças e campanhas políticas, permitindo que eles ajustem sua campanha para melhor se relacionar com os eleitores. Esta área do processamento de linguagem natural é cheia de desafios e, assim como a linguagem natural, a análise de sentimentos é bastante complexa, pois o que é relativamente fácil para os humanos classificarem subjetivamente na comunicação deve ser traduzido em pontuações quantificáveis. Por exemplo, uma palavra que em determinado contexto carrega um sentido positivo, poderá ser utilizada de maneira irônica e facilmente ser interpretada incorretamente por um algoritmo caso o contexto e o tom utilizados não forem levados em consideração.

Neste trabalho, aplicamos técnicas conhecidas como *word embeddings* [Mikolov et al. 2013], *n-grams*, *word replacement*⁴ e múltiplos tipos de regressão e classificação como SVM (*Support Vector Machines*) com *kernels* lineares [Joachims 1998], modelos Bayesianos, *ensembles*⁵ e modelos de classificação baseados em árvores [Tan et al. 2005].

O restante do artigo é estruturado da seguinte forma: a seção 2 descreve os trabalhos relacionados, a seção 3 detalha os conjuntos de dados utilizados, experimentos e a discussão dos resultados e por fim a Seção 4 conclui o nosso trabalho.

2. Trabalhos Relacionados

Muitas iniciativas de pesquisas têm focado na análise de sentimentos das mídias sociais e essas podem ser utilizadas para prever o futuro de indicadores das ações do mercado financeiro. Nesta seção, damos uma visão geral dos estudos relacionados focados em:

⁴*Word Replacement* é uma técnica de substituição de palavras com significado comum como medida de redução do vocabulário. Reduzindo o vocabulário sem perda do significado podemos economizar processamento em casos como análise de frequência e indexação de texto.

⁵Classificadores de sentimentos combinados com o objetivo de prover melhores resultados do que o uso de um único classificador [Fulcher 2008].

(i) análise de sentimentos de mídias sociais como preditores de indicadores do futuro do mercado de ações e (ii) análise de sentimentos baseadas em aspectos.

2.1. Análise de sentimentos de tweets em textos oriundos de domínio financeiro

A publicação mais conhecida nesta área é a de Bollen [Bollen et al. 2010], onde foram investigadas medidas de humor coletivo derivadas de *tweets* e se havia correlação com valor da DJIA (*Down Jones Industrial Average*) ao longo do tempo. A análise de causalidade de Granger e uma Rede Neural *Fuzzy* foram utilizadas para investigar a hipótese de que os estados de humor públicos medidos pela série temporal de humor do OpinionFinder e do Google-Profile of Mood States (GPOMS), foram preditivos de alterações nos valores de fechamento da DJIA. Seus resultados indicaram que a precisão das previsões da DJIA poderiam ser melhoradas pela inclusão de dimensões específicas do humor público.

Smailovic [Smailovic et al. 2014] também analisou se o sentimento expresso em *tweets* que discutiam sobre empresas selecionadas e seus produtos, podem indicar suas mudanças nos preços das ações. Para resolver este problema, uma abordagem de aprendizado ativo foi desenvolvida e aplicada à análise de sentimentos de *stream* de *tweets* no domínio do mercado de ações. O artigo começa apresentando um problema estático de análise de dados do Twitter, explorado a fim de determinar a melhor configuração de pré-processamento dos dados específicos do Twitter para treinar o classificador de sentimento baseado em *Support Vector Machine* (SVM). Na configuração, o teste de causalidade de Granger mostra que os sentimentos em *tweets* relacionados às ações podem ser utilizados como indicadores de movimentos de preços de ações em alguns dias de antecedência, onde os melhores resultados foram obtidos adaptando o classificador SVM para categorizar as postagens do Twitter em três categorias de sentimento: positivo, negativo e neutro. Essas descobertas foram adotadas no desenvolvimento de uma nova abordagem de aprendizagem ativa baseada na *streams* para análise de sentimentos, aplicável na aprendizagem incremental a partir de *stream* de *tweets* financeiros em constante mudança. Para este fim, uma série de experimentos foram conduzidos a fim de determinar a melhor estratégia de consulta para o aprendizado ativo do classificador SVM adaptado à análise de sentimento de *streams* de *tweets* financeiros. Os experimentos na análise de sentimentos do mercado de ações de uma determinada empresa mostraram que mudanças na probabilidade do sentimento positivo podem ser utilizadas como indicadores das mudanças nos preços de fechamento das ações.

