UNIVERSITEIT
GENT

**FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE**

**TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92**

WORKING PAPER

A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies

Dirk Thorleuchter[1]

Dirk Van den Poel[2]

Anita Prinzie[3]

January 2010

2010/632

[1] PhD Candidate, Ghent University & Fraunhofer INT (Germany)
[2] Prof. Dr. Dirk Van den Poel, Professor of Marketing Modeling/analytical Customer Relationship Management, Faculty of Economics and Business Administration, dirk.vandenpoel@ugent.be; more papers about text mining can be obtained from the website: www.crm.UGent.be
[3] Postdoctoral researcher at Ghent University

D/2010/7012/03

A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies

D. Thorleuchter[1], D. Van den Poel[2], and A. Prinzie[2]

[1]Fraunhofer INT, D-53879 Euskirchen, Appelsgarten 2, Germany & PhD Candidate, Ghent University (corresponding author), dirk.thorleuchter@int.fraunhofer.de

[2]Ghent University, Faculty of Economics and Business Administration, B-9000 Gent, Tweekerkenstraat 2, Belgium, dirk.vandenpoel@ugent.be, anita.prinzie@ugent.be

ABSTRACT

The planning of technological research and development (R&D) is demanding in areas with many relationships between technologies. To support decision makers of a government organization with R&D planning in these areas, a methodology to make the technology impact more transparent is introduced. The method shows current technology impact and impact trends from the R&D of an organization's competitors and compares these to the technology impact and impact trends from the organization's own R&D. This way, relative strength, relative weakness, plus parity of the organization's R&D activities in technology pairs can be identified.

A quantitative cross impact analysis (CIA) approach is used to estimate the impact across technologies. Our quantitative CIA approach contrasts to standard qualitative CIA approaches that estimate technology impact by means of literature surveys and expert interviews. In this paper, the impact is computed based on the R&D information regarding the respective organization on one hand, and based on patent data representative regarding R&D information of the organization's competitors on the other hand. As an illustration, the application field 'defence' is used, where many interrelations and interdependencies between defence-based technologies occur. Firstly, an R&D-based and patent-based Compared Cross Impact (CCI) among technologies is computed. Secondly, characteristics of the CCI are identified. Thirdly, the CCI data is presented as a network to show the overall structure and the complex relationships between the technologies. Finally, changes of the CCI are analyzed over time. The results show that the proposed methodology generates useful insights for government organizations to direct technology investments.

.

Key Words: Compared cross impact, Cross impact analysis, Technological impact analysis, R&D, Patent analysis, Defence Taxonomy, Centroid Vector, Machine Learning, Multi Label Classification

Introduction

The planning of research and development (R&D) requires technological trend analysis to ensure an effective investment of limited R&D budgets within organizations [1]. However, trend analysis is a very demanding task in areas where many interrelations and interdependencies between technologies occur because the impact of all related technologies has to be considered. Therefore, analyzing the impact across technologies is helpful for R&D planning and also to develop R&D strategies in these areas.

To support an organization's strategy and R&D planning, the technology impact analysis should be done both for the organization's own R&D activities (from now on referred to as 'internal R&D') as well as for the competitors' R&D activities (from now on referred to as 'external R&D'). By comparing the technology impact from internal R&D to the impact from external R&D, one can portray the advantages and the disadvantages of the internal R&D to competitors' external R&D. This improves the planning of R&D activities [2,3], the systematic identification of R&D priorities [4], the discovery of current technological vacuums [5], and the analysis of technological trends and opportunities [6,7] for the organization at hand.

The internal R&D technology impact analysis focuses on the relationships between technologies of the many simultaneously run R&D projects in the organization's R&D department. Typically, one R&D project deals with several different technologies. Therefore, each internal R&D project is assigned to one or several technologies from a specific technology list or taxonomy [8] by multi-label classification [9]. Analyzing this multiple classification shows which R&D projects are frequently assigned to specific technologies. This enables to calculate the cross impact index estimating the impact across these technologies developed by the organization. A proper calculation of this cross impact index requires a large number of internal R&D projects working on many different technologies. Companies normally do not have a large number of internal R&D projects or they are limited to a small number of technologies. Therefore, our approach focuses specifically on government organizations with a large number of R&D projects (> 100 projects) and a large technological scope (> 20 technologies).

The estimation of the technology impact from internal R&D should be augmented with the analysis of the relationships between technologies of external R&D. After all, no organization is so large that it has enough resources to excel in all technological areas or that it could not benefit from the advice of others [10]. For instance, organizations could learn from small firms, which are often more innovative. Therefore, it is necessary to consider R&D activities related to the internal R&D technologies from other organizations, i.e. external R&D. Patent data are used as representative for external R&D (see Sect. 2.3) because patents normally represent results of R&D projects. If this external R&D is also assigned to several technologies from the above mentioned technology list or taxonomy using multi-label classification then the impact across these technologies can also be estimated for the R&D activities of the organization's competitors.

This paper uses cross impact analysis (CIA) to estimate the impact of each technology on other technologies in a quantitative way as opposed to the more common qualitative approach by means of literature surveys and expert interviews. Our focus on large application fields characterized by a large number of corresponding technologies makes traditional qualitative CIA inappropriate (where a cross impact matrix is constructed by technology experts estimating the initial impact probabilities of each technology and the conditional impact probabilities of each technology pair [11,12,13]). However, in large application fields, a large number of corresponding technologies exists e.g. in the 'defence' application field the European Defence Agency (EDA) taxonomy of technologies consists of more than 200 technologies. To construct a 200-by-200 cross impact matrix n * (n-1) = 200 * 199 = 39.800 estimations are required by human experts. As can be seen from this example, a qualitative CIA approach in large application fields seems infeasible.

In this paper, a quantitative CIA approach is used to compute technology impact estimates that incorporate both internal and external R&D. In contrast to other quantitative CIA approaches which estimate the absolute impact of technologies (see Sect. 2.1), we first focus on technologies from an application field (e.g. 'defence') by assigning both internal R&D from an organization as well as external R&D to these technologies by multi-label classification. Then, we evaluate the relative impact of technologies by comparing the impact from internal R&D to the impact from external R&D, as captured by a new index we developed called the Compared Cross Impact (CCI) index (see Sect. 3). This relative impact shows how a government organization with many R&D projects can profit from the R&D of others (see Sect. 4 and 5).

This paper contributes to previous research in multiple ways. The main contribution of the proposed approach is the new CCI index that identifies relative strength, relative weakness, plus parity of the organization's R&D activities in technology pairs. The second contribution is a method to determine the characteristics of relationships and to show whether two technologies are equally influencing one another (symmetry) or whether the impact of the first technology on the second is different from the impact of the second technology on the first (asymmetry). A third contribution is the presentation of a CCI network graph that shows the overall structure and the complex CCI relationships between several technologies. Finally, changes of the CCI are analyzed over time to discover trends regarding how the technology impact changes over time. They show which technology should receive more or less development and investment. Overall, the results testify to the ability of CCI to generate useful insights for R&D decision makers of organizations.

Background

This approach combines methods from CIA and text classification and it applies them on patent data. The following paragraphs give an overview on existing CIA and text classification methods and on the (dis-) advantages of patent data.

Cross impact analysis

The use of CIA was first mentioned in 1968 [16] and consists of five steps. Firstly, events (e.g. technologies) are defined. Then, the occurrence probabilities and the conditional probabilities between events are estimated in the second and third step. Fourthly, a calibration run is performed to access the consistency / stability of the probabilities and last, the results are evaluated.

In literature, many improved CIA approaches have been introduced. Most of these necessitate the involvement of human experts and are therefore more subjective. The approaches are applied to different areas. Dalkey presents conditions for computing the occurrence probabilities of the first- and second-order [17]. To compute the higher-order probabilities, Duperring and Godet suggest a quadratic programming method [18] and Mitchell provides a linear programming method [19]. Enzer uses CIA to forecast future technologies based on a Delphi survey. Blanning and Reinig use the ratio of experts to define the occurrence probability P(A) (the percentage of all experts who predict the occurrence of A) and the conditional probability P(B|A) (the number of all experts who predict the occurrence of both A and B divided by the number of all experts who predict the occurrence of A) [20].

