

# Chapter 12

## Semi-automatic Technology Roadmapping Composing Method for Multiple Science, Technology, and Innovation Data Incorporation

Yi Zhang, Hongshu Chen and Donghua Zhu

**Abstract** Since its first engagement with industry decades ago, technology roadmapping (TRM) is taking a more and more important role for competitive technical intelligence (CTI) in current R&D planning and innovation tracking. Important topics for both science policy and engineering management researchers involves with approaches that refer to real-world problems, explore value-added information from complex data sets, fuse analytic results and expert knowledge effectively and reasonable, and demonstrate to decision makers visually and understandably. The growing variety of Science, Technology, and Innovation (ST&I) data sources in the Big Data Age increases these challenges and opportunities. Addressing these concerns, this paper attempts to propose a semi-automatic TRM composing method to incorporate multiple ST&I data sources—we design an extendable interface for engaging diverse ST&I data sources and apply the fuzzy set to transfer vague expert knowledge to defined numeric values for automatic TRM generation. We focus on a case study on computer science-related R&D. Empirical data from the United States (US) National Science Foundation (NSF) Award data (innovative research ideas and proposals) and Derwent Innovation Index (DII) patent data source (technical and commercial information) affords vantage points at two stages of R&D process and also provide further capabilities for more ST&I data source incorporation. The understanding gained will also assist in description of computer science macro-trends for R&D decision makers.

---

Y. Zhang (✉) · H. Chen  
Decision Systems & e-Service Intelligence Research Lab,  
Faculty of Engineering and Information Technology,  
Centre for Quantum Computation & Intelligent Systems,  
University of Technology Sydney, Ultimo, Australia  
e-mail: yizhang.bit@gmail.com

Y. Zhang · H. Chen · D. Zhu  
School of Management and Economics, Beijing Institute of Technology,  
Beijing, People's Republic of China

**Keywords** Technology roadmapping · Bibliometrics · Fuzzy set · Competitive technical intelligence

## 12.1 Introduction

The growing variety of data sources in the Big Data Age not only increases the challenges and opportunities for competitive technical intelligence (CTI), but also leads the revolution of business management and decision making. Of significance is “data-driven,” which uses Information Technology (IT) to support rigorous, constant experimentation that guides decisions and innovations (Bughin et al. 2010), and has being characterized as one of the most competitive features in various R&D planning and business models (McAfee et al. 2012).

Technology roadmapping (TRM) approaches, described as a representative, prominent, and flexible instrument for long-range technological forecasting and strategic planning, make good sense to actively incorporate business data into planning procedures (Phaal et al. 2004). Quantitative data, especially Science, Technology, and Innovation (ST&I) data, have already been largely involved in TRM models; however, since diverse emphases and possible time gap between different ST&I data sources, traditional TRM model usually only focuses on single ST&I data source, e.g., publications or patents. Undoubtedly, the booming new data sources, e.g., Twitter, news, customer comments, R&D project proposals, and product reports, also match the concept of ST&I researches perfectly, and the rapid engagement of multiple ST&I data with different formats and emphases introduces new challenges for current studies. Meanwhile, the way that current TRM study used to transfer vague human thoughts to defined numerical values (Lee et al. 2011) is still unfavorable, and the combination of qualitative and quantitative methodologies is also not as smart as what we imagine. At this stage, the emerging concerns are the approaches that refer to real-world problems, explore value-added information from complex data sets, fuse analytic results and expert knowledge effectively and reasonable, and demonstrate to decision makers visually and understandably.

Aiming to provide a solution for the questions—(1) how to incorporate multi-dimensional information from diverse ST&I data sources and (2) how to construct an intelligent TRM method, this paper develops a multiple ST&I data incorporation model and a fuzzy set-based semi-automatic TRM generation model. Based on a ST&I topic identification approach (Zhang et al. 2014b) and a traditional TRM composing method (Zhang et al. 2013), this paper designs a multilayer TRM method to arrange topics and related concepts (e.g., idea, technique, and product) to explore potential relationships. It is a challenge to build up an entire intelligent TRM composing method with novel IT techniques, e.g., training machine to discover potential linguistic relationships between technological components; however, this paper seeks approaches to introduce fuzzy set (Zadeh 1965) to transfer vague expert knowledge into defined numeric values and help automatically locate technological components for TRM composition. The empirical study selects the

United States (US) National Science Foundation (NSF) Award data (innovative research ideas and proposals) and the Derwent Innovation Index (DII) patent data source (technical products), which demonstrates the efficiency and feasibility of our methods and also provides vantage points at the top-bottom stages of R&D process and assists in description of computer science macro-trends for decision makers.

The main contributions of this paper are as follows: (1) We provide an approach to incorporate multiple ST&I data into TRM; (2) we construct a systematic approach to apply fuzzy set to the traditional TRM method and help compose TRM in a semi-automatic function; and (3) the process in which we think and solve problems emphasizes the combination of qualitative and quantitative methodologies and is also adaptive and transferrable to related ST&I researches.

