Measuring Technology Diffusion for the Case of RFID Technology

A Comparison between tf-lag-idf and Topic Modeling

Huseyin Caferoglu, Martin G. Moehrle
University of Bremen, Institute of Project Management and Innovation, IPMI, Bremen, Germany

***Abstract*—When an emergent technology is brought to market, different possibilities regarding its diffusion arise. While some technologies spread quickly across an entire market, other technologies are first established in one market segment, and then move on to further segments. Knowledge about the diffusion of a technology can help managers with their assessment and is therefore of major importance. Recently, [1] suggested a method to measure diffusion by means of an informetric approach, namely the tf-lag-idf. Nevertheless, shortcomings such as a high degree of manual effort and subjective coding decrease the reliability of this method. As an alternative, we develop a method based on topic modeling in accordance with [2] and test our method by using the same dataset as [1]. Applying our method to the case of RFID technology produces application fields such as logistics, payment & finance, or medicine. Comparing the results of topic modeling and tf-lag-idf based on input and output criteria sheds some light on both methods. As a consequence, both approaches enable a semi-automated analysis of diffusion based on text-mining.**

I. Introduction

When an emergent technology is brought to market, different possibilities regarding its diffusion arise. Some technologies spread quickly across the entire market, as the Apple iPhone did, for instance. Other technologies are first established in one market segment, which represents a particular application field, and then diffuse sequentially to other application fields. For example, carbon fiber reinforcement plastics were initially used in the sports and aircraft industries, and then spread to the automotive, bicycle, and wind energy industries [3]. The monitoring of such technological diffusion is an important task for technology managers in companies as well as for scholars working in the respective fields. Knowledge about the topics of a technology`s diffusion may help managers with their assessment and thus lead to a competitive advantage.

A technology`s diffusion can be measured in different ways, e.g. by analyzing the sales volumes of products related to the technology. An early identification can be based on patent data, as the protection of an invention often precedes its realization in a product, and a patent that has to be disclosed to the public contains a lot of useful information [4]. Many approaches have been introduced for measuring technology diffusion by means of patent data. The vast majority of these approaches involves the use of bibliographic meta data, such as classification or citations [5–9]. Recently, [1] presented a promising semantic approach based on a new informetric measure known as tf-lag-idf, revealing the diffusion to application fields in the case of the RFID technology (radio-frequency-identification-device).

While this approach uses more than the meta data, namely the full-text of patents, it is limited by one of its methodical elements: a manual coding of novel terms is required to identify application fields. This has two shortcomings at least: First, manual effort is required, and second, the manual coding (even if supported by a coding scheme) restricts the reliability. The question is, whether the task of diffusion analysis could be automated to a higher degree?

In this work, we develop an alternative to the tf-lag-idf method, based on topic modeling. We test this approach on the same data that was used by [1]. The aim of this paper is to identify the individual advantages and disadvantages of the topic modeling and compare it with the prior method based on the tf-lag-idf.

Our paper is organized as follows: First, we provide the theoretical background regarding the theory of diffusion. Second, we introduce RFID technology as a promising case for evaluating topic modeling and tf-lag-idf, as we use the same patent set as [1]. Third, we explain our method, focusing on topic modeling in accordance with [2] to identify application fields. Fourth, we present the results of the topic modeling, identifying several application fields like payment and finance, logistics, and medicine. Fifth, we discuss the strengths and weaknesses of the methods based on input and output criteria. In our conclusions, we summarize our work and derive recommendations for further research.

II. Diffusion of Technologies: A Theoretical Perspective

Technologies comprise specific knowledge which is represented in products, people (technology experts), processes (machinery) [10]. After its first emergence, a technology usually diffuses to different application fields. Some technologies spread quickly across an entire market, others are first established in one single market segment, and then move to further market segments. The causes for such a sequential establishment are manifold: technical, legal, resource-based as well as marketing-based reasons may be named, which sometimes work separately, sometimes in combination with each other. For instance, a technology has to proof its stability in one application field, before it is accepted in another application field (technical reason). Changes in ruling allow

978-1-890843-37-3 ©2018 PICMET

using a technology in an application field for the first time (legal reason). A limited number of technology experts urges to focus on only one application field (resource-based reason).

