MergerBERT: Predicting Merger Targets and Acquirers from Text via Pretrained Language Models

1 Abstract

We explore the use of a U.S.firm's SEC filings to predict whether the firm will be an acquirer or a target of an acquisition within a year of the filing. Our approach uses transfer learning.

We find that word and phrase features have significant predictive power in models of being an acquirer or a target. In each case, the best performing models involve a different use of text alongside standard financial variables.

We apply the proposed model on several real-world tasks and achieve state-of-the-art performance in almost all of cases.

2 Introduction

Our perform outperforms all other methods.

Mergers are important, but they are also a relatively infrequent event: An average of about 5% of public firms have been acquired every year between 1980 and 2011. Interestingly, but perhaps not surprisingly, mergers are difficult to predict. Table 1 lists prior research aimed at predicting targets where the general conclusion is that predicting target fimms with any accuracy has proven difficult [1]. As the table shows, the explanatory power of the models (typically a logistic regression) is relatively low, with some evidence of interesting time-series properties (mergers come in waves). More directly, merger announcements typically involve a large premium over current prices (between 40% and 50% on average see [9]) and lead to a large and rapid change in market prices suggesting the announcement is news to the market. Accordingly, any improvement in the ability to predict which firms will be involved in a merger deal would prove to be very profitable for an investor in the stock market

In this paper, we predict whether a U.S. firm will participate in a merger or acquisition either as an acquirer or a target by exploiting text data. Specifically, we use the disclosure in the firm's Management Discussion and Analysis (MD&A) section of the annual 10-K filing form. We consider two types of predictions: 1) Will the firm be an acquirer in the subsequent year after the filing? and 2) Will the firm be the target of an acquisition in the subsequent year? We explore the quality of text in making these predictions, and demonstrate how text-based prediction models can offer intuitive hints about upcoming mergers.

The central contributions of this paper are as follows:

- MergerBERT
- A demonstration of the predictive utility of text disclosures for takeover bids
- Modeling innovations for scalable text-driven forecasting.

This paper contributes to the literature that studies the determinants of corporate acquisition decisions. To the best of our knowledge, this is the first paper that uses pretrained models to predict takeover targets and acquirers.

2.1 Related Work

[12] uses variables constructed from text to predict M&A events. However, their approach is substantially different from ours: They build measures of product market similarity across firms from 10-K product descriptions, and employ them as explanatory variables in standard logistic regressions to predict targets and acquirers. [23] uses text regressions to study the predictive power of words and phrases used by the management in their annual 10-K discussion and analysis for takeover events. In comparison, our proposed method MergerBERT is different than these methods in multiple ways: 1-Our method is , 2-yyy.

In general, we contribute to the growing literature that uses data extracted from text to study corporate finance decisions. [18, 14] perform textual analyses of the initial public offering prospectuses to study stock returns of IPO firms. [11, 2, 3] generate text-based measures of financial constraints to study corporate investment, financing decisions, and stock returns, respectively. Our paper shows that information from text when combined with recent deep learning methods is useful to predict mergers and acquisitions, which represent one of the most important investment decisions that firms undertake.

The paper is organized as follows: Section 2 describes the data and the construction of the financial and text variables used for estimation. Section 3 presents the estimation methodology, Section 4 the baseline regression results using standard financial variables, and Section 5 the results incorporating text variables. Section 6 develops and estimates a predictive model that interacts financial and text variables, and Section 7 concludes.

Tobin's Q [?]	The ratio of the market value of company's equity value to book
PPE	The book value of property plant and equipment
Cash balance	Logarithm
Size of Leverage	Book value of debt over book value of assets
Size	Market value of equity
Return on Assets	Operating income divided by year-end book value of assets

Table 1:

Table 2: The regularized and unregularized maximum likelihood estimates for the baseline logistic regression model, and performance on test (out of sample) data. For regularized estimate, $\lambda_1 = xxx$ and $\lambda_2 = xxx$. Table shows only the strongest positive-weighted and negative-weighted year coefficient for each model.

