U. PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

Predicting and understanding rankings of financial analysts

Artur Aiguzhinov

Tese de Doutoramento em Economia

Supervised by:

Carlos Soares
Ana Paula Serra

2015

Resumo

Os rankings dos analistas financeiros são amplamente utilizados por fornecedores de informação financeira para avaliar o desempenho dos analistas em termos das suas recomendações de investimento, preços-alvo, e estimativas de resultados. Por um lado, os analistas no topo do ranking recebem maior reconhecimento entre os seus pares e têm mais atenção de potenciais investidores. Por outro lado, os fornecedores da informação financeira utilizam os rankings para criar recomendações "inteligentes" e vendê-las aos investidores. Assim, os rankings dos analistas são importantes e relevantes para muitos participantes no mercado. Neste trabalho mostramos que é possível prever os rankings e usá-los para delinear uma estratégia de investimento que resulta em retornos ajustados pelo risco positivos acima dos que seriam gerados por uma estratégia de investimento no mercado ou baseada em estimativas de consenso. O trabalho realizado inclui ainda a caracterização do desempenho relativo dos analistas; e a identificação das variáveis que mais contribuíram para alterações nos rankings. Para responder a estas questões, adaptámos e aplicámos um algoritmo de Machine Learning que modeliza a relação entre variáveis económicas e os rankings. O algoritmo baseia-se no conceito de semelhança entre os rankings e utiliza uma probabilidade de "label ranking", obtido a partir da transformação Bayesiana, para prever o ranking mais provável dado um conjunto das variáveis independentes. Em resumo, a contribuição de nossa trabalho é a seguinte: primeiro, mostramos que as estratégias de investimento baseadas nos rankings dos analistas superam aquelas baseadas mas estimativas de consenso dos analistas; segundo, adaptamos um algoritmo de classificação para resolver um problema de previsão dos rankings; terceiro, analisamos a capacidade explicativa das variáveis económicas de estado e identificamos as que determinam alterações nos rankings; finalmente, conseguimos prever os rankings dos analistas e mostramos que as estratégias de investimento baseadas em previsões de rankings superam aqueles baseados em rankings previstas sem modelo.

Abstract

Rankings of financial analysts are widely used by financial research vendors to evaluate the analysts' performance in terms of their recommendations, price target accuracy, and earnings forecasts accuracy. On one hand, top analysts receive a greater acknowledgment among their peers and more attention from the potential investors. On the other hand, financial information vendors utilize the rankings to create smart recommendations and sell them to investors. Thus, rankings of financial analysts are important and relevant for many financial market players. In this work we show that it is possible to predict rankings of financial analysts and use these rankings in active trading strategies with risk-adjusted returns above market returns and those that would result from using consensus estimates. We also address the problem of characterizing the general behavior of analysts' relative performance and identifying the variables that contributed the most to changes in rankings. We solve these tasks by adapting and applying a Machine Learning label ranking algorithm that models a relation between the state variables and the rankings. The algorithm relies on the concept of ranking similarity and uses a label ranking probability, obtained from the Bayesian transformation, to predict the most probable ranking given a set of descriptive independent variables. In summary, the contribution of our research is four-fold. First, we show that trading strategies based on analysts' rankings outperform those based on analysts' consensus estimates; second, we adapt a *naive Bayes* classification algorithm to solve a label ranking problem; third, we analyze state variables and identify the most contributive to changes in rankings; finally, we predict the rankings of financial analysts and show that the trading strategies based on model-predicted rankings outperform those base on a non-model rankings.

Contents

List of Figures

List of Tables

References

Aiguzhinov, A., C. Soares, and A. Serra (2010). A similarity-based adaptation of naive Bayes for label ranking: Application to the metalearning problem of algorithm recommendation. In B. Pfahringer, G. Holmes, and A. Hoffmann (Eds.), *Discovery Science*, Volume 6332 of *Lecture Notes in Computer Science*, pp. 16–26. Springer Berlin Heidelberg.