Additionally, more objective CIA approaches have also been introduced. Caselles-Moncho uses cumulative sales probabilities over time to compute the occurrence probabilities [21]. Jeong and Kim create inference algorithms based on linguistic values and the time lag as fuzzy numbers to compute the conditional probabilities between technologies [11]. A patent based CIA is presented in [1]. The standard assignment of US patents to the United States Patent Classification [22] is used to assign patents to several patent classification codes (PCC). A PCC impact index Impact(A,B) = P(B|A) is proposed to compute the impact of PCC A on PCC B.

Text classification methods

Text classification aims at assigning pre-defined classes (e.g. technology areas) to text documents (e.g. patent descriptions). The most frequently used data mining methods for text classification (categorization) are described in [26]: Naive Bayes is a probabilistic classifier simplifying Bayes' Theorem by naively assuming class conditional independence. The k nearest neighbor (k-NN) classification as instance-based learning algorithm selects documents from the training data which are 'similar' to the target document. Subsequently the class of the target document can be inferred from the class labels of these similar documents. Decision trees [27] are non-parametric classifiers recursively partitioning the observations (patent documents) into subgroups with a more homogeneous response (technology area). C4.5 is a well-known decision tree algorithm. A Support Vector Machine (SVM) is a supervised classification algorithm that determines a hyperplane, which separates the positive examples from the negative examples of the training data. A small number of training examples (support vectors) determine the actual location of the hyperplane. Then, target documents are assigned to one side of this hyperplane. The centroid-based approach [28] describes classes by a centroid vector that summarizes the characteristics of each class, but not by a number of training examples like k-NN and SVM. The assumption of a centroid classifier is that a target document should be assigned a particular class if the similarity of the document vector to the centroid vector of the class is the largest.

Patent data

Patent data are a valuable source of information concerning R&D. The data are useful to researchers for technological decision-making as well as to technology planners for R&D strategy making. Nevertheless, there are some limitations to use patent data because not all inventions are patented [14], the interpretations of patent analyses are not consistent across technology fields [15], and changes in patent law make it difficult to analyze trends over time [14]. However, patents are often used in analyses on technological innovation.

In patent research, statistical data are normally used (e.g. number of patents, application year, registration country, citation information). On the contrary, this research focuses on patent classification data by multiple assignment of patents to technologies and by computing the impact across these technologies. Patent data are used as representative for external R&D. Comparing the external R&D impact to the impact of internal R&D activities from a large organization leads to interesting knowledge for planning and managing R&D activities in this organization.

Methods: A compared R&D-based and patent-based CIA

Overview of proposed CIA

Our proposed quantitative CIA approach to estimate the impact between technologies for organizations with many R&D projects consists of multiple steps as depicted in Fig. 1. In a pre-processing step, internal R&D and external R&D are assigned to specific technologies based on internal R&D project information and patent data respectively. In a second step, the cross impact indexes CI^{6}_{17}(A,B) and CI⅛ṭ$_{7}$(A,B) for each technology pair are calculated. Next, the cross impact indexes CI^{6}_{17}(A,B) and CI⅛ṭ$_{7}$(A,B) are rounded and recoded to boolean cross impact indexes BCI^{6}_{17}(A,B) and BCI⅛ṭ$_{7}$(A,B). In the fourth step, a CCI index CCI(A,B) for each technology pair is calculated and characterized. These CCI scores already provide insights into the organization's relative strength and relative weakness. In the fifth step, a CCI network graph is created visualising the CCI of technologies thereby facilitating the identification of relative strength and relative weakness even more. Steps one to four are discussed in Sect. 3.2 and Sect. 3.3 below. Sect. 3.4 elaborates on step 5. Finally, Sect. 3.5 documents on how the entire five-step approach can be applied on longitudinal data to infer evolution in technology impacts.

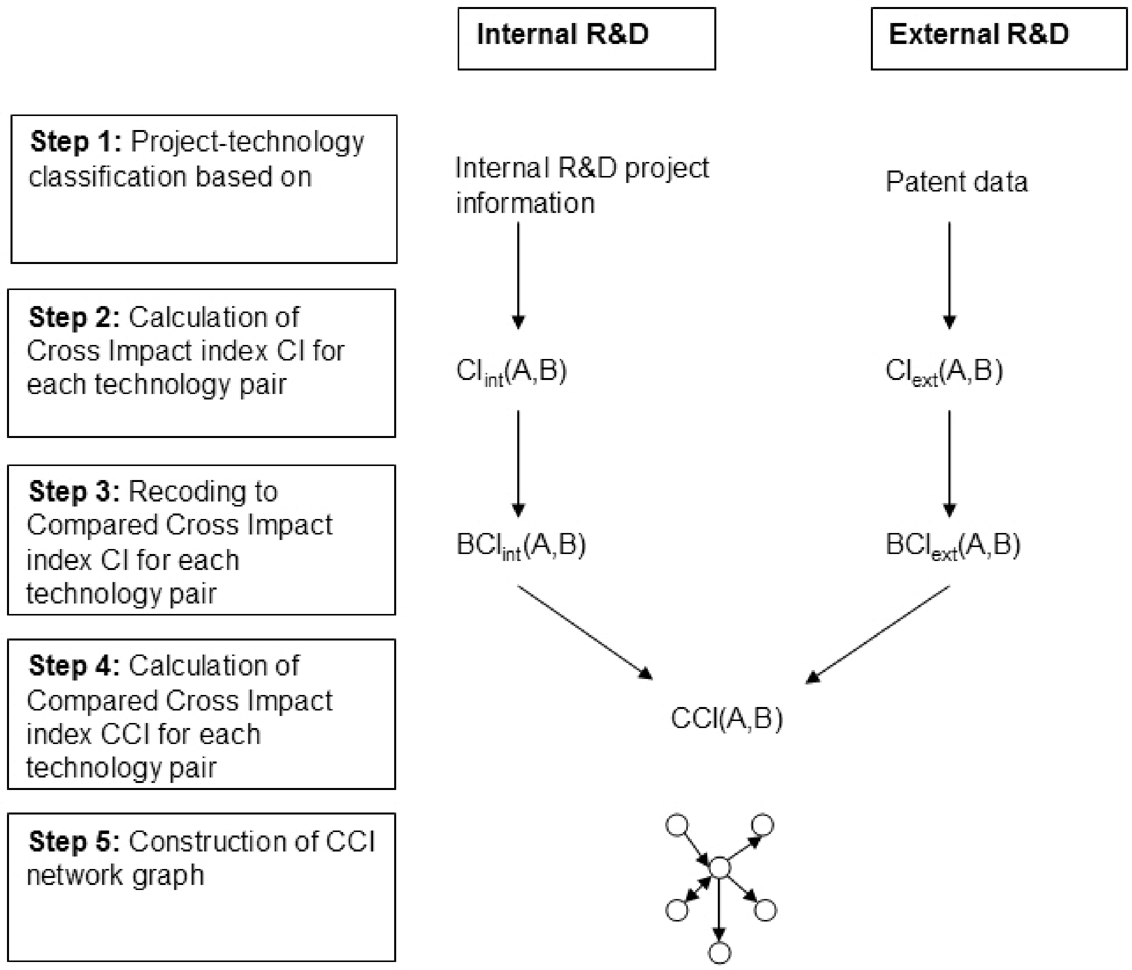

Figure 1: Overview of quantitative CIA-approach.

Estimation of the new compared cross impact index

We adapt the PCC impact index from Sect. 2.1 to a) measure the cross-technology impact of external R&D as reflected by patents and b) measure the cross-technology impact of internal R&D. These modified indices are defined below:

Definition 1. Let $N⅛t_{7}(A)$ be the number of patents (as representative for external R&D) that are associated with technology A and let $N⅛t_{7}(A \cap B)$ be the number of patents associated with both, technology A and B. Then, the cross impact index for external R&D $CI⅛t_{7}(A,B)$ is defined as the conditional probability between technology A and technology B considering patent data.
$CI⅛t_{7}(A,B) = P⅛t_{7}(B|A) = N⅛t_{7}(A \cap B) / N⅛t_{7}(A)$ (1)

In a similar way the cross impact index for external R&D $CI⅛t_{7}(B,A)$ is defined as the conditional probability between technology B and technology A considering patent data.

