THE INTELLECTUAL CORE OF THE ENTREPRENEURSHIP LITERATURE: A LATENT SEMANTIC APPROACH

Xuequn Wang, Dongbei University of Finance & Economics, 217 Jian Shan Street, Dalian, China, 116025, xuequnwang1600@gmail.com
Yibai Li, Washington State University, Todd 420, Pullman, U.S., 99164, yibai@wsu.edu
Jonathan Arthurs, Washington State University, Todd 420, Pullman, U.S., 99164, jarthurs@wsu.edu

ABSTRACT

This study adopts latent semantic analysis to examine a large portion of published entrepreneurship research in order to further the discussion of conceptual convergence of the field and identify the intellectual core of the entrepreneurship discipline. Specifically, the abstracts of all research papers from 1995 through 2010 published in two top entrepreneurship research journals – Journal of Business Venturing and Entrepreneurship Theory and Practice – were analyzed. Our analysis identified seven core research areas: (1) entrepreneurial process and environment; (2) family and ventures; (3) franchising; (4) investment and ventures; (5) startup and resource-related practices; (6) entrepreneurs and ventures; and (7) network and relationship. Our analysis shows that the entrepreneurship academic discipline has maintained a relatively stable research identity that focuses on the entrepreneurial process as well as how ventures interact with entrepreneurs, family, investors, and network.

Keywords: entrepreneurship, literature review, latent semantic analysis, intellectual core

INTRODUCTION

Since the publication of the seminal article by Shane & Venkataraman (2000), entrepreneurship as a field of study has gained increasing legitimacy in the academic community. This article, more than any other, established the entrepreneurial opportunity including its existence, pursuit, and consequent fruition, as the sine qua non of entrepreneurship as a field of study. Successive research has since helped to establish additional boundaries for the field by distinguishing it from closely related areas such as strategy and organizational theory and behavior (Busenitz, West, Shepherd, Nelson, Chandler, & Zacharakis, 2003). While the entrepreneurial opportunity and its concomitant parts (including antecedents and outcomes) remains the point of focus in entrepreneurship, just how the field is developing and its focus is less well known. An interesting paradox concerning entrepreneurship is that although the entrepreneurial opportunity is the foundation for the field, various scholars with vastly different perspectives now publish in the area. These perspectives are also associated with very diverse interests so that the field of entrepreneurship exhibits a fragmented feel. Since entrepreneurship is a very new academic field compared to others, we argue that it is time to take inventory of the increasing research that is being produced to identify areas of convergence in the field. By doing so, we seek to aid the maturing process. The benefit of this maturing process is not necessarily an increasing insularity vis-à-vis other fields of study but rather an increasing reliance on research from other entrepreneurship scholars as the legitimacy of this research increases.
This issue of conceptual maturation and convergence remains a hot topic in entrepreneurship and has led to discussion about what the field should focus on (Davidsson, 2003; Erikson, 2001; Grégoire et al., 2006; Shane & Venkataraman, 2000, 2001; Singh, 2001; Ucbasaran, Westhead, & Wright, 2001; Zahra & Dess, 2001). While many researchers believe that the field of entrepreneurship remains highly fragmented (Aldrich & Baker, 1997; Aldrich, 1992, 2000; Davidsson, Low, & Wright, 2001; Low & MacMillan, 1988; Low, 2001; Wortman, 1987), other research is beginning to help establish the convergence of the field. For example, Grégoire et al. (2006) analyzed the co-citation networks between 1981 and 2004 based on the Frontiers of Entrepreneurship Research series and found that there was some progress of convergence and the field has increasingly begun to rely upon itself. Such effort is important in the sense that the fragmentation may hinder the scholarly development (Brazeal & Herbert, 1999; Venkataraman, 1997). Nevertheless, the debate about the convergence (or maturity) of the field remains important and is likely to continue, and there is a need for empirical analyses to further this discussion.

