Avoiding the Blind Spots: Competitor Identification Using Web Text and Linkage Structure

Gautam Pant
Department of Operations and Information Systems
The University of Utah
gautam.pant@business.utah.edu

Olivia R. Liu Sheng
Department of Operations and Information Systems
The University of Utah
olivia.sheng@business.utah.edu

Abstract

The importance of identifying competitors and of avoiding "competitive blind spots" (Zajac and Bazerman 1991) in marketplace has been well emphasized in research and practice (Walker et al. 2005). However, identification of competitors is non-trivial and requires active monitoring of a focal company's competitive environment. The difficulty in such identification is amplified manifold when there are many more than one focal company of interest (e.g., an analyst exploring the competitive environment of a portfolio containing hundreds of companies) (Ma et al. 2009). Several works have described the difficulties, both from cognitive and procedural standpoint, in identifying competitors. While some have highlighted the role of mental models and taxonomy (Porac and Thomas 1990) in identifying competitive threats, others have highlighted the "managerial myopia" (Bergen and Peteraf 2002) in recognizing competitors.

As the web presence of companies, their clients/consumers, and their suppliers continues to grow, it is increasingly realistic to assume that the real-world competitive relationships are reflected in the text and linkage structure of the relevant pages on the web. However, finding the appropriate web-based cues that effectively signal competitor relationships remains a challenge (Bao et al. 2008, Ma et al. 2009). Typically, competitor identification, which is a necessary first step for competitor analysis and strategy, is based on "supply-side" and the "demand-side" considerations (Chen 1996, Bergen and Peteraf 2002, Desarbo et al. 2006). For example, if two companies depend on similar resources and technology for production (say iron ore) then they may be considered competitors based on the supply-side substitutability. On the other hand if two companies produce products that address similar needs of the consumers (say personal computing) then they may be considered competitors based on demand-side substitutability. Web sites of companies can be expected to receive links from (i.e., in-links) and also link to (i.e., out-links) their supply-side or demand-side relations. Hence an overlap between in-links and out-links of two companies' web sites may be an indication of their substitutability and therefore a signal of potential competitor relationship. The text around a link (or anchor text) has been used effectively as a description of the destination page in web search (Brin and Page 1998, Attardi et al. 1999, Bradshaw and Hammond 2002) and web crawling systems (Pant and Srinivasan 2006). Aggregated anchor text appearing in different pages that link to a company's web site can be used as a summary of how the company is described by others. An overlap between such third party descriptions of two companies may be used as an estimate of their competitive relationship. Similarly, the content of the web site of a company provides a description of the company (e.g., its various

products and services) by the company itself. An overlap between the self descriptions of companies could also be used as a signal of their substitutability. We study the extent to which competitor relationships are discernable from these and other web-based metrics. We explore the notion that such web cues can allow us to discriminate, in a statistically significant manner, between competitors and non-competitors and also allow us to rank the "degree of competitiveness" of various companies to a given focal company.

Data

Our study is based on web cues related with thousands of companies from the Russell 3000 index. We obtain the web site URL for each company from Yahoo Finance. In addition we obtain up to top 500 in-links (about 1.4 million in total) to each of the company web sites using the Yahoo Boss API. Also, we download the pages from each of the web sites using a multi-threaded web crawler that is developed in-house. A list of competitors for each of the companies is obtained from Hoovers for evaluation.

References

Attardi, G., A. Gullí, F. Sebastiani. 1999. Automatic Web page categorization by link and context analysis. *Proceedings of THAI-99, 1st European Symposium on Telematics, Hypermedia and Artificial Intelligence*.

Bao, S., R. Li, Y. Yu, Y. Cao. 2008. Competitor mining with the web. *IEEE Transactions on Knowledge and Data Engineering* **20**(10) 1297–1310.

Bergen, M., M. A. Peteraf. 2002. Competitor identification and competitor analysis: A broad-based managerial approach. *Managerial And Decision Economics* **23** 157–169.

Bradshaw, S., K. Hammond. 2002. Automatically indexing documents: Content vs. reference. *Proc. 6th Intl. Conference on Intelligent User Interfaces*. San Francisco, CA.

Brin, S., L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* **30**(1–7) 107–117.

Chen, M.-J. 1996. Competitor analysis and interfirm rivalry: Toward a theoretical integration. *Academy of Managemenl Review* **21**(1) 100–134.

Desarbo, W. S., R. Grewal, J. Wind. 2006. Who competes with whom? a demand-based perspective for identifying and representing asymmetric competition. *Strategic Management Journal* **27** 101–129.

Ma, Z., G. Pant, O. R. L. Sheng. 2009. Mining competitor relationships from online news: A network-based approach. *Working Paper* .

Pant, G., P. Srinivasan. 2006. Link contexts in classifier-guided topical crawlers. *IEEE Transactions on Knowledge and Data Engineering* **18**(1).

Porac, J. F., H. Thomas. 1990. Taxonomic mental models in competitor definition. *Academy of Managremenl Review* **15**(2) 224–240.

Walker, B. A., D. Kapelianis, M. D. Hutt. 2005. Competitive cognition. *MIT Sloan Management Review* .

Zajac, E. J., M. H. Bazerman. 1991. Blind spots in industry and competitor analysis: Implications of interfirm (mis)perceptions for strategic decisions. *Academy of Management Review* **16**(1) 37–56.