Let $N^{6}_{17}(A)$ be the number of R&D projects (as representative for internal R&D) that are associated with the technology A and let $N^{6}_{17}(A \cap B)$ be the number of R&D projects associated with both, technology A and B. Then, the cross impact index for internal R&D $CI^{6}_{17}(A,B)$ is defined as the conditional probability between technology A and technology B considering internal R&D projects.

$CI^{6}_{17}(A,B) = P^{6}_{17}(B|A) = N^{6}_{17}(A \cap B) / N^{6}_{17}(A)$ (2)

Likewise, the cross impact index for internal R&D $CI^{6}_{17}(B,A)$ is defined as the conditional probability between technology B and technology A considering internal R&D projects.

Result values of $CI⅛t_{7}(A,B)$, $CI⅛t_{7}(B,A)$, $CI^{6}_{17}(A,B)$, and $CI^{6}_{17}(B,A)$ are between 0 and 1. A result value of one means that the first technology has a strong impact on the second technology and a result value of zero means that there is no impact. Two examples to illustrate the meaning of the cross impact index for internal R&D and external R&D are presented. A CI^{6}_{17}(A,B) of 0.25 means that 25% of all internal R&D projects adopting technology A also employ technology B. A $CI^{1/8}_{t_7}$(A,B) of 0.20 means that 20% of all patents related to technology A also refer to technology B.

The estimation of cross impact between technologies is done in two different ways.

Firstly, relationships between technologies are estimated using data regarding R&D activities from an organization. Internal R&D projects are assigned to technologies from a specific technology list or taxonomy (that is normally used in the organization for technology classification). This multiple assignment can be used to compute CI^{6}_{17}(A,B). A proper computation of the cross impact index requires that each technology is associated with many R&D projects from the organization (see Sect. 1). The calculation of the CI^{6}_{17}(A,B) provides organization researchers and research planners with an internal view of the relationships between technologies. However, this internal view does not consider relationships between technologies as apparent from external R&D.

Next, the R&D-technologies multiple assignment and calculation of the cross impact index is repeated for external R&D using patent data instead of internal R&D information. The patent data are assigned to the technologies from the above described technology list or taxonomy. For this, methods from text classification can be used (see Sect. 2.2). This means those patents are considered that are related to at least one technology. The advantages of this patent-based CIA for researchers and technology planners are described in [1]. The disadvantage of patent-based CIA is that it neglects the technological relationships of the internal R&D when assessing the cross-technology impact.

Therefore, a compared R&D-based and patent-based CIA is proposed. Hence, we compute $CI_{int}(A,B)$ and $CI_{ext}(A,B)$. Then, boolean cross impact indexes and cutoff values are defined to decide whether there is an impact of technology A on technology B taking both internal R&D as well as external R&D into account.

Definition 2. Let c_{int} and c_{ext} be the internal and external cutoff percentages respectively. The boolean cross impact index $BCI_{int}(A,B)$ for internal R&D and the boolean cross impact index $BCI_{ext}(A,B)$ for external R&D are defined as follows:

$$BCI_{int}(A,B) = \begin{cases} 1 & (CI_{int}(A,B) \geq c_{int}) \\ 0 & (CI_{int}(A,B) < c_{int}) \end{cases} \tag{3}$$

$$BCI_{ext}(A,B) = \begin{cases} 1 & (CI_{ext}(A,B) \geq c_{ext}) \\ 0 & (CI_{ext}(A,B) < c_{ext}) \end{cases} \tag{4}$$

The cutoff percentage is separately defined for internal and external R&D. This is because the number of internal R&D projects is much smaller than the number of patents. As an example, if $N_{int}(A)$ equals five and $N_{int}(A \cap B)$ equals one then $CI_{int}(A,B)$ equals 0.20. However, this high value does not mean that this technology pair is a focal point in the R&D of the organization and that technology A has an impact on technology B. In contrast to this, a $CI_{ext}(A,B)$ of 0.20 means that 20% of all patents in technology A are also in technology B. Therefore, technology A has an impact on technology B. As seen from this example, it is necessary that cutoff values are separately defined for internal and external R&D e.g. for the case study in Sect. 4, the cutoff percentage for internal R&D c_{int} is set to 0.25 whereas the cutoff percentage for external R&D c_{ext} is set to 0.20.

Definition 3. Starting from the boolean cross impact indexes we define a CCI index CCI(A,B) as the difference between the internal and external boolean cross impact index.

$$CCI(A,B) = BCI^{6}_{17}(A,B) - BCI{}^{1}\!/_{8}t_{7}(A,B) \quad (5)$$

Depending on whether $BCI^{6}_{17}(A,B)$ and $BCI{}^{1}\!/_{8}t_{7}(A,B)$ are zero or one, the result value of CCI(A,B) is negative one, zero, or positive one (see Table 1). If CCI(A,B) equals negative one then a relative weakness in this area is observed for the organization. Technology A has an impact on technology B in the external R&D but not in the internal R&D. The internal R&D does not exploit this technology pair intensively. A potential strategic decision could be to increase R&D activities in this area. Alternatively, to gain strength in this area, the organization could outsource these R&D activities (to buy external R&D know how).

If CCI(A,B) equals positive one then this area can be considered a strength. This occurs, when technology A has an impact on technology B, in the internal R&D but this impact is absent from the competitors' R&D. A potential strategic decision based on this information is presented below: R&D in this area that does not increase value (e.g., it is old-fashioned or no consumers can be identified that are interested in future products from this area) leads to a strategic decision that decreases R&D activities in this area.

A CCI(A,B) of zero leads to two different cases. Firstly, if $BCI{}^{1}\!/_{8}t_{7}(A,B)$ and $BCI^{6}_{17}(A,B)$ equal positive one then technology A has an impact on technology B both in the internal R&D and the external R&D. The R&D activities in this area can be classified as parity. If both cross impact values ($BCI{}^{1}\!/_{8}t_{7}(A,B)$ and $BCI^{6}_{17}(A,B)$) equal zero then there is no impact of technology A on technology B because R&D activities in this area do not intensively occur. Then, the strategic decision to start new internal R&D activities in this area might lead to a strategic decision to start new internal R&D activities in this area might lead to a relative strength in future.

Using a Boolean cross impact index leads to information loss. However, this is more appropriate than using a ratio scale because cutoff values can be determined intuitively (to decide whether there is an impact of technology A on technology B) at an early step and the results are easy to interpret (e.g. a CCI(A,B) of positive one means a relative strength). This makes the approach more transparent to the decision makers. Using a ratio scale instead leads firstly to a normalization of $CI\frac{1}{8}t_{7}$(A,B) and CI^{6}_{17}(A,B) concerning the cutoff values and secondly to a ratio CCI(A,B) score between [-1,..,1]. The higher the CCI(A,B) score the more is the relative strength and the less is the relative weakness. Additionally, the closer the CCI(A,B) is to zero the more is the parity or the probability that there is no impact. Normally, decision makers of organizations preferred results that are easy to interpret created by transparent approaches. Thus, the use of Boolean cross impact indices is preferred in this approach.

Table 1: Result values of CCI(A,B)

$BCI^{6}_{17}(A,B)$	$BCI^{1}/_{8}t_{7}(A,B)$	$CCI(A,B)$
0	0	0 (No impact)
0	1	-1 (Relative weakness)
1	0	1 (Relative strength)
1	1	0 (Parity)

Characteristics of the CCI between technology pairs

The CCI between two technologies can be classified as symmetrical, asymmetrical, or nonexistent. The impact between technology A and B is nonexistent if all four boolean cross impact indexes $BCI_{1/8t_7}(A,B)$, $BCI_{1/8t_7}(B,A)$, $BCI^{6}_{17}(A,B)$ and $BCI^{6}_{17}(B,A)$ equal zero.