Focusing on the conceptual convergence, we try to further the discussion of the convergence and help identify the conceptual core of the field by examining the semantic pattern of the abstracts from previous research. Specifically, we analyze the abstracts of 920 articles published between 1995 and 2010 from Entrepreneurship Theory and Practice and Journal of Business Venturing using latent semantic analysis (LSA). There are several contributions of our study. First, we provide further empirical support for the conceptual convergence and help identify the conceptual core of the entrepreneurship literature and research foci. Specifically, we demonstrate that the entrepreneurship literature indeed maintained a certain level of convergence. Second, we show how the field has changed between 1995 and 2010, further confirming that the field is still in the process of development. Third, we illustrate how LSA can be used to provide the insights of conceptual convergence in the entrepreneurship literature and help further the discussion and debate about the field, as well as how the field has evolved and changed. Essentially, we hope to identify the intellectual core to show the invisible college” of the entrepreneurship literature (Venkataraman, 1997, p. 120).

The rest of this paper is organized as follows. First, we briefly discuss previous literature on convergence. Next, we describe how we apply LSA to a collection of entrepreneurship research. Then, we present the findings, focusing on the conceptual core and areas of focus of the entrepreneurship discipline. We conclude with a summary of findings, limitations of our study, and directions for future research.

LITERATURE REVIEW

Our argument for the importance of conceptual convergence is based on the normal science model from Kuhn (1966), who argues that “research (is) firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice” (Kuhn, 1966, p. 10). Here the main assumption is that “knowledge grows linearly as new data are added to the existing stock of research findings” (Astley, 1985, p. 497). Therefore, convergence is especially important because it shows that a certain field develops and progresses through the accumulation of evidence supporting its core paradigm.
As summarized in Grégoire et al. (2006), there are five arguments about consensus in the field. First, there are needs for consensus about the phenomenon examined in the field, including consensus “on the definition of what the field is and is not, on the definition of the research object, and on its main themes” (Bruyat & Julien, 2001, p. 166). Second, there are needs for consensus about the theories used in entrepreneurship research (Low & MacMillan, 1988). Third, there is a need for the consensus in the field’s purpose, practical impact, and unique contribution (Aldrich & Baker, 1997; Low & MacMillan, 1988). Fourth, there are needs for consensus about the methods and measurement used in the field (Chandler & Lyon, 2001). Fifth, empirical investigation is needed to examine the legitimacy of entrepreneurship (Romano & Ratnatunga, 1997). Therefore, achieving consensus at the field level requires continuous efforts, and it may not be possible to address all of these aspects in one study. In our study, we focus on the first category of consensus and try to understand the conceptual core, which represents the key phenomena in entrepreneurship examined. Such a consensus can help researchers build on each other’s work (Greenfield & Strickon, 1986) and help the field cumulate knowledge (Bruyat & Julien, 2001). In the next part, we discuss the method part of our study.

**METHOD**

**Data Collection**

In order to capture the evolutionary panorama of entrepreneurship research, a time frame from 1995 through 2010, a 16-year period, was utilized. This period allows enough time for a particular research topic to go through a large part of its life cycle. We then collected a total of 920 abstracts from all research articles published in Journal of Business Venturing (JBV) and Entrepreneurship Theory & Practice (ETP) during this period. ETP and JBV are widely recognized as two premier entrepreneurship research journals (Dean, Shook, & Payne, 2007). Additionally, according to citation analysis (Bruton, Ahlstrom, & Li, 2010), these two journals have more impact on the field than other journals (Ratnatunga & Romano, 1997). Although some other top management journals such as the Academy of Management Journal (AMJ) also publish high-quality entrepreneurship research (Busenitz et al., 2003). We limited our sample to JBV and ETP whose mission is to publish only entrepreneurship-related manuscripts (Dean et al., 2007) to ensure that the results of the study reflect only “pure” entrepreneurship research.

**Data Analysis**

The data was analyzed by a text mining technique, more specifically, Latent Semantic Analysis. Latent Semantic Analysis is a type of well-accepted text mining technique (Han, Kamber, & Pei, 2011). Text mining is an umbrella term defined as “… the machine supported analysis of text. It uses techniques from information retrieval, information extraction as well as natural language processing (NLP) and connects them with the algorithms and methods of data mining, machine learning and statistics.” (Hotho et al., 2005). It is also defined as “a process that employs a set of algorithms for converting unstructured text into structured data objects, and the quantitative method that analyze these data objects to discover knowledge” (Delen & Crossland, 2008). In recent years, text mining has been increasingly used for knowledge discovery from scholarly literature (Delen & Crossland, 2008; Jensen, Saric, & Bork, 2006a; Mei & Zhai, 2005; Turban, Sharda, Aronson, & King, 2008). The application of text mining is especially fruitful in biologic
and genetic research. Because the number of articles published in those fields is increasing so quickly it is increasingly infeasible for a researcher to keep up-to-date with all of the relevant literature manually, even on specialized topics (Jensen, Saric, & Bork, 2006b). Geneticists even use text mining to identify the complicated linkages between genes and diseases (Hristovski, Peterlin, Mitchell, & Humphrey, 2005). In summary, text mining empowers researchers to discover knowledge from a large amount of literature in a quantitative manner without researchers’ bias.