3 Experiments

3.1 Financial Model Baseline

Our baseline is logistic regression model with financial explanatory variables. This baseline model uses explanatory variables that are standard in the literature as in Table 3.1. For each of the variables in Table 3.1, we transform them with the z-score to ease interpretation. We use an indicator variable for each calendar year (the year is the year the report is published: 1995 - 2011). Financial variables are all measured for the year-end of the 10-K report.

For our resulting model with intercept, we estimate both an regularized and unregularized laogistic regression models. The regularized and unregularized results are shown in 3.1. Results between regularized and unregularized models are almost indistinguishable in performance, where coefficients are also similar.

For the acquirer prediction task, we achieve a pseudo R^2 just below 0.07, with the year and firm size as the strongest effects. For the target prediction task, we find that this model obtains a pseudo R^2 of 0.0262. This baseline is comparable to, and perhaps stronger than, the values reported by two other studies with overlapping data [4, 10]. According to our results, xxx, yyy, and zzz are significant predictors. Overall, these experiments suggest that the target task is more difficult than the acquirer task.

4 Data

This research is based on two kinds of data: Financial disclosures via company 10-K filings and M&A events.

4.1 Merger Bid Event Data

Our data is obtained from Eikon Thomson Reuters database. We focus on the list of pairs of companies that have made (acquirer) and received (target) a merger offer in a given year from Eikon database. Such offers may eventually be unsuccessful or successful. Each offer occurs on a specific date, and we focus on the period 1995-2019 which overlaps with our disclosure 10-K dataset in Section **??**. We drop cases in which the bidder already owns more than 50% of the target shares prior to the announcement of the bid, in order to exclude acquisitions of minority interest in the target or stock repurchases. We also drop a bid if either of the following conditions are satisfied: 1- The percentage of shares that the bidder is seeking to acquire is less than 50% of the target shares, 2- If such percentage information is missing, and 3- If the fraction of shares held by the bidder after a completed transaction is less than 50%. We also drop observations that SDC labels as block purchases, creeping acquisition, privatization.

This definition of a takeover is standard in the literature (see Betton, Eckbo, and Thorburn, 2008). Over 1995-2019 period, we have xxx takeover bids. However, many of these takeover bids involve non-US or private that do not file 10-K with the US Securities and Exchange Commission (SEC). There are only xxx takeovers where at least one of the parties (target or acquirer) is public. Lastly, we focus on transactions where we have both financial data (via Compustat) and text data (via the 10-K annual reports) so that we can fairly compare our text model with existing studies and baseline approach purely on financial data. The final size of the data is summarized in Table 4.1.

4.2 Financial Data

The financial data is from Compustat. The specific explanatory variables we use are the usual and standard ones in this literature: Tobin's Q (ratio of the market value of company assets to book value), PPE (the book value of property plant and equipment), log of cash balance, the size of leverage (book value of debt over book value of assets), size (market value of equity) and return on assets (operating income divided by year-end book value of

Datasets	Number of Firm-Year Observations	Number of Acquirers	Number of Targets
Training (for parameter estimation)			
Development (for hyperparameter tuning)			
Test (for measuring R^2)			

Table 3: Summary of datasets used in this research

assets). For ease of interpretation we standardize these variables to have mean 0 and variance 1.

4.3 Form 10-K Text Corpus

Our text data comes from the annual report, the 10-K" that each publicly company files with the SEC. Inside the 10-K is a section called Management's Discussion and Analysis (MD&A). The MD&A section is where management reviews the past year's financial and other results and discusses forecasts of the future. Using MD&A section is in line with previous work [23].

From 1995 to 2019 we have xxx firm-year observations.

In this paper, I use machine learning techniques in order to use a new dataset to analyze merger activity - a firm's annual 10K SEC statements. This 10K statement contains description about firm and firm products that will capture difficult to observe variables that could perform quite well with explaining mergers.

I find that the lasso and the ridge regularization techniques have found not only words that align with previous merger theory, but also some interesting variables that were not previously considered. Using this new technique, I obtained a predictive model that yields an R-squared that ranges from 0.02 to 0.07.