Without neglecting the other reasons for a technology's sequential diffusion, we focus on marketing-based reasons, in particular as we see their use for our case later on. For ease of argument, we assume that a technology is applied in one single innovative product, knowing that this is the simplest case. For innovative products, Rogers [11] has presented very early an idealized model of diffusion, he distinguishes six adaptor groups (namely: innovators, early adaptors, early majority, late majority, laggards) according to their willingness to adopt innovations, and suggests that every diffusion process has a penetration phase, in which adaptor groups successively adopt the innovation.

In reality, the idealized Rogers model is often difficult to depict. In marketing, for example, markets are segmented in target groups that are defined according to group-specific needs and characteristics [12]. Once this market segmentation has been carried out, companies develop a strategy for entering the market (figure 1), considering how many market segments they wish to address and which market segments they intend to enter in what sequence? In case of (i) single-segment concentration the company focuses on one specific market segment which it provides with one particular product. Thus, in accordance with Rogers' model, the diffusion of the technology takes place within the respective market segment (intra-application field diffusion). For (ii) selective specialization, different products of the same technology serve a multitude of market segments, either simultaneously or successively. Consequently, this enables the technology to diffuse within one market segment (intra-application field) as well as across various market segments (spill-over-effect) (inter-application field diffusion). Finally, in (iii) full market coverage, a company penetrates the entire market (all market segments), supplying it with specific products of the same technology. Similar to (ii), there are options for companies to proceed simultaneously or successively, so that intra- and inter-application field diffusion can take place. As usually all companies that wish to introduce one new technology at a time largely adhere to this segmenting and targeting behavior, the sequential introduction of a technology in different application fields is plausible on an industry level as well.

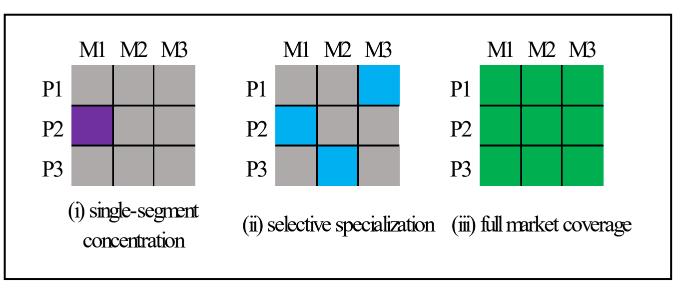

Fig. 1. Segmentation of markets. Source: [12, p.269]

III. Data

For our analysis, we select a case technology and retrieve a dataset. Considering the aim of our analysis, we have to ensure the comparability of topic modeling and tf-lag-idf. Hence, we decide to select the same case technology (RFID) and dataset as [1] used in their paper. In the following, we briefly present the motivation for selecting RFID technology and describe how the data was retrieved. For a detailed explanation of the case selection and data collection we refer to [1].

RFID is short for radio-frequency identification device and "*designed as wireless identification systems to store and broadcast information while tracking things or people*" [13, p.53]. In the context of our analysis, RFID represents a promising technology, as it is well researched and documented - from a primarily technical perspective as well as from a patent-based one. [1] explain their case selection on the basis of two aspects: First, they argue that the RFID technology is a developing technology which generates numerous patents. "*Although the origins of RFID date back to the first half of the 20th century [14], and the first commercial application of RFID already took place the 1980s, the world-wide expansion of RFID technology did not set in before the late 1990s [15]. Between the early 1990s and the end of the 21st century's first decade, RFID evolved in three phases, which could be labelled as technology development and improvement in the early 1990s, economies of scale and standardisation in the late 1990s and the early 2000s, and adoption from the mid-2000s onward [16].*" [1, p.3]. Second, since RFID is an enabling technology, [1] assume to find several application fields by means of semantic patent analysis. In retrospective, the results of their paper prove that the RFID dataset is suitable for the discovery of various application fields.

The search string initiated by the authors produces a pool of approximately 37.000 granted patents in the United States Patent and Trademark Office (USPTO) Patent and Full-Text Image Database (PatFT). For our analysis, we exclusively select patents from USPC 705 ("Data processing: financial, business practice, management, or cost/price determination) which date back to the period from 1998 to 2014. We expect to find potential application specifications in these patents especially, as they describe the use of RFIDs in different contexts, and do not focus on primarily technical aspects. Ultimately, our established patent pool consists of 3,740 business method patents. At this point, we shall not go into further detail on the data basis. For additional quality measures (e.g. precision or recall) regarding the dataset we refer to [1].