A Brief Introduction to Latent Semantic Analysis

Among all kinds of text mining techniques, Latent Semantic Analysis (LSA) is a special mathematical and statistical method used to identify the latent concepts within the textual data at the semantic level (Hossain, Prybutok, & Evangelopoulos, 2011). In contrast to many other text mining techniques which analyze textual data at the syntax level by simply counting the occurrence of particular words, LSA is a methodology that can extract the contextual-usage meaning of words and obtain approximate estimates of meaning similarities among words within the given textual data, thus providing the information at the semantic level (Hossain et al., 2011). LSA has numerous applications in natural language processing, search engine and library indexing and many other areas (Hossain et al., 2011).

LSA simulates the way the human brain distills meaning from text (Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008). LSA is capable of identifying underlying concepts within textual data for its particular mathematical method considers not only the word frequency per se but also the contexts in which the particular word is embedded (Sidorova et al., 2008). LSA is based on the fact that multiple words may share the same meaning and one word may mean different things in different contexts. The main concept behind LSA is similar to that of factor analysis. The words that share the same meaning will “load” to their common underlying concept, just like multiple manifest variables will load to their latent factor in factor analysis; one word may “load” to multiple latent concepts other than its main underlying concept, just like one manifest variable may have cross-loadings with multiple latent factors in the factor analysis. Therefore, we can easily draw an analogy between LSA and factor analysis. In fact, LSA, traditional factor analysis and principle component analysis all share the same mathematical method, the singular value decomposition (SVD). The main purpose of SVD in these methods is to reduce the dimensionality of the original data.

LSA generates two sets of loadings, one for the terms (or words) and one for the documents (the abstracts in this study). The term loading shows how individual terms or words load to different latent concepts. Higher term loading reflects the greater chance that the particular term is truly associated with a certain latent concept. Likewise, the document loadings shows how different documents load to different latent concepts. Higher document loading means a greater likelihood that the particular document is truly talking about a certain latent concept. LSA also generates an eigenvalue matrix which shows the importance of all identified latent concepts. A higher eigenvalue is associated with a greater importance of particular latent concepts. It is of the researchers’ choice to decide the cut-off point for eigenvalue, a point below which a latent concept is too “trivial” to be considered in the study. Researchers can adjust the cut-off point to
get different level of the aggregation. At a lower level of aggregation, factors will reveal common research themes and, at a higher level of aggregation, key research areas.

**Factor Interpretation**

Similar to classical factor analysis, we related each factor to its high-loading terms and documents to assist factor interpretation. For each solution, we created a table containing all high-loading terms and documents sorted by absolute loading. We then used these terms and documents to interpret and label the factor. The task of labeling the factors was conducted independently by two researchers. The process consisted of examining the terms and documents (abstracts) related to a particular factor, interpreting the underlying area, and determining an appropriate label. After independently labeling all factors for each solution, the researchers compared their resulting factor labels. 90 percent or more of the factors were given practically identical labels; the researchers then reconciled their differences regarding the other 10 percent or less without controversy.

**Measuring the Strength of Research Themes**

In order to assess how different research themes change over time, we measured the strength of a research theme as the number of documents that load highly on the corresponding factor. These document counts were calculated for four different 4-year segments. In order to calculate the theme strength in a specific 4-year segment, only those documents that were published during the corresponding time period were considered. Results of the analysis are discussed in the next section.

**RESULTS**

**Overview**

As the number of factors decreases, these narrow directions become converged into broader research themes. Among different solutions, we find the 7-factor solution to provide a particularly interesting insight into the key research themes of the entrepreneurship literature, and will explain it in more detail in the discussion section.

**Evolution of the Field**

We examined the dynamics of the entrepreneurship research based on the total number of publications corresponding to each of the research themes during time periods. Our analysis shows the relative strength of the corresponding research themes during different time periods and provides insights into how the nature of the entrepreneurship research has changed over time.