[12] We examine how product similarity and competition influence mergers and acquisitions and the ability of firms to exploit product market synergies through asset complementarities. Using novel text-based analysis of firm 10K product descriptions, we find three key results. 1) Firms are more likely to enter mergers with firms whose language describing their assets is similar. 2) Transactions in competitive product markets with similar acquirer and target firms experience increased stock returns and real longer-term gains in

cash flows and higher growth in their product descriptions. 3) These gains are higher when the target is less similar to the acquirer's closest rivals, and when firms have the potential for unique products. Our findings are consistent with firms merging and buying assets to exploit asset complementarities and to create new products to increase product differentiation.

Our analysis is two-fold. We apply modified BERT to financial text prediction. We also present analysis similar to [19].

5 Introduction

This result is in line with the Q-theory of mergers and acquisitions [15], which predicts that profitable companies are eager to acquire poor-performing firms and generate operational gains by putting their assets to a more productive use.

[17]

[16] This study examines the predictive power of textual information from S-1 filings in explaining IPO underpricing. Our empirical approach differs from previous research, as we utilize several machine learning algorithms to predict whether an IPO will be underpriced, or not. We analyze a large sample of 2,481 U.S. IPOs from 1997 to 2016, and we find that textual information can effectively complement traditional financial variables in terms of prediction accuracy. In fact, models that use both textual data and financial variables as inputs have superior performance compared to models using a single type of input. We attribute our findings to the fact that textual information can reduce the ex-ante valuation uncertainty of IPO firms, thus leading to more accurate estimates.

[23]

[5]

Mergers and acquisitions (M&As) play a key role in the economy. At the aggregate level, M&A transactions represent the main mechanism for consolidation and restructuring within industries, and their value as a fraction of U.S. GDP is substantial; 5.8% between 1980 and 2011 (according to M&A volume is computed from SDC Platinum data as the aggregate value of completed acquisitions of U.S. target companies). At the individual company level, takeovers constitute major investment decisions and an effective way to discipline ineffcient managers. Given their importance in the economy, it is no surprise that M&As have attracted a great deal of attention among researchers|there is a wide body of theoretical and empirical research surrounding mergers.

5.1 Related Work

Using textual analysis to identify merger participants: Evidence from the U.S. banking industry [?].

Confining value from neural networks A sectoral study prediction of takeover targets in the US technology sector

Language models have become a key step to achieve state-of-the art results in many different Natural Language Processing (NLP) tasks. Leveraging the huge amount of unlabeled texts nowadays available, they provide an efficient way to pre-train continuous word representations that can be fine-tuned for a downstream task, along with their contextualization at the sentence level. This has been widely demonstrated for English using contextualized representations [6, 20, 13, 21, 7, 25].

Natural language processing tasks, such as question answering, machine translation, reading comprehension, and summarization, are typically approached with supervised learning on task specific datasets. We demonstrate that language models begin to learn these tasks without any explicit supervision when trained on a new dataset of millions of webpages called WebText. When conditioned on a document plus questions, the answers generated by the language model reach 55 F1 on the CoQA dataset - matching or exceeding the performance of 3 out of 4 baseline systems without using the 127,000+ training examples. The capacity of the language model is essential to the success of zero-shot task transfer and increasing it improves performance in a log-linear fashion across tasks. Our largest model, GPT-2, is a 1.5B parameter Transformer that achieves state of the art results on 7 out of 8 tested language modeling datasets in a zero-shot setting but still underfits WebText. Samples from the model reflect these improvements and contain coherent paragraphs of text. These findings suggest a promising path towards building language processing systems which learn to perform tasks from their naturally occurring demonstrations [22]

There are previous methods that use Machine learning to predict stock markets. [8] proposes a deep learning method for event-driven stock market prediction. First, events are extracted from news text, and represented as dense vectors, trained using a novel neural tensor network. Second, a deep convolutional neural network is used to model both short-term and long-term influences of events on stock price movements.

The use of robo-readers to analyze news texts is an emerging technology trend in computational finance. In recent research, a substantial effort has been invested to develop sophisticated financial polarity-lexicons that can be used to investigate how financial sentiments relate to future company performance. However, based on experience from other fields, where sentiment analysis is commonly applied, it is well-known that the overall semantic orientation of a sentence may differ from the prior polarity of individual words.