This paper is organized as follows: The Related Works section reviews previous studies on TRM. In the Methodology section, we present the detailed research method for the semi-automatic TRM composing method, involving with a multiple ST&I data incorporation model and a fuzzy set-based semi-automatic TRM composing model. The Empirical Study section follows, using the US NSF Awards and DII patent data as the case. Finally, we conclude our current research and outline future work.

## 12.2 Related Works

This section reviews related literatures that include TRM and related ST&I data incorporation studies and also summarizes the limitations of these previous works.

Based on the significant work of Phaal et al. (2004), who summarized previous TRM methods and constructed an effective qualitative composing model, TRM research has already been extended from qualitative study only to a combination of qualitative and quantitative methodologies. Representatively, Huang et al. (2014) introduced a bibliometric technique-based four-dimensional TRM for the science and technology planning of China's solar cell industry; Zhang et al. (2014c) combined TRM model with Triple Helix innovation and Semantic TRIZ concepts and presented an empirical study on China's dye-sensitized solar cell industry; Lee et al. (2015) proposed a scenario-based TRM that involved with a plan assessment map and an activity assessment map for organizational plans, which also engaged the Bayesian network for topology and a causal relations definition. Geum et al. (2015) added the association rule mining to TRM to identify relationships between items on different layers of TRM.

The current TRM model has been combined with various concepts and methods, applying into real ST&I assessment, forecasting, and planning. According to previous studies, it is promising to conclude the benefits of current TRM studies as follows: (1) TRM is a visual model to present content, which enables easy understanding for both macro-, meso-, and micro-level problems (Zhang et al. 2013); (2) the hierarchical structure of TRM helps indicate potential relationships between items on different layers (Geum et al. 2015); (3) TRM is able to take dimensional impact factors into consideration, e.g., time, science policy, market pull, and

technique push (Lee and Park 2005; Robinson and Propp 2008; Huang et al. 2014); and (4) previous TRM studies have perfect adaptability for general publication and patent data sources and match requests from different industrial domains.

Multiple ST&I data incorporation for TRM always holds great interests for ST&I analyzers. Data incorporation, known as data integration or data fusion, is one important branch of data mining and relies on strong computer science and IT background. Current research is to combine data residing at different sources and provide users with an unified view of these data (Lenzerini 2002). However, these data incorporation studies focus on data structure problem and ignore underlying insights and interactions of different data sources, which in contrast are the key factors for ST&I analyses. On the other hand, it has been a long time since the idea of multiple ST&I data incorporation generated, but until now related researches are rare. Robinson and Propp (2008) designed a multipath roadmapping framework where four layers—research lines, experimental integration, integrated platform/product, and application area—were used to emphasize the different phases of technology development. Zhang et al. (2013) engaged core terms derived from Web of Science (WoS) publications and International Classification Code (IPC) retrieved from the United States Patent Trademark Office (USPTO) into one TRM model to highlight the technological researches and applications.

Another tough task of current TRM study is the approach transferring vague human thoughts to defined numerical values (Lee et al. 2011), and this issue will heavily influence the efficiency and accuracy of TRM auto-generation process. In this context, fuzzy set would be a helpful instrument to minimize expert aid and consumed time, but maximize the usage of expert knowledge. Lu et al. (2011) constructed a novel group decision-making method with fuzzy hierarchical criteria for theme-based comprehensive evaluation in new product development, which calculated ranking results by fusing all assessment data from human beings and machines. Lee et al. (2011) introduced a fuzzy analytic hierarchy process to TRM model, which was definitely an innovative attempt for the combination of fuzzy concepts and TRM, although the empirical study only applied to a small-range data set (five sub-technologies of hydrogen energy) with expert ranking.

### 12.3 Methodology

On the purpose of providing solutions for constructing a multilayer TRM to reveal multidimensional information from emphasis-differed ST&I data sources and to explore insights for technical intelligence understandings, this paper constructs a TRM method for multiple ST&I data incorporation with the following steps:

1. Inputs: Grouped Topics—we focus on raw ST&I data and retrieve meaningful phrases and terms via a Term Clumping process (Zhang et al. 2014a) and then apply a K-means-based clustering approach (Zhang et al. 2014b) to identify hot research topics as key technological components;

2. Step 1: Multiple ST&I data incorporation model—we construct a designing process for multiple ST&I data incorporation that introduces Technology Readiness Level to analyze the emphases of ST&I data and proposes a questionnaire for expert consultation.
3. Step 2: Fuzzy set-based semi-automatic TRM generation model—in order to combine the qualitative and quantitative methodologies, we engage experts to evaluate topics by removing meaningless topics, consolidating duplicate topics, and highlighting significant ones, and we also evaluate each topic and group them into specified fuzzy sets, after which we generate the TRM in an automatic manner;
4. Outputs: TRM.

Note that the both steps would think about the combination of qualitative and quantitative methodologies. The step 1 is to consider the diverse emphases of applied ST&I data and data analyst would take an active role for the design process, while domain experts would dominate the step 2 for evaluating selected topic candidates. The framework of the TRM method for multiple ST&I data incorporation is given in Fig. 12.1.