Examination of paper counts suggests that different research themes change differently over time. Some research themes, such as entrepreneurs and ventures (F7.7) and investment and ventures (F7.4), consistently increases over time. On the other hand, franchising (F7.3) keeps decreasing over the same time period. For the rest of the research themes, such as entrepreneurial process
and environment (F7.1), family and ventures (F7.2), startup and resource-related practices (F7.5), and networks and relationships (F7.7), their overall trend is to increase between 1995 and 2010.

**DISCUSSIONS**

**A Multilevel Perspective of Entrepreneurship Research**

Viewing entrepreneurship research at different levels provides valuable insights into the relationship between the diversity of the entrepreneurship literature and its identity. At the lowest level, each paper represents fundamental units of entrepreneurship research (see Figure 1). These papers cover different topics and adopt various theoretical and methodological approaches. Therefore, it can be difficult to get a sense of the entrepreneurship field and key themes by reading these individual papers. On the other hand, by viewing these papers at a higher level of semantic aggregation, the entrepreneurship field can be represented by a large number of research themes.

![Figure 1. A Multilevel Perspective of the Entrepreneurship Research (Adapted from Sidorova et al. (2008))]($)

Therefore, we decrease the number of factors, as shown in the 7-factor solution. This 7-factor solution can help us gain a more concise view of entrepreneurship research and reveal its identity (see the Identity Level in Figure 2). Here a research area or theme usually includes multiple research topics, and an individual research topic may be related to more than one research area.

**Interpreting the Identity of Entrepreneurship Research**

The research themes identified through the 7-factor solution provide a useful and parsimonious view of the conceptual core of entrepreneurship literature (Figure 3). Four of the seven factors from the 7-factor solution can be interpreted as research focusing on the interaction of ventures with different stakeholders: entrepreneurs and ventures (F7.6), family and ventures (F7.2), investment and ventures (F7.4) and networks and relationships (F7.7). Specifically, F7.6: Entrepreneurs and ventures mainly focuses on entrepreneurs’ perception, intention and behaviors before, during, and after venture creation. This research theme includes research topics such as entrepreneurs’ cognition, women entrepreneurs, risk behaviors, and so on. F7.2: Family and
Ventures examines the role of family on ventures and related issues of family business. Here, research focuses on topics such as succession, family representation, family business creation, and so on. Next, F7.4 Investment and Ventures deals with how investors interact with ventures. Research from this theme examines investors’ evaluation, IPO, venture capital investment, and so on. Finally, F7.7 Network and Relationships examines the role of the relationships with entrepreneurs’ team, between entrepreneurs and investors, as well as entrepreneurs’ networks on ventures. Examples of research topics include teams’ cohesion and conflict, the impact of networks on venture creation, network structure and so on.

While these four research themes examine how, and to what effect, different stakeholders interact with ventures, the remaining three research themes deal with different aspects of the entrepreneurial process. Specifically, F7.3 Franchising deals with various issues related to franchising, such as franchising decisions, international franchising, franchisee success and so on. F7.5 Startup and Resource-related Practices examines entrepreneurs’ startups as well as their various practices (e.g., alliance, internalization) to acquire different kinds of resources (e.g., knowledge, fund). Examples include infusion of external capital, alliance intention, knowledge acquisition, technology strategies and so on. Lastly, F7.1 Entrepreneurial Process and Environment examines the process of entrepreneurship as well as how the environment impacts this process. Research topics include the effects of the national culture on entrepreneurial orientation, the determinants of corruption in transition economies, economic freedom and entrepreneurial outcomes, and so on.

Combining the interpretations of the seven research areas, the intellectual core of the entrepreneurship discipline can be summarized as follows (See Figure 2):

The Entrepreneurship academic discipline focuses on the entrepreneurial process as well as how ventures interact with entrepreneurs, family, investors, and networks.

**CONCLUSIONS**
In this study, we attempt to facilitate conceptual convergence in the entrepreneurship discipline by examining a large body of published research and empirically identify key research themes and topics. The key contribution of our study is that it identifies the conceptual core and themes of entrepreneurship research in the form of seven key research areas. The results of our study provide an empirical foundation for the consensus about the conceptual convergence in the entrepreneurship literature. The results of our study concerning the intellectual core can be used to better communicate the nature of entrepreneurship research within the entrepreneurship community and to its external stakeholders.

References