Otherwise, if $BCI_{1/8t_7}(A,B)$ equals $BCI_{1/8t_7}(B,A)$ and $BCI^{6}_{17}(A,B)$ equals $BCI^{6}_{17}(B,A)$ then there is an impact of technology A on technology B and a similar impact of technology B on technology A. In this case, the CCI is classified as symmetrical. In the other case, the CCI between two technologies can be classified as having a asymmetrical impact. An example for this is a relative strength concerning CCI(A,B) and a relative weakness concerning CCI(B,A). These characteristics are used to build a CCI network graph (see Sect. 3.4).

CCI network graph

The CCI calculates the relationship between two technologies considering both internal and external R&D. However, each technology can affect two or more technologies and vice versa. Therefore, it is useful to identify the complex relation among three or more technologies. To visually express the relationships between several technologies network analysis - as well-known technique from graph theory [23] - is used. In this graph, each node represents a technology and each edge represents the CCI between two technologies. The direction of the edge shows the direction of the asymmetrical or symmetrical impact.

With the network graph, influencing and influenced technologies can be identified. For example, a technology might influence several other technologies or may be influence by several technologies. For a technology, that influences a large number of related technologies, an increased development and investment also probably increases strength in the related technologies. Additionally, forecasting future trends is easier in technologies that are influenced by a small number of other technologies.

A sequential impact between several technologies (where technology A has an impact on technology B and technology B has an impact on technology C) also can be found in the network graph. Then, the strategic decision to start new internal R&D activities in technology A might lead to an increased strength in technology C.

As an example, Fig. 2 shows a symmetrical relative strength between A and B and it also shows an asymmetrical relationship between A and C as well as between B and C. The impact of C on B represents a parity and the impact of B on C represents a relative strength. Additionally, a relative weakness is seen concerning the impact of C on A and no impact is seen of A on C. Further, a 3-element long sequential relative strength A → B → C can be seen. Last, technology A influences B and is influenced by B and C.

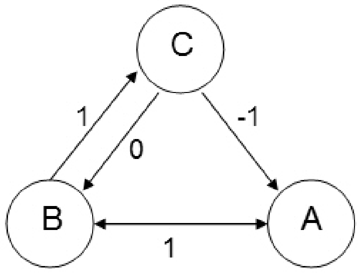

Figure 2: Example for a CCI network graph.

Changes of CCI

The CCI constantly changes over time because it is based on the cross impact with regard to internal R&D and external R&D. It is characteristic for R&D activities of organizations that many new R&D projects start and many existing R&D projects are completed every year. A new R&D project often focuses on a different technology combination and therefore, the impact across technologies changes over time. It is also characteristic for patent data that the impact across technologies changes because of the change in customer needs and the occurrence of new technologies.

The change of the cross impact between technologies concerning internal R&D can be computed by using information from the R&D program of the organization in a specific year. An R&D program is the collection of all active R&D projects. Using this yearly internal R&D information, the cross impact between technologies in a specific year can be identified and used in the CIA approach. Additionally, by collecting the patents that are registered in a specific year the cross impact between technologies in a specific year concerning patent data can be identified.

Then, the CCI and the degree of change can be computed using the proposed compared CIA approach. As an example, CCI(A,B) for 2006 equals positive one and CCI(A,B) for 2007 equals zero with BCI^{6}_{17}(A,B) and $BCI^{1/8}_{t7}$(A,B) both being positive one. Then, the relative strength in the impact of technology A on technology B has become parity. For strategic decision making, this information could be interesting because it shows the impact trend and therefore, it shows which technology should receive more or less development and investment in the future.

Case study ’defence’ - data collection and text classification

Application field

In the last years, the rising asymmetrical threat is causing governments to pay more attention to defence, especially in technological areas. New and ever more complex tasks in areas concerned with defence against these new types of threats require additional R&D of new techniques. For this reason, European governments and the European Union are increasingly funding defence-based technological R&D. For example, the EDA was established in 2004 and coordinates defence-based R&D between State Members of the European Union. Because of growing budgets in the field of defence-based R&D, one can monitor an increasing number of research projects and an increasing collaboration especially between defence-based R&D and (civil) security-based R&D. This leads to a continuous change of the defence-related technological landscape: the appearance of many new technologies and new interrelations and interdependencies between technologies [24]. This is partly due to applied science R&D projects often using several technologies to create a defence application [25].

Technology collection

For this case study, technologies from the application field ’defence’ are needed. A well-known European technology taxonomy in this field is from the EDA. The EDA taxonomy of technologies (CAPTECH) contains about 200 defence-based technologies that are assigned to 32 technology areas. Additionally, EDA provides detailed descriptions for each technology. For this case study, we use all 32 technology areas from this taxonomy as described in Table 2.

Table 2: List of technology areas from EDA taxonomy of defence-based technologies

Number	*Technology area*
A01	Structural & Smart Materials & Structural Mechanics
A02	Signature Related Materials
A03	Electronic Materials Technology
A04	Photonic/Optical Materials & Device Technology
A05	Electronic, Electrical & Electromechanical Device Technology
A06	Energetic Materials and Plasma Technology
A07	Chemical, Biological & Medical Materials
A08	Computing Technologies & Mathematical Techniques
A09	Information and Signal Processing Technology
A10	Human Sciences
A11	Operating Environment Technology
A12	Mechanical, Thermal & Fluid Related Technologies & Devices
B01	Lethality & Platform Protection
B02	Propulsion and Powerplants
B03	Design Technologies for Platforms and Weapons
B04	Electronic Warfare and Directed Energy Technologies
B05	Signature Control and Signature Reduction
B06	Sensor Systems
B07	Guidance and Control systems for Weapons and Platforms
B08	Simulators, Trainers and Synthetic Environments
B09	Integrated Systems Technology
B10	Communications and CIS-related Technologies
B11	Personnel Protection Systems
B12	Manufacturing Processes/Design Tools/Techniques
C01	Defence Analysis
C02	Integrated Platforms
C03	Weapons
C04	Installations and Facilities
C05	Equipped Personnel
C06	Miscellaneous Defence Functions and Policy Support
C07	Battlespace Information
C08	Business Process

Collection of internal R&D

We use R&D projects from the German Ministry of Defence (GE MoD) as internal R&D information. 985 R&D projects from the GE MoD have been identified. The projects are already manually assigned to technologies and therefore also to the technology areas of the EDA taxonomy by use of multi-label classification. This means, each project is assigned to one or several technology areas from the EDA taxonomy.

Collection of external R&D

Patent data are collected from the United States Patent and Trademark Office (USPTO). We use the Derwent Innovations Index to extract patent numbers, titles, and abstracts from the 182,928 patents from the year 2007. Patents from the GE MoD are not collected as well as patents from other organizations and companies where the R&D behind this patent is funded by the GE MoD. Then, patents are assigned to none, one or several technology areas of the EDA taxonomy by use of multi-label text classification.

Centroid-based patent classification

In this case study, we opt for centroid-based patent classification. Below, we substantiate this methodological choice. Centroid-based classifiers have been widely used in many web applications and previous work [29] has shown that the prediction accuracy of centroid-based classifiers is significantly lower than other approaches (e.g., SVM). However, two advantages are important in practice. Firstly, the centroid-based algorithm has a very intuitive meaning [30], which is important because classification results are used as decision support for managers and decision makers of the GE MoD. Secondly, the computational complexity of this centroid-based approach is important given the large number of patents (182,928 patents from the year 2007) and the large number of classes / training examples in the application field ’defence’ (32 technology areas / 200 technology descriptions). In the training phase, the centroid-based algorithm has a linear-time complexity that depends on the number of training examples. We also observe a linear complexity in the classification phase that depends on the number of classes. Hence, the overall computational complexity of the centroid-based algorithm is very low.

Each technology area consists of several technologies (see Sect. 4.2). To acquire training examples for each technology area, we use the descriptions of the respective defence-based technologies from each technology area as reflected in the EDA taxonomy of technologies. Then, terms and term frequencies are extracted and term vectors in a vector space model [31] are built for each training example. For each technology area, we build the centroid vector of all term vectors that are assigned to the technology area. For this, we use tokenization [32], stop word filtering (by use of domain specific stop word list), stemming (by use of Porter stemmer [33]), and manual extraction of prevalent features [34] that are characteristic for a technology area. This centroid vector is used to describe the corresponding technology area.