### ***12.3.1 Inputs: Grouped Topics***

How to retrieve meaningful phrases and terms from raw ST&I data and identify valuable topics via clustering analysis are definitely interesting research questions, but they are not the foci of this paper. Briefly, we apply a revised version of the Term Clumping process (Zhang et al. 2014a) to retrieve phrases and terms from ST&I textual data by term removal, consolidation, and clustering, and then, a K-means-based clustering approach (Zhang et al. 2014b) is used to group related linguistic elements, e.g., phrases and terms, or records, into meaningful topics. We define these grouped topics as technological components, which reflect scientific or technical information of ST&I data.

### ***12.3.2 Step 1: Multiple ST&I Data Integration***

It is a general understanding that the emphases of different ST&I data are diverse. As shown in Fig. 12.2, we summarize the emphases of the selected mainstream ST&I data sources as below:

1. Academic proposal, e.g., NSF proposals, is usually granted by national government to support academic institutions for basic research, whose content focuses on new ideas, concepts, and unrealized innovative actions. Discoveries derived from academic proposal would be an express way to dive into innovation and promising for both academies and industries.

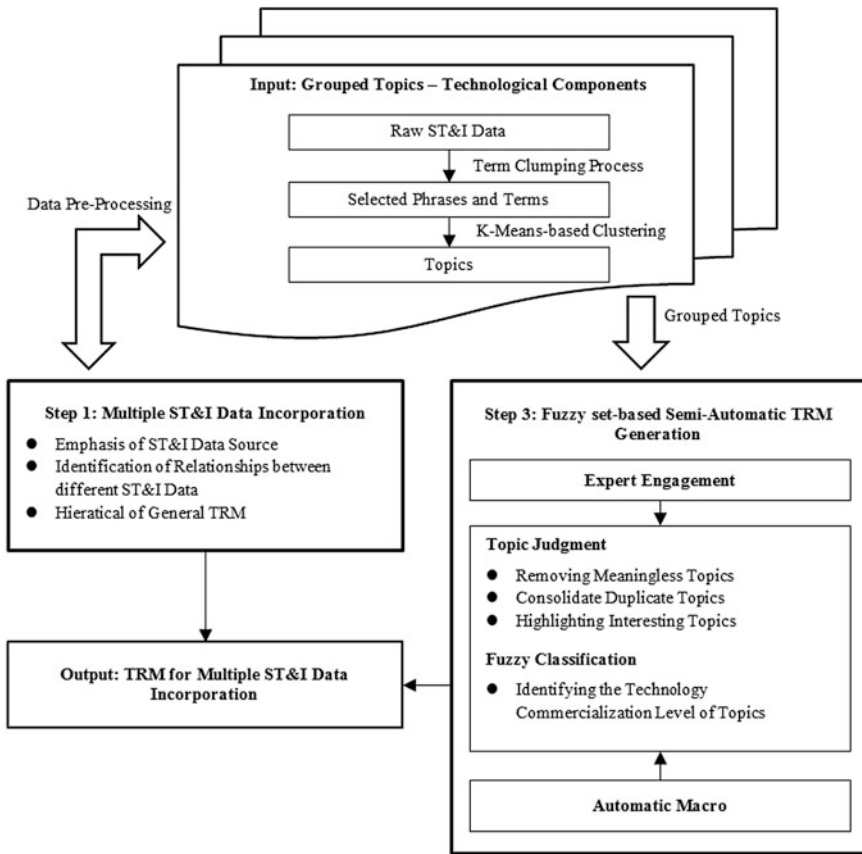


Fig. 12.1 Framework of technology roadmapping method for multiple ST&I data incorporation

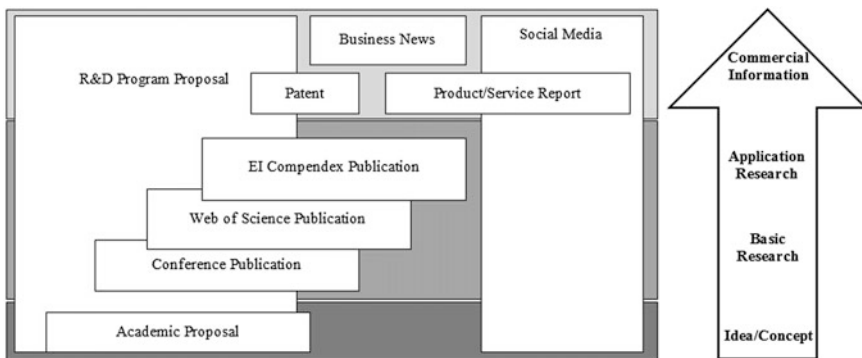


Fig. 12.2 Emphasis of selected mainstream ST&I data sources

2. Publication contributes to both basic research and application research, but, in detail, conference paper, e.g., IEEE paper, mostly presents draft research frameworks, experimental results, or mature ideas, while publications in Web of Science including SCI and SSCI data, and the EI Compendex indexed papers emphasize fundamental research and application research, respectively.
3. Patent, e.g., DII patent, contributes to actual applications or products, the same as detailed technical report or guidebook of business services. In particular, patent terms should be vague and address legal effects (with patent barrier), and, comparably, academic terms are more clear and direct.
4. Business news, e.g., Factiva data, highlights social significances, where common technical terms and modifying adjectives will heavily influence the description.
5. Social media data, e.g., Twitter, are similar to the content of news, but it presents information in a more informal way. Sentimental analysis, sometimes, is applied to understand real meanings of these free texts.
6. National R&D program proposal, comparing with the five ST&I data above, is a complex one, which would include all aspects of ST&I emphases. Therefore, it would be not a smart option to incorporate national R&D programs with other ST&I data, but one feasible solution is to apply national R&D program proposal as a comparable case for other ST&I data.