IV. Method

In our approach, we apply topic modeling to the same dataset as provided by [1], in order to enable a comparison of the results. For a more facile understanding of our work, we first present the two semantic approaches in principle, and subsequently introduce our three-step method for topic modeling.

A. Methods in principle

In our analysis, we compare topic modeling and the tf-lag-idf approach. For this purpose, it is important to give an overview of the two methods in principle. Both approaches can be organized in three generic steps (figure 2), the first of which they have in common. For both methods a term-document matrix is needed, containing terms from relevant and cleaned patent parts along with measures for these terms. While topic modeling makes use of uni-grams, tf-lag-idf relies on bi-grams. The second step consists in applying the specific algorithms.

- The first algorithm is called topic modeling, which is an unsupervised learning method applied to big text corpora in order to discover underlying topics [2]. In contrast to supervised models, this does not involve predefined labels or classes. Instead, the unsupervised model aims to identify structures and patterns in data. One popular form of topic model is known as latent dirichlet allocation (LDA) and presented by [2]. The LDA is a generative bayesian model which assumes that a document is a mixture of latent topics, and that a topic is a distribution over a fixed vocabulary (corpora) [2]. To perform an LDA, the analyst has to select a number of topics in advance. As a result, the algorithm produces a word to topic probability matrix as well as a topic to document probability matrix. In the word to topic probability matrix terms are assigned to topics and they are listed in descending order based on the probability of belonging to the topic. To characterize a topic and give it a headline, it is useful to examine some of the high rated words and search for a common category. In our analysis, we make use of the LDA in its basic form.

- The second algorithm is the tf-lag-idf which is a novel informetric measure developed through modification of the tf-idf and first presented by [1]. Whereas in case of the tf-idf, the term frequency and inverse document frequency relate to the same period of examination, the tf-lag-idf is extended by modifying the inverse document frequency in a way that it relates to the preceding period (j-1). Hence, the main emphasis of the tf-lag-idf is on detecting novel as well as interesting terms which occur frequently in the observed year, but rarely or not at all in the preceding year [1]. In order to calculate the tf-lag-idf, the analyst has to divide a dataset into time slices and calculate the tf-lag-idf for each term per year as follows [1]:

$$\boxed{tf\text{-}lag\text{-}idf_{ij} = tf_{ij} * idf_{ij-1} = tf_{ij} * \log\left(\frac{S_{j-1}}{df_{ij-1}}\right)} \qquad (1)$$

With tf_{ij} = term frequency of concept i in period j; idf_{ij-1} = inverse document frequency of concept i in period j-1; df_{ij-1} = document frequency of concept i in period j-1; S_{j-1} = Number of patents in period j-1.

The third step requires manual operations. Topic modeling produces term clusters (topics), and the analyst has to decide upon the meaning of these topics. In contrast, tf-lag-idf offers unconnected bigrams and the analyst (preferably a domain expert) has to assign them to (predefined or emergent) application fields. In their paper [1] extract for each year the top 20 bi-grams. They focus on bi-grams, which can be interpreted clearly, and cluster the remaining bi-grams to identify the main application areas. In cases of doubt, there is the possibility to retrieve the patents in which the bi-grams are used, and get some more information about the application fields from them. Doing so, one may develop a coding scheme for the assignment of bi-grams to application fields.

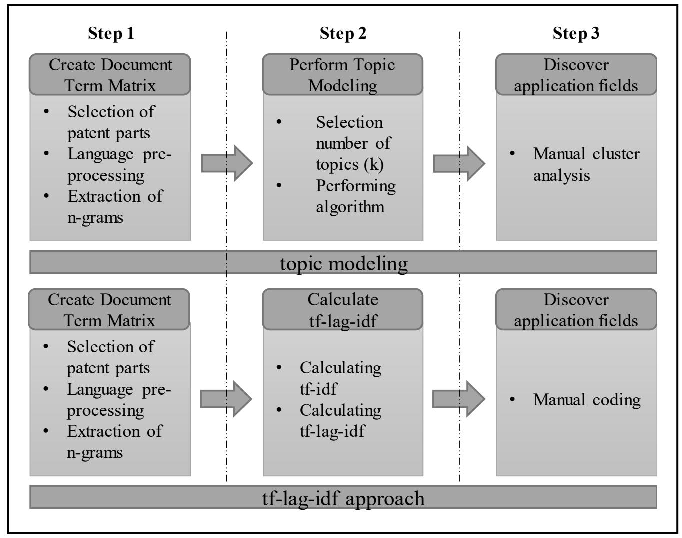

Fig. 2. Process for topic modeling and tf-lag-idf approach in three steps. Source: authors.