For classification, patents are used as test examples (see Sect. 4.4). Patent descriptions of these examples are prepared and terms and term frequencies are extracted for each patent. Then, these terms are used to create term vectors. Each term vector from the test examples is compared to each centroid vector using a similarity measure. Here, Jaccard’s coefficient measure [31] is selected because it handles well vectors of different length; e.g. the term vector might have a different length than the centroid vector to which it is compared.

To identify whether a term vector from a test example (patent) is similar to a centroid vector representing a technology area a maximal distance to the centroid vector is determined. A term vector from a test example is defined as similar to a centroid vector if the corresponding Jaccard's coefficient measure is greater than or equal to a user-set α (alpha-cut method [35]). A term vector from a test example (patent) is simply assigned to all classes of its similar centroids. As a result, one can identify none, one or several corresponding technology areas for a given patent. For the case study 'defence' α is set to 0.15 to balance the type I and type II error. If the percentage of α is too small then probably patents are falsely assigned to technology areas (type I error). If the percentage of α is too large then patents are probably not assigned to the technology areas they belong to (type II error).

Results and Discussions

CCI between technology areas

Table 3 shows the results of the case study 'defence'. The technology area pairs are ordered by the CCI score. Within CCI score the technology area pairs are ordered by R&D-based cross impact score CI$^{6}_{17}$(A,B) if CCI(A,B) equals positive one or zero, otherwise they are ordered by the patent-based cross impact score CI⅛ṭ$_{7}$(A,B). The influencing technology area is represented by 'Techn. area A' and the influenced technology area is represented by 'Techn. area B'. Table 3 does not consider technology area pairs with no impact. For each technology area pair, R&D-based and patent-based cross impacts are computed and rounded, i.e. CI$^{6}_{17}$(A,B) and CI⅛ṭ$_{7}$(A,B) respectively. R&D-based cross impacts scores CI$^{6}_{17}$(A,B) exceeding the 0.25 threshold are indicated in bold face and patent-based cross impact scores CI⅛ṭ$_{7}$(A,B) exceeding the 0.20 threshold are indicated in italics. Next, the boolean cross impact scores BCI$^{6}_{17}$(A,B) and BCI⅛ṭ$_{7}$(A,B) are calculated. The BCI$^{6}_{17}$(A,B) and BCI⅛ṭ$_{7}$(A,B) are positive one if the CI$^{6}_{17}$(A,B) and CI⅛ṭ$_{7}$(A,B) are at least 0.25 and 0.20 respectively as described in Sect. 3.2. Finally, the CCI scores are computed. The last column shows that CCI(A,B) is classified as symmetrical or asymmetrical as described in Sect. 3.3.

Table 3: Technology area pairs with high cross impact in 2007

Techn. area A	*Techn. area B*	CI^{6}_{17} *(A,B)*	BCI^{6}_{17} *(A,B)*	CI^{r}_{8t7} *(A,B)*	BCI^{r}_{8t7} *(A,B)*	*CCI (A,B)*	*Sym. Asym.*
B02	A05	**0.54**	1	0.07	0	1	S
B07	C03	**0.39**	1	0.02	0	1	A
A04	B07	**0.34**	1	0.13	0	1	A
C05	B11	**0.32**	1	0.08	0	1	A
A05	B02	**0.29**	1	0.01	0	1	S
A07	A04	**0.25**	1	0.08	0	1	A
B10	B07	0.03	0	*0.26*	1	-1	S
A05	C05	0.16	0	*0.23*	1	-1	A
A05	B10	0.07	0	*0.21*	1	-1	A
B07	B10	0.11	0	*0.20*	1	-1	S
A02	B05	**0.92**	1	*0.58*	1	0	S
A03	A05	**0.86**	1	*0.30*	1	0	S
B05	A02	**0.62**	1	*0.46*	1	0	S
B04	A05	**0.61**	1	*0.35*	1	0	S
A12	B02	**0.58**	1	*0.22*	1	0	A
B08	A08	**0.53**	1	*0.26*	1	0	S
B01	A01	**0.42**	1	*0.27*	1	0	A
B06	A09	**0.38**	1	*0.23*	1	0	A
B07	C02	**0.34**	1	*0.23*	1	0	A
A05	A03	**0.32**	1	*0.26*	1	0	S
A08	B08	**0.31**	1	*0.22*	1	0	S
A05	B06	**0.27**	1	*0.20*	1	0	A
A05	B04	**0.26**	1	*0.21*	1	0	S

For example, let us consider row 1 in Table 3. The number of R&D projects / patents in the technology area B02 ’Propulsion and Powerplants’ is 37 for R&D projects and 563 for patents. The number of R&D projects and patents included both in technology area B02 and A05 ’Electronic, Electrical & Electromechanical Device Technology’ is 20 for R&D projects and 39 for patents. Table 4 explains the calculation of the R&D-based cross impact score CI$^{6}_{17}$(A,B), the patent-based cross impact score CI⅛t$_{7}$(A,B), the boolean cross impact scores BCI$^{6}_{17}$(A,B) and BCI⅛t$_{7}$(A,B) using a cutoff of 0.25 and 0.20 respectively and finally the CCI score CCI(A,B). CCI(A,B) is classified as symmetrical because BCI$^{6}_{17}$(A,B) equals BCI$^{6}_{17}$(B,A) and BCI⅛t$_{7}$(A,B) equals BCI⅛t$_{7}$(B,A).

Table 4: Explanation of calculation of cross impact scores and CCI score for row 1 of Table 3

$CI^{6}_{17}(A,B)$	$= N^{6}_{17}(A \cap B) / N^{6}_{17}(A)$
	$= 20 / 37$
	$= 0.54$
$CI^{18}_{17}(A,B)$	$= N^{18}_{17}(A \cap B) / N^{18}_{17}(A)$
	$= 39 / 563$
	$= 0.07$
$BCI^{6}_{17}(A,B)$	$= 1$
$BCI^{18}_{17}(A,B)$	$= 0$
$CCI(A,B)$	$= BCI^{6}_{17}(A,B) - BCI^{18}_{17}(A,B)$
	$= 1 - 0$
	$= 1$

Relative strength

In the case study, a relative strength for the R&D of the GE MoD can be seen in various technology area pairs where the CCI(A,B) equals positive one (see Table 3). Here, the R&D-based cross impact score $CI^{6}_{17}(A,B)$ is greater than or equal to the internal cutoff value and the patent-based cross impact score $CI^{1/8}_{t_7}(A,B)$ is smaller than the external cutoff value. Below, we describe these technology area pairs.

A focal point of the GE MoD is the R&D to create a MEE (More Electric Engine). 54% R&D projects in the technology area B02 (Propulsion and Powerplants) are also assigned to A05 (Electronic, Electrical & Electromechanical Device Technology) and 29% vice versa. The external R&D is not focused on the combination of these two technology areas B02 and A05. A further core theme is the R&D in fibre optic gyroscope technology for navigation. 34% of all R&D projects from 'Photonic/Optical Materials & Device Technology' (A04) also are assigned to B07 (Guidance and Control systems for Weapons and Platforms). An impact of technology area C05 on technology area B11 can be seen. This is because research in the technology area C05 'Personnel Equipment' is mainly focused on the technology area B11 'Personnel Protection Systems' e.g. to provide significant survivability to the German infantryman. Therefore, 32% of all R&D projects in technology area C05 are also assigned to technology area B11. Additionally, the intensive R&D in guidance and control systems for weapons to reduce collateral damage leads to an impact of technology area B07 on technology area C03 and the intensive R&D for a chemical oxygen iodine laser leads to an impact of technology area A07 on technology area A04. Together with expert knowledge (e.g. the fact that R&D in chemical oxygen iodine lasers probably does not increase value because it might be old-fashioned concerning fibre lasers), an advise can be given to decrease these R&D activities.

These results show that the GE MoD has strength in several technology area pairs and that other organizations (e.g. competitors) do not have strength in these technology area pairs as apparent from the small patent-based cross impact scores CI$^{1}/_{8}t_{7}$(A,B). An organization should aim to build on its relative strength specifically when R&D in these technology area pairs increases value. As such, knowledge about own relative strength and its competitors' relative weakness can be used for R&D planning and strategic decision-making.