The time gap problem is always the barrier that prevents incorporating multiple ST&I data effectively. It takes time to transfer an innovative idea into feasible plans and valid experiments, and related patents and mature products would be years or even decades after the idea firstly comes out. The time would be shorter and more unpredictable for emerging technologies. In this context, expert knowledge would be easier to deal with this situation than quantitative computation. Referring to the technology readiness levels (TRL) (Mankins 1995), we design a TRL scale for ST&I data sources, as shown in Table 12.1.

Although the definition of each TRL is still a fuzzy concept and its actual applications would depend on specified technological domains of ST&I data, Table 12.1 proposes a manner to transfer the emphases of ST&I data into an operable format for expert consultation. According to Table 12.1, we would recommend selecting ST&I data on neighbor TRLs, since the possible time gap would be tiny and could be ignored. However, if the time gap is able to be fully handled and considered, a comparison between the ST&I data sources at TRL 1 and TRL 5 would also make good sense to see an entire technological evolutionary pathway. Furthermore, based on the revised TRL scale, we design a questionnaire for multiple ST&I data incorporation in Table 12.2.

We attempt to provide a logical work flow to organize expert in a workshop process. The questions in Table 12.2 aim to lead the discussions and gain valuable information effectively. Note that what we list are only options for possible case studies, and we need to revise and refine this questionnaire to match actual requirements.

**Table 12.1** Technology readiness level scale for ST&I data source

TRL	Definition	Emphases of ST&I data
1	Technology concept and/or application formulated	Innovative ideas/concepts
2	Analytic and experimental critical function and/or characteristic proof-of-concept	Basic research
3	Component and/or breadboard and/or system/subsystem model and/or prototype demonstration in a laboratory or relevant environment	Application research
4	Actual system completed and “capability qualified” through test and demonstration	Application research/products
5	Actual system “capability proven” through successful mission operation	Products

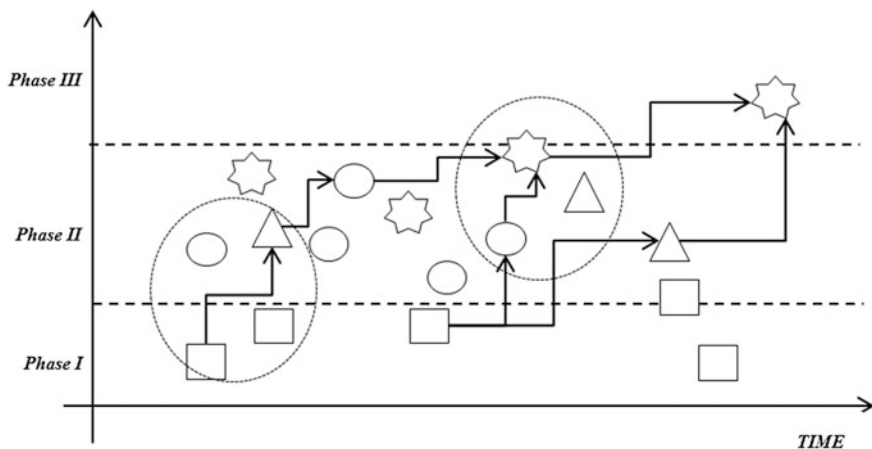
**Table 12.2** The questionnaire for multiple ST&I data incorporation

No.	Focus	Questions
1	ST&I data	Which ST&I data sources shall we use?
		Is there any possible conflict between applied ST&I data?
2	Data structure	How to unify the structure of applied ST&I data?
3	Emphases	What are our emphases of this study?
		Do we need to forecast?
		How to rank these emphases and what is the priority one?
4	Time interval	Shall we look backward, forward, or both?
		How long the time interval will be?
5	TRL	How the TRL distance between the emphases of applied ST&I data?
		How to evaluate the possible time gap among applied ST&I data?
		Is the time gap at an acceptable level?
6	Layer	Shall we divide applied ST&I data into separated layers or construct a multilayer TRM?
		What is the criterion for classifying layers?

Based on the hierarchical landscape and the expression form of components in our previous TRM models (Zhang et al. 2013; Zhou et al. 2014), we enrich the structure by the following: (1) distinguishing the scope of ST&I data sources with the shape of components, (2) softening the definition of *Y* axis to reserve an interface with the configuration of fuzzy set, and (3) defining linkages between components with multifactors, e.g., semantic similarity, time, science, policy. A sample of TRM for multiple ST&I data incorporation is given as in Fig. 12.3.