B. *Applying topic modeling for technology diffusion*

After providing a general overview, we go into greater detail for topic modelling, once again following the three steps described above (figure 2): First, we create a term-document matrix by pre-processing and extracting uni-grams. Second, we apply topic modeling to reveal latent topics in our patent set. Third, we identify application fields by means of manual qualitative cluster analysis.

1) *Create a Term-Document Matrix*

The first step aims to create a term-document matrix consisting of uni-grams. We divide this procedure into three sub-steps, namely (i) selection of patent parts, (ii) language pre-processing, and (iii) extraction of uni-grams.

In the first sub-step, we focus on the selection of textual parts from patents for the extraction of uni-grams ("bag-of-word"). To guarantee the comparability of our analysis with the work of [1], we decide to select titles, abstracts and claims. In fact, the selection of these particular elements is recommended by many authors, as they describe the essential content of a patent [17–20].

In order to proceed with the language pre-processing, we clean (by filtering out stop words) and unify (by lemmatizing the terms) the textual parts. For this purpose we use the PatVisor® which, for instance, eliminates stop words and patent specific words as well as numbers, punctuation, and symbols [21]. A lemmatizer is used to obtain word stems.

TABLE I. RESULTS FROM TOPIC MODELING. THE FIRST AND SECOND COLUMN CONTAIN THE CLUSTERS AND ASSIGNED TOPIC NUMBERS. THE FOLLOWING COLUMNS ENTAIL FOR EACH TOPIC FIVE UNI-GRAMS SORTED IN DESCENDING ORDER (FROM LEFT TO RIGHT) BASED ON THEIR PROBABILITY OF BELONGING TO A TOPIC. SOURCE: AUTHORS

cluster	topic	uni-gram				
security & authentication	topic 4	authentication	token	secure	security	encrypt
	topic 14	identifier	request	receive	authorization	associate
	topic 27	code	identification	database	system	unique
	topic 40	electronic	access	network	entity	system
technological infrastructure	topic 1	location	determine	associate	system	area
	topic 24	mobile	communication	wireless	server	portable
	topic 33	signal	reader	radio	frequency	identification
	topic 35	server	computer	element	software	system
finance & payment	topic 2	machine	cash	include	check	operative
	topic 6	card	credit	plurality	prepaid	computer
	topic 9	account	financial	amount	fund	consumer
	topic 21	payment	point	sale	bill	platform
	topic 25	transaction	system	pointsale	include	processor
medicine	topic 8	medication	prescription	drug	dose	time
	topic 20	patient	medical	healthcare	health	system
	topic 34	record	system	patient	computer	medical
logistic	topic 15	container	asset	system	label	waste
	topic 16	order	delivery	return	recipient	mail
	topic 17	item	inventory	system	plurality	shop
	topic 28	package	track	system	material	ship
trade	topic 13	customer	system	associate	program	include
	topic 26	merchant	system	computerbase	base	processor
	topic 31	service	consumer	message	provider	offer
	topic 37	reward	coupon	computer	discount	purchase
	topic 39	product	purchase	price	store	computer
transportation	topic 10	vehicle	toll	domain	passenger	performance
	topic 38	sensor	monitor	park	system	space

According to the recommendations given by [22], we filter out uni-grams which occur singly. This reduces the computational burden during the topic modeling without affecting topic distinctiveness and quality.

Similar to the work of [2], we use uni-grams (bag-of-words) to carry out the Latent Dirichlet Allocation, accepting that we ignore the information which the word order delivers. Using the PatVisor®, we retrieve uni-grams as a term document matrix.

2) *Perform topic modeling*

In the second sub-step, we carry out topic modeling in accordance with [2], using the topicmodels package in software R[1]. The aim of the second step is to create a word to topic probability matrix as well as a document to topic probability matrix.