Relative weakness

Besides relative strength, Table 3 also portrays a relative weakness for the R&D of the GE MoD in technology area pairs where the CCI score equals negative one. This is the case when the R&D-based cross impact score CI$^{6}{}_{17}$(A,B) is smaller than the internal cutoff value of 0.25 and the patent-based cross impact score CI$^{1}/_{8}t_{7}$(A,B) is greater than or equal to the external cutoff value of 0.20. Below, we describe these relative weakness technology area pairs.

In patent data, a symmetrical impact of navigation technology on communication technology can be found as described in [1]. Here in this case study, we also identify a symmetrical patent-based impact of B07 'Guidance and Control systems for Weapons and Platforms' (that includes e.g. navigation technology) and B10 'Communications and CIS-related Technologies'. However, only a small number of the GE MoD's R&D projects that are assigned to technology area B07 are also assigned to the technology area B10 and vice versa. Further results of the case study are the patent-based impact of A05 (Electronic, Electrical & Electromechanical Device Technology) on C05 (Equipped Personnel) and on B10 (Communications and CIS-related Technologies). Here, it can also be seen that only a small number of internal R&D projects from A05 are assigned to C05 or B10.

These results show that the GE MoD does not have strength in several technology area pairs. However, other organizations often do have R&D projects in these technology area pairs as reflected by the patent-based cross impact score CI$^{1}/_{8}t_{7}$(A,B) exceeding the 0.20 threshold. An organization should aim to reduce its relative weaknesses specifically when R&D in these technology area pairs increase value. If the GE MoD has strength in a technology area like B07 then it can easily gain strength in a technology area like B10 in which it has relative weakness e.g. by R&D outsourcing. From this ’defence’ application it is clear that the knowledge about these relative weaknesses and about the possibilities to bridge these gaps can be used for R&D planning and strategic decision making.

Parity technology area pairs

The case study also identifies R&D technology area pairs being both focal to the GE MoD as well as to other organizations. These technology area pairs appear as third group in Table 3 where both the R&D-based cross impact score CI^{6}_{17}(A,B) as well as the patent-based cross impact score CI⅛t_{7}(A,B) is greater than or equal to the internal or external cutoff value, respectively. Some technology area pairs have a large R&D-based and a large patent-based cross impact e.g. many R&D projects in A02 ’Signature Related Materials’ are also assigned to B05 ’Signature Control and Signature Reduction’. This is because the centroid vectors of A02 and B05 contain similar features. Then, the R&D-based and the patent-based cross impact score are both high and the CCI score equals zero. Further examples for centroid vectors with similar features are the technology area pair A03 ’Electronic Materials Technology’ and A05 ’Electronic, Electrical & Electromechanical Device Technology’ as well as the impact of technology area A12 ’Mechanical, Thermal & Fluid Related Technologies & Devices’ on B02 ’Propulsion and Powerplants’.

Another core theme of GE MoD is R&D for an intelligent smart sensor. Therefore, many R&D projects from technology area B06 ’Sensor Systems’ are also assigned to technology area A09 ’Information and Signal Processing Technology’. The R&D activities can be classified as parity because a patent-based impact of B06 on A09 is also observed. These results show that the GE MoD has strength in several technology area pairs in which other organizations also have strength. A strategic decision to decrease development and investment in a parity technology area pair probably leads to a relative weakness in the future. Therefore, this information can be used for R&D program planning and strategic decision making.

Technology area pairs with no impact

Technology area pairs with no impact are not listed in Table 3 because the R&D-based and patent-based cross impact scores are smaller than the corresponding cutoff values. However, they represent potential future strengths if they receive more development and investment from the GE MoD in the future. An example for using these technology area pairs in R&D planning is given in Sect. 5.3

Characteristics of the CCI between technology area pairs

Table 3 presents examples of symmetrical (S) and asymmetrical (A) impacts. The technology area impact between A05 and B02 is symmetrical. This means that the GE MoD portrays relative strength both for the (A05, B02) technology area pair as for the (B02, A05) pair. If the CCI score is negative one then a symmetrical cross impact can be observed between technology areas B07 and B10. Hence, the GE MoD has a relative weakness in the technology area pair (B07, B10) as well as and in the technology area pair (B10, B07). Additionally, (A03, A05) and (A05, B04) are examples of symmetrical parity cross impacts where the corresponding CCI score is zero.

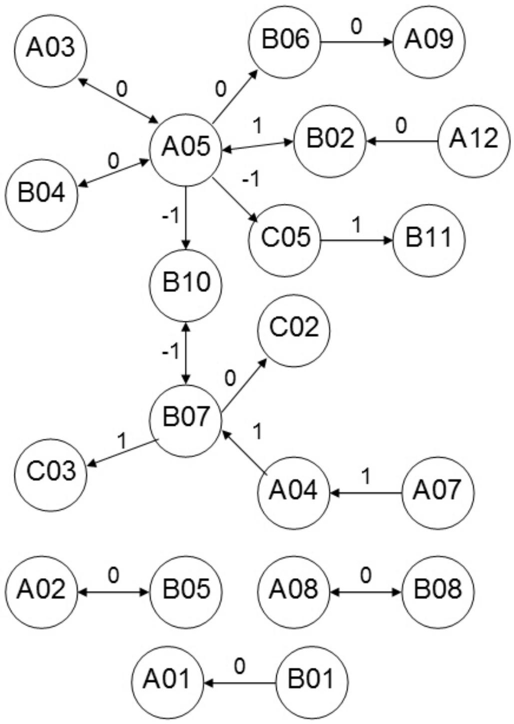

Figure 3: CCI network graph of EDA-technology areas in 2007.

CCI network graph

Based on the results of the case study, a CCI network graph of EDA technology areas is presented in Fig. 3 showing technology area impacts in 2007. With this CCI network graph, the overall structure and the complex relationships between several EDA technology areas can be shown. Each node represents an EDA technology area. Every edge is annotated with its corresponding CCI score that classifies the impact between two technology areas as an R&D-based cross impact; i.e. relative strength (1), a patent-based cross impact; i.e. relative weakness (-1) or an R&D-based and patent-based cross impact; i.e. parity (0). Additionally, the direction of each arrow characterizes the technology area impact as symmetrical or asymmetrical.

The CCI network graph shows the impact of two or more technology areas on a specific technology area. For example the impacts of A03, B02, and B04 on A05 are all symmetrical but differ in whether they display relative strength: CCI(B02,A05) = 1, or parity of the GE MoD: CCI(A03,A05) = 0 and CCI(B04,A05) = 0.

The CCI network graph also reveals the direction of technology area impacts. This way the influencing technology areas and the influenced technology areas can be identified. As an example, the technology area A05 influences six technology areas (A03, B02, B04, B06, B10, and C05). However, it is also influenced by three technology areas (A03, B02, and B04). Non-influenced technology areas are A07, A12, and B01. Each of them only influences one technology area. Additionally, A01, A09, B11, C02, and C03 are examples for non-influencing technology areas.

Three islands can be found in the CCI network graph. The two symmetrical parity technology area pairs are A02 and B05 as well as A08 and B08. The third island represents the asymmetrical impact of B01 on A01.

Sequential impacts between technology areas can be detected as well. For example, a sequential impact starts with A07 via A04 via B07 and it ends with C03. All these impacts are asymmetrical and every corresponding CCI score equals one. This means that the sequential impact represents a relative strength. A further sequential impact with different CCI scores is A12, B02, A05, B10, B07, and C03. Examples for a symmetrical sequential impact are B02, A05, and B04 as well as B02, A05, and A03.