In Fig. 12.3, time is marked as the *X* axis, while *Y* axis is used for multilayer demonstration. Shaped components indicate topics derived from different ST&I data, and they are located among the multiple layers of TRM, and grouped components





**Fig. 12.3** The sample of TRM for multiple ST&I data incorporation

would share similar topics with “some” possible linkages. It is also available to locate components on separated layers to highlight their origin ST&I data and relationships between different layers. In usual, experts will help to identify the macro-level linkage between components to track possible evolutionary pathways.

### 12.3.3 Step 2: Fuzzy Set-Based Semi-automatic TRM Generation

The balance between qualitative and quantitative methodologies in TRM composing model is definitely an intriguing and complicated research topic at the current time. Expert knowledge has been largely engaged with results refinement, component allocation, and the understanding of TRM for decision making (Zhang et al. 2015). At this stage, we aim to minimize the aid of experts and maximize the usage of expert knowledge in the limited time and scope, so that the fuzzy set can be considered as an effective tool to deal with this issue. As mentioned above, the emphases of ST&I data, the definition of TRL, the evaluation of the time gap, and the criterion for classifying the layers of TRM also depend on subjective expert knowledge. These vague elements definitely match the main concept of fuzzy set and afford possibilities for the engagement of fuzzy set.

As a general definition, we denote “all components” of TRM as the universe  $X = \{x_1, x_2, \dots, x_i, \dots, x_{n-1}, x_n\}$  and “each phase/layer” as a fuzzy set  $A_j$  defined on the  $X$  where  $j \in [1, m]$ . The membership function  $A_j(x_i)$  is considered as the degree that the component  $x_i$  belongs to the phase/layer  $A_j$  and will be decided depending on research purposes and empirical data. The detailed steps are outlined below:

- Considering specific case, to identify  $X$ ,  $A_j$  and  $A_j(x_i)$ ;
- For each component  $x_i$ , experts would help classify into one of the fuzzy sets  $A_j$  and mark a membership grade  $A_j(x_i)$  for the selected fuzzy set;
- Based on  $A_j(x_i)$ , to calculate the  $X(x_i)$  for each component and set as the  $Y$  value;
- To generate TRM automatically via macros.

#### ***12.3.4 Outputs: TRM for Multiple ST&I Data Incorporation***

After our 2-step method, we integrate multiple ST&I data sources, fuse the analytic results with expert knowledge, and generate the graphic TRM as our final output.

### **12.4 Empirical Study**

Computer science would not be still considered as an emerging technology as what we did decades ago, but it has been integrated with IT techniques and various engineering applications and has become a fundamental instrument for multidisciplinary researches. This paper focused on the technology commercialization studies for computer science, and our purpose was to incorporate multiple ST&I data to track the evolutionary pathway of computer science technologies for commercialization. Since the limited condition for time and data sources, this paper only chose the NSF Award data (granted proposals) and the DII patent data for empirical study. Our consideration was the two data sources concentrated on the innovative ideas and mature technical products, respectively, and the contrast on the technology commercialization level would be better to indicate the importance of information fusion for multiple ST&I and also to demonstrate benefits of our TRM method. In addition, we built up our expert base with the help of twelve experts (Associate Professors, Lecturers, Researchers, and Ph.D. Candidates) from the Centre for Quantum Computation & Intelligent System, University of Technology Sydney, Australia, and the Knowledge Management and Data Analysis Laboratory, Beijing Institute of Technology, China.

#### ***12.4.1 Step 1: Incorporation of NSF Awards and DII Patents***

The importance of NSF Awards and DII patents has been discussed separately in our previous studies (Zhang et al. 2014b, c), as shown in Table 12.2, and academic proposal is at the bottom of TRL, while patent belongs to the top of TRL; thus, the

**Table 12.3** Steps of term clumping processing

	Step	NSF awards		DII—2013 <sup>a</sup>	
		#T <sup>b</sup>	#A	#T	#A
1	Number of raw records	12,915		44,141	
2	Natural language processing via VantagePoint (2015)	254,992	17,859	706,739	154,791
3	Basic cleaning with thesaurus	214,172	16,208	679,736	131,577
4	Fuzzy matching	184,767	15,309	–	–
5	Pruning <sup>c</sup>	42,819	2470	19,926	19,930
6	Extra fuzzy matching	40,179	2395	15,510	16,603
7	Computer science based common term cleaning	30,015	2311	14,029	16,529

<sup>a</sup>Considering the larger amount of DII data, we divided it by year for the term clumping process, and therefore, we only present DII data in 2013 as a sample; also, we applied pruning to DII first and, then, fuzzy matching

<sup>b</sup>#T = number of title terms and #A = number of abstract terms

<sup>c</sup>In the pruning process, we removed terms appearing in only one record in the NSF awards and DII titles, but removed terms appearing in less than five records in DII abstracts

large distance between TRL scale introduces not only challenges but also promising insights, and the comparison would help identify the technology commercialization trend in a specified time interval. In this case, we aimed to track the rapid technological changes occurring with the coming of the Big Data Age—how innovative ideas boomed and how applicable techniques evolved; at this stage, we set the time interval from 2009 to 2013 to highlight Big Data-related techniques and their changes. Moreover, we consulted domain experts and decided to apply an integrated multilayer TRM to emphasize the contrast between NSF Awards and DII patents, and the layers included “basic research—TRL 1 and 2,” “application research—TRL 3 and 4,” and “products—TRL 5.”