In order to perform this topic modeling, it is necessary to specify the number of topics (k) beforehand. There are many measures which support analysts in determining the number of topics, e.g. perplexity [2, 23]. Nevertheless, some authors argue "*that traditional metrics do not capture whether topics are coherent or not*" [24, p.8] and recommend to "*focus on evaluations that depend on real-world task performance rather than optimizing likelihood-based measures*" [24, p.8]. An example for such a real world task performance is used by [25]. The authors use a trial an error approach to ensure a topic coherence. Based on that, we decide also to use a trial-and-error approach to specify the number of topics. For this purpose, we perform topic modeling for several numbers of topics, and verify the suitability for our purposes by analyzing the topic coherence, knowing that words with a high probability of belonging to a topic give a good impression of the topic's content. Consequently, we decide to use 40 topics for our analysis.

3) *Discover application fields*

Finally, in the third step we detect potential application fields on the basis of the retrieved topics. Many application fields are not represented by a single topic but by several ones. Hence, we carry out a manual cluster analysis in which we qualitatively analyze each topic (words with a high probabilities) and cluster these in an appropriate way. By doing so, we identify seven clusters.

V. RESULTS

Our analysis provides insights into the spectrum of application fields of RFID technology. By means of topic modeling we identify the following seven clusters, each of which contains a variety of topics: (i) technological infrastructure, (ii) security & authentication, (iii) finance & payment, (iv) medicine, (v) logistic, (vi) trade, (vii) transportation and (viii) others. In table 1, we present these clusters and their related topics by showing the top five uni-grams with a high probability of belonging to a topic. For example, the cluster (i) technological infrastructure contains uni-grams like mobile, communication, signal, or reader. We do not think that this cluster describes an application field but technical aspects of the RFID technology. In contrast to this,

[1] The package topicmodels uses the C code for LDA by David M. Blei and co-authors to estimate and fit a latent dirichlet allocation model with the VEM (variational empirical bayes) algorithm

the clusters (ii) security & authentication, (iii) finance & payment, (iv) medicine, (v) logistic, (vi) trade, (vii) transportation do represent application fields. Finally, the cluster (viii) others contains several topics that we are unable to assign to any specific application field. However, we identify some indicators of further application fields, such as gaming or insurance.

To perform a time series analysis, we combine the results of topic modeling with the application date of the patents (figure 3). Only patents are included which could be assigned to one of six application fields. The time series analysis does not reveal when the diffusion started. Instead it enables us to make few, albeit rahter vague, statements about trends:

- Trend 1: In the first examined years, the cluster of technological infrastructure seems to account for a large portion of the patents. However, there is an indication that its importance decreases over the following years.
- Trend 2: At the beginning, logistics had a portion of 20 % in 1998, increasing to 30 % in 2001, decreasing slowly and constantly over the course of time to 10 % in 2014.
- Trend 3: In the case of payment & finance, there are hints to a continuous increase in importance over the years.
- Trend 4: It appears that the cluster security & authentication is of constant importance in the context of RFID technology.

In summary, our method reveals coherent topics and helps us gain a quick overview regarding the application fields of RFID. Furthermore, a time series analysis enables us to make some trend statements. Nevertheless, not every topic could be assigned to a particular application field, and our method does not reveal when the diffusion started. Ultimately, the given trend statements should be handled with caution, as they represent weak signals at most.

VI. Comparison

After the topic modeling and presentation of the results, we are now going to compare our method with [1], on the basis of certain input and output criteria. As concerns input criteria, we compare how much effort the analyst has to put into the execution of the different methods, focusing on time aspects and the degree of automation. Regarding output criteria, we compare the results produced by both methods in terms of interpretation possibilities as well as individual advantages and disadvantages of the methods.

In order to figure out differences regarding input criteria, we take a closer look at the steps comprised in the two methods (figure 2). The first step is quite similar in both methods: The analyst selects parts of the text, carries out language pre-processing, and extracts n-gram models. In step 2, certain differences occur: In case of the tf-lag-idf approach, the analyst has to calculate a tf-idf for the first observation year and then the tf-lag-idf for each bi-gram of all subsequent years. In the case of topic modeling, however, the LDA algorithm has to be applied on R without considerable intervention. The analyst has to decide how many topics (k) to use by checking the results of different k values for topic coherence. In step 3, both methods are aimed at detecting application fields. In case of the tf-lag-idf approach, manual coding is carried out by assigning the identified terms to application fields according to a subjective assessment. In the context of topic modeling, a manual cluster analysis is performed, in which the detected topics are clustered by the analyst according to application fields, in order to avoid overlaps. Based on the aforementioned observations, we conclude that topic modeling is (i) less time consuming and (ii) more automated than the tf-lag-idf approach.