As such, the CCI network graph facilitates the detection of asymmetrical / symmetrical relative strengths or relative weaknesses by showing the structure and the complex relationship between several technology areas. This is helpful information for research planning and strategic decision making. Searching for the edges annotated with -1 immediately indicates for which technology area pairs the GE MoD has relative weakness: in the technology area pairs (A05, C05), (A05, B10), (B10, B07). Likewise scanning for the edges annotated with +1 points out in which technology area pairs GE MoD excels: in (B02, A05), (C05, B11), (B07, C03), (B07, A04) and (A04, A07). From the CCI network graph it is apparent that the GE MoD’s relative strengths are located along the B07 star whereas its relative weaknesses are mainly located along the A05 star. In general, an organization should aim a) to build on its relative strengths and b) to reduce its relative weaknesses. As to the former, the GE MoD should investigate whether it could extend the sequential impact A07 → A04 → B07 → C03. New relative strengths could be (*, A07), (A07, *), (*, A04), (A04, *), (*, B07), (B07,*), (*, C03), or (C03, *) with * referring to any technology area being part of the technology area pair with no impact (see Sect. 5.1.4). The advantage of building upon existing relative strengths stems from the fact that the organization already has experience with one of the technology areas belonging to the new relative strength technology area pair.

Besides building on its existing relative strengths, GE MoD should equally investigate whether it could connect its relative strengths. For the GE MoD turning one of the technology area pairs with no impact (C03, A05), (C03, B02), (C03, C05), (A05, A07), (B02, A07), and (B11, A07) in a relative strength would build on its sequential relative strength at the same time. Given that the GE MoD would gain strength in the technology area pair (C03, C05), the sequential relative strength A07 → A04 → B07 → C03 could be extended with C03 → C05 → B11 to form a 6-element long sequential relative strength A07 → A04 → B07 → C03 → C05 → B11. As such, the GE MoD should initially focus on turning specifically technology area pairs with no impact in a relative strength by increasing development and investment. If it is not possible to gain strength in the technology area pair e.g. (C03, C05) then turning two technology area pairs with no impact (C03, x) and (x, C05) into a relative strength also builds on its sequential relative strength, e.g. x could be technology area A06. This would establish the 7-element long sequential relative strength A07 → A04 → B07 → C03 → A06 → C05 → B11. As such, the GE MoD should initially focus on turning the relative weaknesses in parity technology area pairs and the technology area pairs with no impact in relative strengths.

In summary, the above illustrates how the CCI network graph allows guiding research planning and strategic decision making.

Changes of the CCI

In Table 3 the CCI is computed by use of R&D information and patent data from year 2007. However, technology areas / technologies change and therefore, the CCI as well as the R&D-

based and patent-based cross impact also change. To analyze this change over time, two technology area pairs have been tracked for years 2004 to 2008.

The technology area B06 (Sensor System) has an impact on technology area A09 (Information and Signal Processing Technology) because of R&D for smart (intelligent) sensors. The patent-based cross impact shows a nearly increasing trend from 2004 to 2008 (see Table 5). In the R&D of the GE MoD smart sensor activities become a focal point since 2006. Given that the internal and external cutoff values were set to 0.25 and to 0.20 then no impact of B06 on A09 can be seen in 2004. There is a relative weakness in 2005 because the CI⅛t$_{7}$(B06,A09) is exceeding the threshold (printed in italics). This has led to an increased development and investment by the GE MoD and since 2006 the smart sensor R&D activities can be classified as being at parity because the CI^{6}_{17}(B06,A09) is exceeding threshold (in bold print). An advice for 2008 probably can be that the GE MoD should cut back investment a little bit in this technology area pair to keep the parity with a smaller investment.

A further example is the R&D to create a MEE, which is a focal point in the R&D of the GE MoD since 2005. Patents that deal with electronic, electrical or electromechanical device technology (A05) are normally assigned to other applications (communication, computer systems etc.) but not to propulsions and powerplants (B02). Therefore, a small patent-based cross impact CI⅛t$_{7}$(A05,B02) can be seen from Table 6. Given that the internal and external cutoff values were set to 0.25 and to 0.20 there is no impact of A05 on B02 in 2004, but since 2005 the R&D activities combining A05 and B02 can be classified as a relative strength. This example shows how an increased development and investment in 2005 turn a technology area pair with no impact into a relative strength and it also shows that the value of CI^{6}_{17}(A05, B02) in 2008 is much larger than 0.25. An advice for 2008 probably can be that the GE MoD should cut back investment a little bit by reducing the number of R&D projects in this technology area pair to keep the relative strength with a smaller investment.

Table 5: Change of the (compared) cross impact of technology area B06 on technology area A09 from years 2004 to 2008

	2004	*2005*	*2006*	*2007*	*2008*
CI^{6}_{17}(B06, A09)	0.14	0.14	**0.32**	**0.38**	**0.48**
CI⅛t_{7}(B06, A09)	0.19	*0.20*	*0.24*	*0.23*	*0.24*
BCI^{6}_{17}(B06, A09)	0	0	1	1	1
BCI⅛t_{7}(B06, A09)	0	1	1	1	1
CCI(B06, A09)	0	-1	0	0	0

Table 6: Change of the (compared) cross impact of technology area A05 on technology area B02 from years 2004 to 2008

	2004	*2005*	*2006*	*2007*	*2008*
CI^{6}_{17}(A05, B02)	0.12	**0.28**	**0.28**	**0.29**	**0.36**
$CI⅛t_{7}$(A05, B02)	<0.01	0.01	<0.01	0.01	0.01
BCI^{6}_{17}(A05, B02)	0	1	1	1	1
$BCI⅛t_{7}$(A05, B02)	0	0	0	0	0
CCI(A05, B02)	0	1	1	1	1

Summary and conclusions

This paper introduced an analytical Cross Impact Analysis (CIA) approach to support strategy making and R&D planning for organizations with many R&D projects in areas with many relationships between technologies. Unlike traditional qualitative CIA approaches the newly proposed quantitative CIA approach is able to show relative technology impacts and trends for a large number of R&D projects. The quantitative CIA analyzes the cross impacts between selected technologies based on R&D information of a organization. Additionally, the cross impacts between these technologies based on patent data are computed. Both internal and external cross impacts are compared to compute the relative impact between technology pairs as measured by the newly introduced Compared Cross Impact (CCI) index. CCI indices of positive one point out in which technology pairs the organization excels whereas CCI indices of negative one signify technology pairs in which the organization has relative weakness. Comparing CCI(A,B) to CCI(B,A) indicates whether two technologies are equally influencing one another (symmetrical) or whether the impact between two technologies is different (asymmetrical). As such, symmetrical / asymmetrical relative strengths and relative weaknesses are identified for the organization by inspecting the CCI values. However, to facilitate the detection of symmetrical / asymmetrical relative strengths and relative weaknesses a CCI network graph is introduced as an exploratory management tool supporting the organization's strategy making and R&D planning. The CCI network graph visualizes the overall structure and the complex relationships between several technologies from the organization's perspective. In a glance, managers can detect (sequential) relative strengths and relative weaknesses from the CCI network graph. Finally, the analysis of changes in the CCI values for technology pairs over time reveals trends in technology impacts thereby signaling which technologies should receive more or less development and investment. Overall, the quantitative CIA approach shows that the CCI supports strategy making and R&D planning for organizations with many R&D projects in areas with many relationships between technologies.

The results of the case study show technology impacts and current trends from the application field 'defence'. The selected R&D information from the German Ministry of Defence (GE MoD) is manually assigned to technology areas from the European Defence Agency (EDA) taxonomy of technologies. Patent data are assigned to these technology areas by use of a centroid-based multi-label text classification approach. The R&D-based cross impact $CI^{6}_{17}(A,B)$ is compared to the patent-based cross impact $CI^{1}/_{8}t_{7}(A,B)$ and summarized in the new CCI index CCI(A,B). The CCI between technology area pairs can be used by the GE MoD for research planning and strategy making. For example, the GE MoD has a very strong relative strength in the 'electronic, electrical & electromechanical device technology' (A05) and the 'propulsion and powerplants' (B02) technology area pair. Regardless of the GE MoD's experience with the 'electronic, electrical & electromechanical device technology' (A05), it has a serious relative weakness in the technology area pair 'electronic, electrical & electromechanical device technology' (A05) and 'communications and CIS-related technologies (B10). The construction of the CCI network graph suggested several ways to extend the GE MoD relative strengths as pinpointed technology area pairs with no impact to turn into relative strengths. The analysis of the change in CCI showed that the GE MoD excels in the 'electronic, electrical & electromechanical device technology' (A05) and the 'propulsion and powerplants' (B02) technology areas since 2005. Overall, the 'defence' application illustrates how the compared R&D-based and patent-based cross impact analysis can support an organization's strategy making and R&D planning.