We grouped the topics of NSF Awards and DII patents separately. We selected 12,915 granted proposals under the Division of Computer and Communication Foundation in the NSF Awards (Zhang et al. 2014b) and 177,974 DII patents with the field Topic and Subject Category as “computer science,” and the field Basic Patent Country and Priority Country as “US.” The revised term Clumping steps were applied for feature extraction, and the process is given in Table 12.3.

We, then, applied the K-means-based clustering model (Zhang et al. 2014b) to group topics and acquire 54 topics from the NSF Awards and 44 ones from the DII patents.

### 12.4.2 Step 2: Fuzzy Set-Based Semi-automatic TRM Generation

According to the three layers for the technology commercialization study, we let the power set  $A = \{A_1, A_2, A_3\}$  and introduced Gaussian distribution to define the

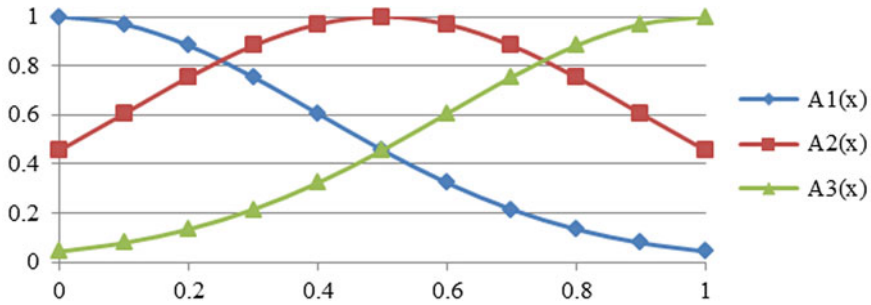


Fig. 12.4 Distribution curves of member functions

membership functions. The three membership functions are provided below, and the distribution curves are shown in Fig. 12.4.

$$A_1(x): X \sim N\left(0, \frac{1}{2\pi}\right), x \in [0, 1]$$

$$A_2(x): X \sim N\left(\frac{1}{2}, \frac{1}{2\pi}\right), x \in [0, 1]$$

$$A_3(x): X \sim N\left(1, \frac{1}{2\pi}\right), x \in [0, 1]$$

As shown in Fig. 12.4, we divided the universe into three intervals, where [0, 0.25] was mapped to “basic research,” while [0.25, 0.75] and [0.75, 1] were for “application research” and “products,” respectively. The experts firstly classified specified topic  $x_i$  into one fuzzy set  $A_j$  and marked the membership grade  $A_j(x_i)$  to the fuzzy set. Then, we calculated the  $X(x_i)$  and assigned the topic  $x_i$  into one of the three fuzzy sets. As a sample, we list parts of the marked topics in Table 12.4, and the generated TRM is given in Fig. 12.5.

Although the experts only assigned each topic into one fuzzy set, we were able to calculate its membership grade for each fuzzy set. As an example, the topic “Video Frames” was assigned to  $A_2(x)$  with the membership grade 0.81, and we got the  $X(x)$  as 0.76, so according to the membership function, it was definitely possible to calculate the membership grade vector of “Video Frames” as (0.18, 0.81, 0.83). At this stage, we re-assigned the 98 topics with the following rules: (1) to classify the topic to the fuzzy set as the First Preference with the largest membership grade, if same membership grades occurred, we preferred to choose the fuzzy set for a lower TRL scale and (2) if the second largest membership grade was not less than 0.7, we set the related fuzzy set as the Second Preference. In this consideration, we found 51 topics in  $A_2(x)$  (42 NSF topics) and 47 topics in  $A_3(x)$  (35 DII topics) with First Preference and three topics in  $A_1(x)$ , all of which belonged to NSF Awards, 13 topics in  $A_2(x)$  (8 NSF topics), and 20 topics in  $A_3(x)$  (13 NSF topics) with Second Preference.

**Table 12.4** Big data-related topics with membership grades of three fuzzy sets

Year	Topic	Topic description	Data	$X(x)$	$A(x)$
2009	Adaptive grasping	Adaptive grasping, automatic speech recognition, empirical mechanism design, hierarchical visual categorization, infinite Bayesian networks	NSF	0.63	$A_2(x) = 0.95$
2010	Reading data	Reading data, RFID tag, tag memory, configuration data, service provider	DII	0.88	$A_3(x) = 0.96$
2011	Solving large systems	Linear equations, parallel strategy, recursive divide, solving large systems	NSF	0.67	$A_2(x) = 0.91$
2011	Remote location	Remote location, retail establishment, source code, information source, navigation database, road sign	DII	0.74	$A_2(x) = 0.83$
2012	Real time	Real time, telecommunication network, advertisement server, mobile communication facility data, monetization platform	DII	0.67	$A_2(x) = 0.91$
2012	Large asynchronous multichannel audio corpora	Large asynchronous multichannel audio corpora, novel speech processing advancements, robotic intelligence	NSF	0.43	$A_2(x) = 0.98$
2013	Video frames	Video frames, encoding video frame, improving video quality, live multicast system, severe degradation	DII	0.76	$A_2(x) = 0.81$
2013	Big data	Algorithm foundation, big data, parsimonious model, mathematical problems	NSF	0.5	$A_2(x) = 1$