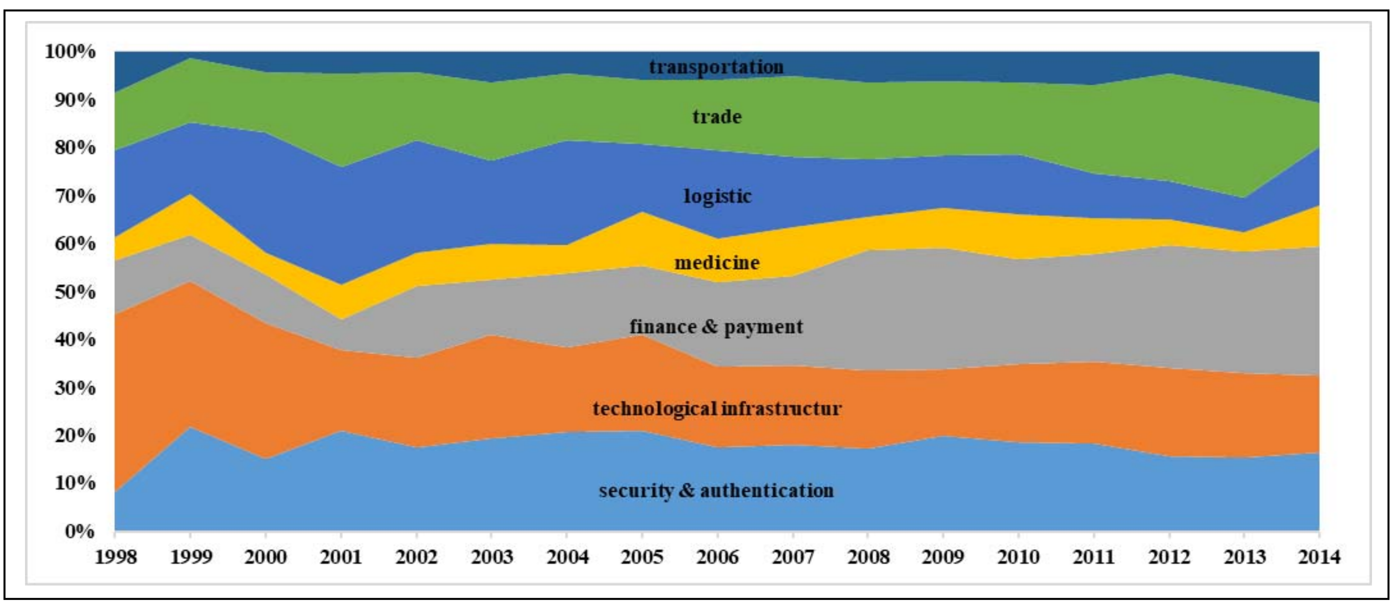

Fig. 3. A time series analysis oft he clusters between 1998-2014. Only patents are included which could be assigned to one of six application fields Source: authors

TABLE II. COMPARING TOPIC MODELING AND TF-LAG-IDF BASED ON INPUT AND OUTPUT CRITERIA. SOURCE: AUTHORS.

<table>
<tr><th colspan="2"></th><th>topic modeling</th><th>tf-lag-idf</th></tr>
<tr><td rowspan="2">input</td><td>human intervention</td><td>low</td><td>medium</td></tr>
<tr><td>degree of automation</td><td>high</td><td>medium</td></tr>
<tr><td rowspan="5">output</td><td rowspan="2">advantage</td><td>overview of technology field</td><td>novel application fields</td></tr>
<tr><td>more intuitive identification of application field</td><td>manageable number of bi-grams to be inspected</td></tr>
<tr><td rowspan="3">disadvantage</td><td>No convincing rule for the number of topics which should be extracted</td><td>Some identified bi-grams may represent other aspects than application fields</td></tr>
<tr><td>Several application fields may be mixed in one topic</td><td rowspan="2">no information about the development of application field over time</td></tr>
<tr><td>One application field may be spread over several topics</td></tr>
</table>