Pagolu [Pagolu et al. 2016] observou quão bem as alterações nos preços das ações de uma empresa, os aumentos e quedas, estão correlacionados com a opinião pública sendo expressada em *tweets* sobre aquela companhia. O artigo empregou duas representações textuais, *Word2Vec* e *N-gram*, para análise do sentimento público em *tweets*. Os autores aplicaram análise de sentimentos e princípios de aprendizado de máquina supervisionados aos *tweets* e analisaram a correlação entre as alterações do mercado de ação de uma empresa e o sentimento de *tweets*. Eles mostraram que notícias e *tweets* positivos nas mídias sociais sobre uma empresa encorajariam as pessoas a investir nas ações daquela empresa e, como resultado, o preço das ações da empresa aumentaria. Além disso, este estudo demonstrou uma forte correlação entre os aumentos e quedas nos preços das ações com os sentimentos do público em *tweets*.

Dada a ligação entre sentimento e a dinâmica do mercado, a análise de sentimento público torna-se um método poderoso para prever as reações do mercado. No entanto,

a precisão das abordagens de sentimento baseada em aprendizado de máquina raramente excede setenta por cento [Takala et al. 2014].

2.2. Análise de sentimentos baseadas em aspectos

Sistemas de análise de sentimentos baseados em aspectos tratam um conjunto de textos (por exemplo, revisão de produtos ou mensagens de mídias sociais) discutindo uma entidade em particular (por exemplo, um novo restaurante) para detectar os principais aspectos (características) desta entidade (por exemplo, comida ou serviço). O objetivo é estimar o sentimento médio dos textos por aspectos (por exemplo, quão positivas ou negativas as opiniões são, em média, para cada aspecto)[Liu 2012].

Um dos primeiros estudos sobre a análise de sentimentos baseadas em aspectos é de [Hu and Liu 2004], que seguiu uma abordagem baseada em frequência. A ideia é, basicamente, que os nomes frequentemente mencionados são mais propensos a serem um aspecto. Para compensar os erros resultantes causados por terem ignorado os nomes pouco frequentes, os autores sugeriram explorar palavras de opinião para encontrar os aspectos. Para esta parte, eles propuseram considerar a opinião mais próxima. Esta ideia foi utilizada em documentos de [Zhuang et al. 2006].

As principais abordagens utilizam classificadores com recursos manuais caros baseados em *n-grams* e léxico de sentimentos. Por exemplo, Popescu and Etzioni [Popescu and Etzioni 2005] sugeriram o uso de *parsers* sintáticos para remoção de termos nominais que frequentemente não são aspectos. Beneficiando-se de estatísticas sobre o uso de substantivos da língua inglesa, Scaffidi [Scaffidi et al. 2007] melhorou a abordagem geral de confiar em nomes frequentes para extrair aspectos. Long [Long et al. 2010] adicionou o uso de informações de distância e palavras dependentes (adjetivos).

3. Avaliação Experimental

3.1. Datasets

O *dataset* utilizado foi disponibilizado pela competição *Open Challenge – Financial Opinion Mining in FiQA (2018)*¹ - Tarefa 1, a qual é relacionada à análise de sentimento financeiro baseada em aspectos. O conjunto de dados consiste de 675 mensagens de *microblogs* e 438 notícias e manchetes para a fase treino (ver Tabela 1). Para a fase de teste o *dataset* possui 99 mensagens de *microblogs* e 93 notícias e manchetes. Os dados foram extraídos de websites como StockTwits³, Reddit⁶, Wikinews⁷ entre outras páginas de domínio financeiro.

Tabela 1. Estatísticas do dataset

	Instâncias	Positivas	Negativas	Neutras
Mensagens de Microblogs	675	440	234	1
Manchetes e Notícias	438	282	144	12

A Tabela 1 refere-se às estatísticas sobre os dados de treino. A coluna Instâncias refere-se à quantidade de sentenças disponíveis no *dataset*. Positivas, Negativas e Neutras correspondem às instâncias cuja pontuação dos sentimentos eram positivas, negativas ou zero, respectivamente.

⁶<https://www.reddit.com/>

⁷<https://www.wikinews.org/>

Sobre os aspectos, o *dataset* possui 95 classes de aspectos para as notícias e manchetes e 83 classes de aspectos para as mensagens de *microblogs*. Dentre estas classes, as mais comuns nas notícias e manchetes são: 'corporate', 'm&a' e 'stock' com 327, 106 e 101 classificações, respectivamente. As classes mais comuns dentre as mensagens de *microblogs* são: 'stock', 'price action', e 'bullish' com 546, 379 e 203 classificações respectivamente. Os *datasets* possuem muitas classes com apenas uma mensagem representante daquele aspecto. Por exemplo, o conjunto de mensagens de notícias e manchetes possui 28 classes e o de mensagens de *microblogs* possui 29 classes com apenas um representante de cada aspecto.