### 12.4.3 Findings for the Commercialization of the Computer Science-Related Technologies

In our previous studies, single ST&I data-based TRM took active role in technical intelligence studies, which hold more benefits on exploring inner features of related technologies and identifying technology development chains. However, the incorporation of multiple ST&I data sources makes possible to stand on a higher macro-level to understand technology evolutionary pathways and to discover the gaps during technology development and transfer. We attempted to understand the insights of Fig. 12.5 for R&D plan and technology management concerns, and our findings are given as below:

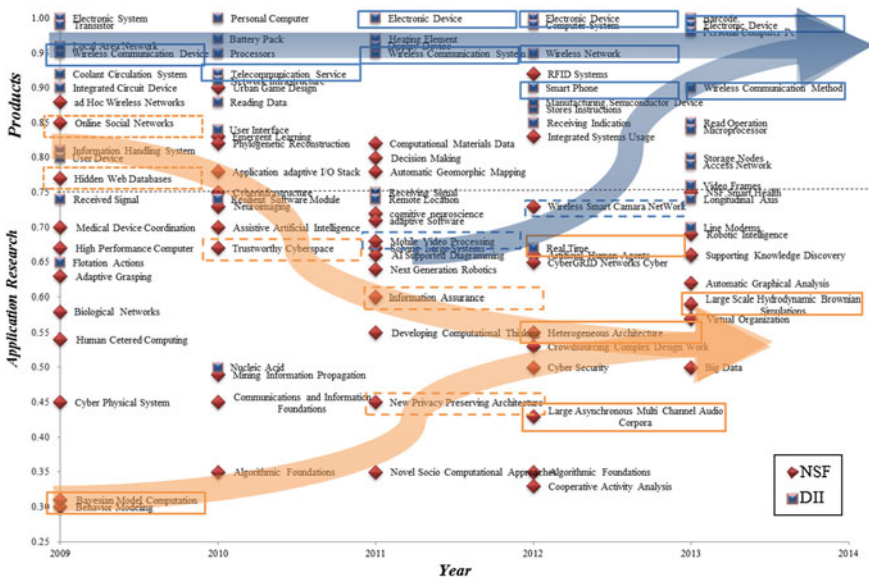


Fig. 12.5 Technology roadmapping for computer science from 2009 to 2013 (based on the incorporation of NSF awards and DII patents)

**The “mobile device” and related techniques were, are, and still will be a hot commercial target in the near future.**

Obviously, the “mobile device” and related techniques (marked as the blue solid box) keep being identified as the hot topics from 2009 to 2013 in the DII patents, which means keen competitions occur or will occur in this filed, and the inventors (including the commercial firms) are seeking the intellectual property protection from the patents. At the same time, not only the mature products reflected in the DII patents, but also some undertaking researches in the NSF researches (marked as the blue dashed box) could be addressed in Fig. 12.5, e.g., “mobile video processing” and “wireless smart camera networks.” Therefore, it is reasonable to conjecture that innovation and advanced techniques will also be engaged and transferred as a strong technical support for the follow-up developments.

**Big Data is not a creation, but a result of technology evolution and fusion, all related techniques of which are able to track down the origins.**

Big Data is an unavoidable topic in recent years. Social media, e.g., Twitter, Facebook, is more popular than any other periods in the history, and the boom of various new techniques, e.g., MapReduce, Hadoop, also illustrates revolutionary changes. In this situation, the voice that highlights the new creation of the Big Data-related techniques would be enough to have its supporters. However, as shown in Fig. 12.5, it is definite to declare that the Big Data could be considered as the results of technology evolution or fusion, and all related techniques are able to

track down the origins. Extending the discussion by Zhang et al. (2014b), social media-led online social networks and web data (marked as the orange-dotted box) constitute parts of the foundation of Big Data, and the coming Big Data age also increases the concerns on the information security (marked as the orange-dashed box) rapidly. On the other hand, the efforts on the improvement of existed algorithms (marked as the orange solid box) have never been stopped, which compose the mainstream techniques of the Big Data Age.

In addition, if we narrow down the scope to technology commercialization, current Big Data-related research still concentrates on the NSF Awards and stands at the fundamental stages that include constructing concepts (e.g., Trust Worth Cyberspace, Real Time) and algorithms (e.g., Bayesian Network Computing, Large Asynchronous Multi-Channel Audio Corpora, Large Scale Hydrodynamic Brownian Simulations), and collecting data and applying it to experimental applications, while Big Data-related business models and real-world applications are crude, even there are no direct related topics in the DII patents. Thus, we should imagine that it will take time to transfer a new technique to commercial practices, and this kind of attempts would be an obvious trend in the near future.