In terms of output criteria, we compare the results produced by both methods to shed light on major differences. This comparison of the results shows that by means of tf-lag-idf [1] we manage to detect ten application fields, whereas we identify six application fields and some further indicators of possible application fields such as gaming or insurance by topic modeling. This difference in the results can be traced back to the methods' two chief characteristics. First, the tf-lag-idf approach reveals novel application fields for every observed year, whereas topic modeling is very explorative in character, and yields information regarding all topics dealt with in the respective technology field over time. Second, topic modeling is not specifically designed for detecting application fields, which leads to the discovery of coherent topics, indicating several application fields. Therefore, several topics cannot be assigned to any specific application field. With tf-lag-idf, on the other hand, the subjective assessment performed by the analyst in the course of manual coding ensures a formation of clusters related to clearly defined individual application fields, which leads to the detection of a larger number of application fields.

Summing up, we come to the conclusion that both methods are marked by different characteristics and thus have varying advantages and disadvantages as regard the detection of application fields. The tf-lag-idf is well suited for detecting novel application fields, enabling the analyst to every application field by a manageable number of bi-grams. In contrast, topic modeling is suitable for gaining an overview of the application spectrum of a technology field. It yields information regarding all topics dealt with in the technology field over time. In consequence, our recommendation to analysts is to select a method according to their purposes, or to combine both methods.

VII. CONCLUSIONS

This paper discusses topic modeling as an alternative to the tf-lag-idf approach for measuring a technology`s diffusion. We develop a method in accordance with the latent dirichlet allocation model by [2] to perform topic modeling. By applying our method to the same case (technology and dataset) as [1], we identify several application fields for RFID. Finally, we discuss the results of a tf-lag-idf approach and topic modeling regarding input and output criteria to establish specific profiles of both methods.

The aforementioned comparison leads to three insights: (i) In consideration of input criteria, we identify topic modeling as being less time consuming and more automated than the tf-lag-idf approach. (ii) In terms of output criteria, the strength of topic modeling lies in its capability to provide an overview of the entire observation field by detecting topics, including weak information on the development of these topics over time. However, it is not possible to assign topics and application fields one by one, manual inspection is still needed. In contrast to this, the strength of the tf-lag-idf approach lies in its ability to detect novel application fields – albeit with the shortcoming that little information is provided regarding the development of these application fields over time.(iii) We conclude that both methods are marked by different characteristics and thus have their particular advantages and disadvantages in terms of detecting application fields. For this reason, our recommendation to analysts is to select a method according to their purposes, or to combine both methods.

Seen from a managerial point of view, there are two significant implications: First, we believe that our method is well suited for application in management, as it provides a facile and fast way to gain an overview of any technology field. Second, we conclude that a hybrid model (combined tf-lag-idf and topic modeling) enables instructive insights into the detection of technology dynamics. This might also help scholars in the analysis of a technology`s diffusion.

Our study is characterized by two limitations: (i) the first limitation relates to both methods simultaneously. Since both methods are based on text mining, biases caused by synonyms and homonyms (double meanings of words) may occur. (ii) Since topic modeling is designed for identifying topics rather than application fields, it is possible to identify coherent topics involving terms that point to various application fields. Therefore, these topics cannot be assigned unambiguously to any specific application field.

Our work suggests options for further research. (i) In terms of the tf-lag-idf, there is need to expand the method in order to gain insights into the development of the detected application fields. (ii) As regards the topic modeling approach, a time- slice based approach thereof should be applied to enable the direct detection of novel application fields. Such approaches have already been suggested by [26].

REFERENCES

[1] M.G. Moehrle, M. Wustmans, J.M. Gerken, "How business methods accompany technological innovations - a case study using semantic patent analysis and a novel informetric measure", R&D Management 7, 2017, p. 379.

[2] D.M. Blei, A.Y. Ng, M.I. Jordan, "Latent dirichlet allocation", Journal of machine Learning research 3, 2003, pp. 993–1022.

[3] J. Frischkorn, "Technologieorientierte Wettbewerbspositionen und Patentportfolios". Dissertation, 2017.

[4] D. Alberts, C.B. Yang, D. Fobare-DePonio, K. Koubek, S. Robins, M. Rodgers, E. Simmons, D. DeMarco, "Introduction to Patent Searching", in: M. Lupu, K. Mayer, J. Tait, A.J. Trippe (Eds.), Current Challenges in Patent Information Retrieval, Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 3–43.

[5] H. Ernst, "The Use of Patent Data for Technological Forecasting: The Diffusion of CNC-Technology in the Machine Tool Industry", Small Business Economics 9, 1997, pp. 361–381.