This paper contributed to previous technology impact research in four ways: 1) the introduction of a CCI measure, 2) the characterization of technology impacts as symmetrical or asymmetrical, 3) the presentation of the CCI network graph as exploratory management tool, and 4) the analysis of changes in CCI to discover trends in technology impact. Still there are at least two avenues for future research. The most important avenue of research relates to granularity. The case study focuses on the impact between 32 technology areas. However, a more detailed view at the technology level rather than at the technology *area* level could lead to better R&D planning support and better strategic decision making. Therefore, future research should aim at assigning R&D projects to technologies rather than technology areas. In the case study, internal R&D projects and patent data should be assigned to the 200 defence-based technologies from EDA taxonomy. Then, a more detailed view on the technological landscape in the 'defence' application field could be provided. A second avenue of further research could take the occurrence of new technologies into account. This research focuses on computing the impacts between technologies or technology areas. It does not consider the computation of the occurrence probability of new technologies or technology areas. This could be an interesting topic for future research.

References

[1] C. Choi, S. Kim, Y.Park, A patent-based cross impact analysis for quantitative estimation of technological impact: The case of information and communication technology, Technol. Forecast. Soc. Change 74 (2007) 1296-1314.

[2] F. Narin, E. Noma, Patents as indicators of corporate technological strength, Res. Policy 16 (2/4) (1987) 143-155.

[3] M.E. Mogee, R.G. Kolar, International patent analysis as a tool for corporate technology analysis and planning, Technol. Anal. Strat. Manag. 6 (4) (1994) 485-503.

[4] M. Hirschey, V.J. Richardson, Are scientific indicators of patent quality useful to investors? J. Empir. Finance 11 (1) (2004) 91-107.

[5] B. Yoon, C. Yoon, Y. Park, On the development and application of a self-organizing feature map-based patent map, R&D Manage. 32 (4) (2002) 291-300.

[6] B. Yoon, Y. Park, A text-mining-based patent network: analytical tool for high-technology trend, J. High Technol. Managem. Res. 15 (1) (2004) 37-50.

[7] B. Yoon, Y. Park, A systematic approach for identifying technology opportunities: keyword-based morphology analysis, Technol. Forecast. Soc. Change 72 (2) (2004) 145-160.

[8] D. Thorleuchter, Finding Technological Ideas and Inventions with Text Mining and Technique Philosophy, in: C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), Data Analysis, Machine Learning, and Applications, Springer, Berlin, Heidelberg, 2008, pp. 413-420.

[9] G. Tsoumakas, I. Katakis, Multi Label Classification: An Overview, International Journal of Data Warehousing and Mining 3 (3) (2007) 1-13.

[10] C. Trumbach, D. Payne, A. Kongthon, Technology mining for small firms: Knowledge prospecting for competitive advantage, Technol. Forecast. Soc. Change 73 (2006) 937-949.

[11] G.H. Jeong, S.H. Kim, Aqualitative cross-impact approach to find the key technology, Technol. Forecast. Soc. Change 55 (3) (1997) 203-214.

[12] S. Enzer, Cross-impact techniques in technology assessment, Futures 4 (1) (1972) 30-51.

[13] A. Schuler,W.A. Thompson, I. Vertinsky, Y. Ziv, Cross impact analysis of technological innovation and development in the softwood lumber

industry in Canada: a structural modeling approach, IEEE Trans. Eng.

Manage. 38 (3) (1991) 224-236.

[14] H. Dernis, D. Guellec, Using patent counts for cross-country comparisons of technology output, Special Issue on New Science and Technology Indicators, STI Review, vol. 27, OECD, Paris, 2002, 129-146.

[15] M. Trajtenberg, A penny for your quotes: patent citations and the value of innovations, in: A. Jaffe, M. Trajtenberg (Eds.), Patents, Citations and Innovations, MIT Press, Cambridge, MA, 2002.

[16] T. Gordon, H. Haywood, Initial experiments with the cross impact matrix method of forecasting, Futures 1 (2) (1968) 100-116.

[17] N.C. Dalkey, An elementary cross-impact model, Technol. Forecast. Soc. Change 3 (1972) 341-351.

[18] J.C. Duperrin, M. Godet, SMIC 74 - a method for constructing and ranking scenarios, Futures 7 (4) (1975) 302-312.

[19] R.B. Mitchell, J. Tydeman, R. Curnow, Scenario generation: limitations and developments in cross-impact analysis, Futures 9 (3) (1977) 205-215.

[20] R.W. Blanning, B.A. Reinig, Cross-impact analysis using group decision support systems: an application to the future of Hong Kong, Futures 31 (1) (1999) 39-56.

[21] A. Caselles-Moncho, An empirical comparison of cross-impact models for forecasting sales, Int. J. Forecast. 2 (3) (1986) 295-303.

[22] USPTO, Overview of the classification system, USPTO (2009), Available at, http://www.uspto.gov/.

[23] D. Knoke, J. Kuklinski, Network Analysis, Sage, London, 1982.

[24] H.J Lange, H.P. Ohly, J. Reichertz, Auf der Suche nach neuer Sicherheit: Fakten, Theorien und Folgen, VS Verlag, 2008, p. 11.

[25] D. Thorleuchter, D. Van den Poel, A. Prinzie, Mining Innovative Ideas to Support new Product Research and Development, in: H. Locarek-Junge, C. Weihs (Eds.), Classification as a Tool for Research, Springer, Berlin, Heidelberg, 2010 (To Appear).

[26] A. Hotho, A. Nürnberger, G. Paaß, A Brief Survey of Text Mining, LDV Forum 20 (1) (2005) 19-26.

[27] L. Breiman, Classification and regression trees, in: L. Breiman, J.H. Friedman, R. A. Olshen, C. J. Stone (Eds.), The Wadsworth statistics/probability series, Wadsworth International Group, Belmont, CA, 1984, p. 358.

[28] E.S. Han, G. Karypis, Centroid-Based Document Classification: Analysis and Experimental Results, in: Principles of Data Mining and Knowledge Discovery, Springer, Berlin, Heidelberg, 2000, pp. 116-123.

[29] V. Tam, A. Santoso, R. Setiono, A comparative study of centroid-based, neighborhood-based and statistical approaches for effective document categorization, 16th International Conference on Pattern Recogniton, Iv:235-238, 2002.

[30] S.M. Weiss, N. Indurkhya, T. Zhang, F.J. Damerau, Text Mining, Springer, Berlin, 2005, p.113.

[31] R. Ferber, Information Retrieval, dpunkt.verlag, Heidelberg, 2003, p. 78.

[32] D. Thorleuchter, D. Van den Poel, A. Prinzie, Extracting Consumers Needs for New Products – A Web Mining Approach, in: Proc. WKDD 2010, IEEE Computer Society, Los Alamitos, CA, 2010, p. 441.

[33] M.F. Porter, An algorithm for suffx stripping, Program 14 (3) (1980) 130-137.

[34] K. Coussement, D. Van den Poel, Integrating the voice of customers through call center emails into a decision support system for churn prediction, Information & Management 45 (2008) 165.

[35] A. J. Abebe, V. Guinot, D. P. Solomatine, Fuzzy alpha-cut vs. Monte Carlo techniques in assessing uncertainty in model parameters, in: Proc. 4-th International Conference on Hydroinformatics, Iowa City, USA, 2000.

Dirk Thorleuchter is scientist at Fraunhofer-Institute for Technological Trend Analysis, Euskirchen, Germany. His research interests include text mining and trend analysis for technological R&D planning. Additionally, he consults the Armament Directorate of the German Ministry of Defence concerning new technology trends in security and defence.

Dirk Van den Poel is professor of marketing at the Faculty of Economics and Business Administration of Ghent University, Belgium. His main interest fields are the quantitative analysis of consumer behavior, data mining, text mining, and operations research.

Anita Prinzie is a post doctoral researcher in Economics and Business Administration at Ghent University, Belgium. Her main research interests include choice analysis and classification.