**The process of technology commercialization is much faster than that several years ago.**

It is a common sense that the NSF interests include various fundamental researches that hold potential capabilities on further innovation, and the DII topics only concentrate on applicable techniques with any commercial benefits, e.g., software or hardware techniques. However, considering the three fuzzy sets for technology commercialization studies, only few topics belong to the set “basic research,” and most of them are on the medium level between basic research and products. The possible explanation could be that the current process of technology commercialization is much faster than that several years ago, and new techniques could be used to solve real-world problems in a short time, or as we say, the experimental time is engaging into the commercialization process. In addition, more and more innovations are originated from real-world needs, which might be another strong driving force.

## 12.5 Discussion and Conclusions

Highlighting real-world needs and the engagement of emphasis-differed multiple ST&I data sources, this paper proposes an effective method to (1) incorporate multiple ST&I data to explore value-added information for R&D plans and technology innovation management and (2) introduce the fuzzy set concept to fuse the analytic results and expert knowledge smoothly and, then, help generate the TRM in a semi-automatic model. The thinking that combines qualitative and quantitative

methodologies runs through the whole paper, the attempts on which provide great potential for related expert systems or decision-making processes.

We anticipate further study to look into the following directions: (1) to introduce novel IT techniques, e.g., machine learning, to enrich the semi-automatic model to an entire automatic composing model, and (2) to apply the multiple ST&I data incorporation for real-world applications. In addition, we will also consider the influences resulting from different empirical domains, e.g., emerging technology, social science, and mixed data with multidisciplinary, and address the concerns with more experiments.

**Acknowledgments** This work is supported by the Australian Research Council (ARC) under discovery grant DP140101366 and National High Technology Research and Development Program of China (Grant No. 2014AA015105).

## References

- Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly*, *56*, 75–86.
- Geum, Y., Lee, H., Lee, Y., & Park, Y. (2015). Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping. *Technological Forecasting and Social Change*, *91*, 264–279.
- Huang, L., Zhang, Y., Guo, Y., Zhu, D., & Porter, A. L. (2014). Four dimensional science and technology planning: A new approach based on bibliometrics and technology roadmapping. *Technological Forecasting and Social Change*, *81*, 39–48.
- Lee, S., Mogi, G., Lee, S., & Kim, J. (2011). Prioritizing the weights of hydrogen energy technologies in the sector of the hydrogen economy by using a fuzzy AHP approach. *International Journal of Hydrogen Energy*, *36*, 1897–1902.
- Lee, S., & Park, Y. (2005). Customization of technology roadmaps according to roadmapping purposes: Overall process and detailed modules. *Technological Forecasting and Social Change*, *72*, 567–583.
- Lee, C., Song, B., & Park, Y. (2015). An instrument for scenario-based technology roadmapping: How to assess the impacts of future changes on organisational plans. *Technological Forecasting and Social Change*, *90*, 285–301.
- Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 233–246). ACM.
- Lu, J., Ma, J., Zhang, G., Zhu, Y., Zeng, X., & Koehl, L. (2011). Theme-based comprehensive evaluation in new product development using fuzzy hierarchical criteria group decision-making method. *IEEE Transactions on Industrial Electronics*, *58*, 2236–2246.
- Mankins, J. C. (1995). Technology readiness levels. *White Paper*, April 6.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review*, *90*, 61–67.
- Phaal, R., Farrukh, C. J., & Probert, D. R. (2004). Technology roadmapping—A planning framework for evolution and revolution. *Technological Forecasting and Social Change*, *71*, 5–26.
- Robinson, D. K., & Propp, T. (2008). Multi-path mapping for alignment strategies in emerging science and technologies. *Technological Forecasting and Social Change*, *75*, 517–538.
- VantagePoint. (2015). GA, USA: Search Tech Inc. Available: [www.theVantagePoint.com](http://www.theVantagePoint.com). Accessed April 10, 2015.



- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.
- Zhang, Y., Guo, Y., Wang, X., Zhu, D., & Porter, A. L. (2013). A hybrid visualisation model for technology roadmapping: Bibliometrics, qualitative methodology and empirical study. *Technology Analysis & Strategic Management*, 25, 707–724.
- Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014a). “Term clumping” for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change*, 85, 26–39.
- Zhang, Y., Robinson, D., Porter, A. L., Zhu, D., Zhang, G., & Lu, J. (2015). Technology roadmapping for competitive technical intelligence. *Technological Forecasting and Social Change* (to appear).
- Zhang, Y., Zhang, G., Porter, A. L., Zhu, D., & Lu, J. (2014b). Science, technology & innovation textual data-oriented topic analysis and forecasting: Methodology and a case study. In *Proceedings of 5th International Conference on Future-Oriented Technology Analysis*. Brussels, Belgium.
- Zhang, Y., Zhou, X., Porter, A. L., Gomila, J. M. V., & Yan, A. (2014c). Triple Helix innovation in China’s dye-sensitized solar cell industry: Hybrid methods with semantic TRIZ and technology roadmapping. *Scientometrics*, 99, 55–75.
- Zhou, X., Zhang, Y., Porter, A. L., Guo, Y., & Zhu, D. (2014). A patent analysis method to trace technology evolutionary pathways. *Scientometrics*, 100, 705–721.