[6] S.-B. Chang, S.-M. Chang, W.-Y. Guh, "Exploring the Technology Diffusion Trajectories and Groups of Basic Patents of Business Methods": Using the Patent Citation Network, in: D.F. Kocaoglu (Ed.), Management of converging technologies: PICMET '07, 5 - 9 August 2007, Portland, Oregon, USA ; proceedings, IEEE, Piscataway, NJ, 2007, pp. 1784–1789.

[7] M.H. Fallah, E. Fishman, R.R. Reilly, "Forward patent citations as predictive measures for diffusion of emerging technologies", in: PICMET '09 - 2009 Portland International Conference on Management of Engineering & Technology, IEEE, 2009, pp. 420–427.

[8] S.-B. Chang, K.-K. Lai, S.-M. Chang, "Exploring technology diffusion and classification of business methods", Technological Forecasting and Social Change 76, 2009, pp. 107–117.

[9] N. Duch-Brown, M.T. Costa-Campi, "The diffusion of patented oil and gas technology with environmental uses", Energy Policy 83, 2015, pp. 267–276.

[10] G. DOSI, : Technological paradigms and technological trajectories. In: Research Policy 11, 1982, Nr. 3, S. 147–162

[11] E.M. Rogers, "Diffusion of innovations", Free Press, New York, London, Toronto, Sydney, 2003.

[12] P. Kotler, K.L. Keller, "Marketing management", 13th ed., Pearson Prentice Hall, Upper Saddle River, NJ, 2009.

[13] C.V. Trappey, H.-Y. Wu, F. Taghaboni-Dutta, A.J.C. Trappey, "Using patent data for technology forecasting", Advanced Engineering Informatics 25, 2011, pp. 53–64.

[14] J. Landt, "The history of RFID", IEEE Potentials 24, 2005, pp. 8–11.

[15] M. Bhuptani, S. Moradpour, "RFID field guide": Deploying radio frequency identification systems, Prentice Hall PTR, Upper Saddle River, NJ, 2005.

[16] R. Bunduchi, C. Weisshaar, A.U. Smart, "Mapping the benefits and costs associated with process innovation", Technovation 31, 2011, pp. 505–521.

[17] Y.-H. Tseng, C.-J. Lin, Y.-I. Lin, "Text mining techniques for patent analysis", Information Processing & Management 43, 2007, pp. 1216–1247.

[18] J. Yoon, H. Park, K. Kim, "Identifying technological competition trends for R&D planning using dynamic patent maps", Scientometrics 94, 2013, pp. 313–331.

[19] T. Daim, N. Preschitschek, H. Niemann, J. Leker, M. G. Moehrle, "Anticipating industry convergence", Foresight 15, 2013, pp. 446–464.

[20] M.G. Moehrle, F. Passing, "Applying an anchor based patent mapping approach", World Patent Information 45, 2016, pp. 1–9.

[21] L. Walter, F.C. Schnittker, "Patentmanagement": Recherche - Analyse - Strategie, 2016.

[22] K. Lu, X. Cai, I. Ajiferuke, D. Wolfram, "Vocabulary size and its effect on topic representation", Information Processing & Management 53, 2017, pp. 653–665.

[23] W. Zhao, J.J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, W. Zou, "A heuristic approach to determine an appropriate number of topics in topic modeling", BMC bioinformatics 16 Suppl 13, 2015, S8.

[24] J. CHANG; S. GERRISH; C. WANG; J.L. BOYD-GRABER ; D.M. BLEI: Reading Tea Leaves : How Humans Interpret Topic Models. In: BENGIO, Y.; SCHUURMANS, D.; LAFFERTY, J. D.; WILLIAMS, C. K. I.; CULOTTA, A. : Advances in Neural Information Processing Systems 22 : Curran Associates, Inc, 2009, S. 288–296

[25] A. SUOMINEN ; H. TOIVANEN; M. SEPPÄNEN,: Firms' knowledge profiles : Mapping patent data with unsupervised learning. In: Technological Forecasting and Social Change 115 (2017), S. 131–142

[26] D.M. BLEI, ; J.D. LAFFERTY: Dynamic topic models. In: COHEN, William : Proceedings of the 23rd international conference on Machine learning. New York, NY : ACM, 2006, S. 113–120