O *dataset* de treino contém os seguintes atributos:

- **Target:** Nas mensagens de *microblog* este refere-se às siglas das ações das companhias sem a *cashtag* (\$). Nas notícias e manchetes este refere-se à companhia principal citada na sentença.
- **Sentence:** A mensagem ou notícia/manchete na qual o sentimento é expresso.
- **Snippets:** A principal parte da sentença.
- **Aspects:** Os aspectos associados às sentenças.
- **Sentiment Score:** O sentimento da sentença em uma escala entre -1 e 1, onde 1 é muito positivo, -1 é muito negativo e 0 representa um sentimento neutro.

3.2. Engenharia de atributos

Durante o desenvolvimento do sistema nós testamos um conjunto de técnicas e parâmetros na tentativa de alcançar um modelo preditivo com o maior desempenho. As técnicas utilizadas são descritas abaixo:

1. *N-Grams*

É uma sequência de palavras em um texto/sentença. Nós fizemos o uso de unigrams, bigrams, and trigrams no *grid search* (estimação dos parâmetros dos algoritmos de classificação utilizados) para identificação de qual teria melhores resultados.

2. *Tokenização*

Foi utilizada uma biblioteca chamada Unitok⁸ [Michelfeit et al. 2014] para tokenizar as sentenças. Esta biblioteca já realiza a tratativa de URLs, endereços de e-mails, DNS e endereços de IPs, além de reconhecimento de abreviações da língua inglesa.

3. *Word Replacements*

Foram testadas a substituição e a não substituição das siglas das ações das companhias apresentadas nas sentenças, bem como a substituição de palavras próximas de excelentes e horríveis. Para a substituição das palavras positivas nós utilizamos as palavras mais similares, baseadas na similaridade de cosseno para identificação da polaridade, neste caso utilizamos a média das 10 palavras mais similares. Com a sigla das ações, foi testada a substituição das mesmas por uma palavra comum (por exemplo, substituímos a sigla \$APPL pela palavra 'company') e também testamos a sua remoção das sentenças.

4. *Word Embeddings*

Assim como muitas tarefas envolvendo o processamento de linguagem natural, nós decidimos pelo uso de *word embeddings*, que possuem melhor performance

⁸<http://corpus.tools/wiki/Unitok>

que o tradicional *bag-of-words*, que dificilmente captura a semântica das palavras. Como esta técnica foi amplamente utilizada pelos vencedores da *Task 5* do *SemEval 2017* [Cortis et al. 2017], nós decidimos pelo uso da técnica em nossos experimentos através do uso do modelo *word2vec*⁹. Nós escolhemos por utilizar o *word2vec* por conta do seu tamanho (3 milhões de entradas) e por conta que o mesmo é treinado no Google News¹⁰ utilizando o método de Mikolov [Mikolov et al. 2013].

5. TF-IDF (*Term Frequency - Inverse Document Frequency*)

Foi utilizado o balanceamento TF-IDF [Salton and Buckley 1988] para redução do impacto da frequência de palavras em ambos os casos, regressão e classificação.

3.3. Configuração experimental

Na fase de treino, nós dividimos os dados utilizando a proporção de 80%-20% de maneira aleatória, onde a parte que continha 80% dos dados era utilizada para o treino dos algoritmos e a parte com 20% dos dados era utilizada para testar e validar nossos algoritmos. Os resultados foram avaliados utilizando o Erro Quadrático Médio (MSE - *Mean Squared Error*) e o Coeficiente de Determinação (*R Squared*) para a análise dos sentimentos, e a precisão, revocação (*recall*) e F1-Score para a tarefa de classificação dos aspectos. Nós utilizamos, em ambos os casos, algoritmos baseados em Máquinas de Vetor de Suporte (SVM - *Support Vector Machine*) do sklearn¹¹, então nós aplicamos o algoritmo de *Support Vector Regression* para a tarefa de análise de sentimentos e o algoritmo de *Support Vector Classifier* para a tarefa de classificação.

3.4. Resultados e discussão

Nesta seção apresentamos os resultados da análise de sentimento e classificação de aspectos realizada no *Open Challenge – Financial Opinion Mining in FiQA (2018)* - Tarefa 1. Em nossos testes, obtivemos os seguintes resultados, de acordo com as métricas informadas pela organização – ver Tabela 2.

Tabela 2. Resultados da classificação de aspectos no "conjunto de testes não oficial". Os dados de treinamento foram divididos randômicamente utilizando a proporção 80%-20%, onde a porção que continha 80% dos dados foi utilizada para treinar o algoritmo e os 20% para teste e validação do algoritmo

	Precisão	Revocação	F1-Score
Mensagens de <i>Microblogs</i>	0.6673	0.5592	0.5775
Manchetes e Notícias	0.4992	0.4	0.4240

De acordo com os resultados oficiais¹², nós atingimos as seguintes pontuações sobre o desafio de análise de sentimentos – ver Tabela 3.