Financial risk, defined as the chance to deviate from return expectations, is most commonly measured with volatility. Due to its value for investment decision making, volatility prediction is probably among the most important tasks in finance and risk management. Although evidence exists that enriching purely financial models with natural language information can improve predictions of volatility, this task is still comparably underexplored. We introduce PRoFET, the first neural model for volatility prediction jointly exploiting both semantic language representations and a comprehensive set of financial features. As language data, we use transcripts from quarterly recurring events, so-called earnings calls; in these calls, the performance of publicly traded companies is summarized and prognosticated by their management. We show that our proposed architecture, which models verbal context with an attention mechanism, significantly outperforms the previous state-of-the-art and other strong baselines. Finally, we visualize this attention mechanism on the token-level, thus aiding interpretability and providing a use case of PRoFET as a tool for investment decision support [24].

References

[1] Sandra Betton, B. Eckbo, and Karin Thorburn. Corporate takeovers. *Handbook of Corporate Finance: Empirical Corporate Finance*, 2, 05 2008.

[2] Andriy Bodnaruk, Tim Loughran, and Bill McDonald. Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, 50(4):623–646, 2015.

[3] M Buehlmaier and TM Whited. Looking for risk in words: a narrative approach to measuring the pricing implications of financial constraints. 2015.

[4] K. J. Martijn Cremers, Vinay B. Nair, and Kose John. Takeovers and the Cross-Section of Returns. *The Review of Financial Studies*, 22(4):1409–1445, 04 2008.

[5] Linlin Cui, Jianhua Chen, and Wentao Wu. Predicting secondary equity offerings (seos) using machine learning. In M. Arif Wani, Mehmed M. Kantardzic, Moamar Sayed Mouchaweh, João Gama, and

Edwin Lughofer, editors, *17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA, December 17-20, 2018*, pages 1219–1224. IEEE, 2018.

[6] Andrew M. Dai and Quoc V. Le. Semi-supervised sequence learning. NIPS'15, page 3079–3087, Cambridge, MA, USA, 2015. MIT Press.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[8] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, page 2327–2333. AAAI Press, 2015.

[9] B. Espen Eckbo. Corporate takeovers and economic efficiency. *Annual Review of Financial Economics*, 6(1):51–74, 2014.

[10] ALEX EDMANS, ITAY GOLDSTEIN, and WEI JIANG. The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance*, 67(3):933–971, 2012.

[11] Gerard Hoberg and Vojislav Maksimovic. Redefining Financial Constraints: A Text-Based Analysis. *Review of Financial Studies*, 28(5):1312–1352, 2015.

[12] Gerard Hoberg and Gordon Phillips. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies*, 23(10):3773–3811, 08 2010.

[13] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[14] Narasimhan Jegadeesh and Di Wu. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712 – 729, 2013.

[15] Boyan Jovanovic and Peter L. Rousseau. The q-theory of mergers. *American Economic Review*, 92(2):198–204, May 2002.

[16] Apostolos Katsafados, Ion Androutsopoulos, Ilias Chalkidis, Manos Fergadiotis, George Leledakis, and Emmanouil Pyrgiotakis. Textual information and ipo underpricing: A machine learning approach, 10 2020.

[17] Apostolos G. Katsafados, Ion Androutsopoulos, Ilias Chalkidis, Emmanouel Fergadiotis, George N. Leledakis, and Emmanouil G. Pyrgiotakis. Using textual analysis to identify merger participants: Evidence from the U.S. banking industry. MPRA Paper 96893, University Library of Munich, Germany, November 2019.

[18] Tim Loughran and Bill McDonald. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2):307–326, 2013.

[19] R. Moriarty, H. Ly, E. Lan, and S. K. McIntosh. Deal or no deal: Predicting mergers and acquisitions at scale. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 5552–5558, Dec 2019.

[20] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[21] A. Radford. Improving language understanding by generative pre-training. 2018.

[22] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[23] Bryan R. Routledge, Stefano Sacchetto, and N. A. Smith. Predicting merger targets and acquirers from text. 2013.

[24] Christoph Kilian Theil, Samuel Broscheit, and Heiner Stuckenschmidt. Profet: Predicting the risk of firms from event transcripts. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5211–5217. International Joint Conferences on Artificial Intelligence Organization, 7 2019.

[25] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, pages 5753–5763. Curran Associates, Inc., 2019.