Black, F. and R. Litterman (1992). Global portfolio optimization. *Financial Analysts Journal 48*(5), 28–43.

Bouckaert, R. R. (2005). Naive Bayes classifiers that perform well with continuous variables. In *AI 2004: Advances in Artificial Intelligence*, Volume 3339 of *Lecture Notes in Computer Science*, pp. 85–116.

Brazdil, P., C. Soares, and J. Costa (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning 50*(3), 251–277.

Brazdil, P., C. Soares, C. Giraud-Carrier, and R. Vilalta (2009). *Metalearning Applications to Data Mining*. Springer.

Cheng, W., K. Dembczynski, and E. Hüllermeier (2010). Label ranking methods based on the Plackett-Luce model. In J. Fürnkranz and T. Joachims (Eds.), *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Omnipress.

Cheng, W., J. Hühn, and E. Hüllermeier (2009). Decision tree and instance-based learning for label ranking. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, New York, NY, USA, pp. 161–168. ACM.

Cheng, W. and E. Hüllermeier (2009). A new instance-based label ranking approach using the Mallows model. In W. Yu, H. He, and N. Zhang (Eds.), *Advances in Neural Networks – ISNN 2009*, Volume 5551 of *Lecture Notes in Computer Science*, pp. 707–716. Springer Berlin Heidelberg.

de Sá, C., C. Soares, A. Jorge, P. Azevedo, and J. Costa (2011). Mining association rules for label ranking. In J. Huang, L. Cao, and J. Srivastava (Eds.), *Advances in Knowledge*

Discovery and Data Mining, Volume 6635 of *Lecture Notes in Computer Science*, pp. 432–443. Springer Berlin Heidelberg.

Dekel, O., C. Manning, and Y. Singer (2004). Log-linear models for label ranking. In *Advances in Neural Information Processing Systems 16*.

Domingos, P. and M. Pazzani (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning 29*(2), 103–130.

Grbovic, M., N. Djuric, and S. Vucetic (2012). Learning from pairwise preference data using gaussian mixture model. *Preference Learning: Problems and Applications in AI*, 33.

Gurrieri, M., X. Siebert, P. Fortemps, S. Greco, and R. Słowiński (2012). Label ranking: A new rule-based label ranking method. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. Yager (Eds.), *Advances on Computational Intelligence*, Volume 297 of *Communications in Computer and Information Science*, pp. 613–623. Springer Berlin Heidelberg.

Har-Peled, S., D. Roth, and D. Zimak (2002). Constraint classification: A new approach to multiclass classification. In N. Cesa-Bianchi, M. Numao, and R. Reischuk (Eds.), *Algorithmic Learning Theory*, Volume 2533 of *Lecture Notes in Computer Science*, pp. 365–379. Springer Berlin Heidelberg.

Hüllermeier, E., J. Fürnkranz, W. Cheng, and K. Brinker (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence 172*(2008), 1897–1916.

Lebanon, G. and J. Lafferty (2002). Cranking: Combining rankings using conditional probability models on permutations. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 370. Morgan Kaufmann Publishers Inc.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ribeiro, G., W. Duivesteijn, C. Soares, and A. Knobbe (2012). Multilayer perceptron for label ranking. In A. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2012*, Volume 7553 of *Lecture Notes in Computer Science*, pp. 25–32. Springer Berlin Heidelberg.

Todorovski, L., H. Blockeel, and S. Dzeroski (2002). Ranking with predictive clustering trees. In T. Elomaa, H. Mannila, and H. Toivonen (Eds.), *Machine Learning: ECML 2002*, Lecture Notes in Computer Science, pp. 444–455. Springer Berlin Heidelberg.

Vembu, S. and T. Gärtner (2010). Label ranking algorithms: A survey. In J. Fürnkranz and E. Hüllermeier (Eds.), *Preference Learning*, pp. 45–64. Springer.

Vogt, M., J. Godden, and J. Bajorath (2007). Bayesian interpretation of a distance function for navigating high-dimensional descriptor spaces. *Journal of Chemical Information and Modeling 47*(1), 39–46.