⁹<https://code.google.com/archive/p/word2vec/>

¹⁰<https://news.google.com/>

¹¹<http://scikit-learn.org/stable/index.html>

¹²Até a data da publicação deste artigo, os organizadores ainda não haviam aberto ao público e divulgado a classificação real e os valores para as tarefas de classificação de aspectos e análise de sentimentos.

Tabela 3. Resultados Oficiais - Análise de Sentimentos

	MSE	R Squared	Cosseno
Mensagens de <i>Microblogs</i>	0.206794	0.1665593177	0.415379
Manchetes e Notícias	0.0958436	0.1642305099	0.533388

De acordo com os resultados oficiais¹², nosso sistema obteve o melhor *Mean Squared Error* da competição na utilização do *dataset* de manchetes e notícias com o valor de 0.0958436.

Nossa abordagem, nesta competição, foi a única que não utilizou de recursos de *Deep Learning* e obteve resultados satisfatórios tendo em vista o tamanho do conjunto de dados fornecido e as técnicas aplicadas. Durante o desenvolvimento do nosso sistema participante, cujo foco foi obter bons resultados utilizando técnicas simples de aprendizado de máquina, nós testamos outros algoritmos como as árvores de regressão com níveis de profundidade 3, 5 e 10, alguns algoritmos de *ensembles*¹³ como o Ada Boost e Random Forest. Nós também testamos a realização de um *ensemble*¹³ utilizando modelo de Ada Boost com SVM e extraíndo a média dos resultados dos modelos, porém nossos melhores resultados foram obtidos através do uso de SVMs.

4. Conclusões

Neste artigo, nós descrevemos a nossa solução implementada para a competição Open Challenge – Financial Opinion Mining in FiQA (2018)¹ - Tarefa 1, com o objetivo de prever sentimentos (positivos, negativos e neutros) em relação à uma marca alvo de manchetes, notícias e mensagens de *microblogs* financeiras, além de prever os aspectos relacionados à cada mensagem. A organização não divulgou o *rank* oficial, apenas as métricas analisadas e o resultado de cada métrica, onde obtivemos o melhor *Mean Squared Error* do desafio ao analisar o *dataset* de manchetes e notícias. Nossa abordagem é baseada na utilização de *Support Vector Machine*, uma técnica de aprendizado supervisionado utilizando de um *grid search* para obter o melhor classificador/regressor para as configurações testadas. Nós mostramos que é possível obter bons resultados utilizando abordagens simples de aprendizado de máquina aliadas à uma boa seleção de atributos.

References

- Bollen, J., Mao, H., and Zeng, X. (2010). Twitter mood predicts the stock market. *CoRR*, abs/1010.3003.
- Cortis, K., Freitas, A., Daudert, T., Huerlimann, M., Zarrouk, M., Handschuh, S., and Davis, B. (2017). Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 519–535. Association for Computational Linguistics.
- Fulcher, J. (2008). *Computational Intelligence: An Introduction*, pages 3–78. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Goonatilake, R., Herath, A., Herath, S., Herath, S., and Herath, J. (2007). Intrusion detection using the chi-square goodness-of-fit test for information assurance, network, forensics and software security. *Journal of Computing Sciences in Colleges*, 23(1):255–263.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 168–177, New York, NY, USA. ACM.

¹³Técnica que permite treinar diferentes classificadores de sentimentos e então combinar as saídas com uma regra de combinação, por exemplo voto majoritário.

- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine learning: ECML-98*, pages 137–142.
- Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Long, C., Zhang, J., and Zhut, X. (2010). A review selection approach for accurate feature rating estimation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10*, pages 766–774, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Michelfeit, J., Pomikálek, J., and Suchomel, V. (2014). Text tokenisation using unitok. In Horák, A. and Rychlý, P., editors, *RASLAN 2014*, pages 71–75, Brno, Czech Republic. Tribun EU.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Pagolu, V. S., Challa, K. N. R., Panda, G., and Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. *CoRR*, abs/1610.09225.
- Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics.
- Popescu, A.-M. and Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 339–346, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.*, 24(5):513–523.
- Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., and Jin, C. (2007). Red opal: Product-feature scoring from reviews. In *Proceedings of the 8th ACM Conference on Electronic Commerce, EC '07*, pages 182–191, New York, NY, USA. ACM.
- Smailovic, J., Grcar, M., Lavrac, N., and Znidarsic, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, 285:181 – 203. Processing and Mining Complex Data Streams.
- Takala, P., Malo, P., Sinha, A., and Ahlgren, O. (2014). Gold-standard for topic-specific sentiment analysis of economic texts. In Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35(3):399–433.
- Zhuang, L., Jing, F., and Zhu, X.-Y. (2006). Movie review mining and summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM '06*, pages 43–50, New York, NY, USA. ACM.