Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 788–795
Expert Systems with Applications
www.elsevier.com/locate/eswa

Grouping of TRIZ Inventive Principles to facilitate automatic patent classification

He Cong *, Loh Han Tong

Department of Mechanical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

Abstract

Automatic patent classification facilitates searching for previous patent documents. For TRIZ users, they would like to search for patents based on the solutions (TRIZ Inventive Principles) to the Contradictions addressed in the patents, which is different from traditional searching for prior arts based on the application fields of the inventions. For this purpose, a TRIZ-based patent classification expert system is needed. To facilitate automatic classification of patent documents according to Inventive Principles (IPs) for TRIZ users, we analyze the original 40 IPs proposed by Altshuller. Seven IPs are defined as Obscure IPs, the other 33 as Distinct IPs. Furthermore, two kinds of similarity among the Distinct IPs are defined: text similarity and meaning similarity. Then the 40 IPs are grouped into 22 new classes. Automatic classification based on 674 patent documents associated with these 22 new classes is tested and analyzed, with two issues of multi-label classification and class imbalance addressed.

Keywords: TRIZ; Inventive Principles; Similarity; Patent document; Automatic classification

1. Introduction

In 1965, Altshuller, the father of TRIZ analyzed 40,000 patents. From them, he recognized that the same fundamental problems in one area had been addressed by many inventions in other technological areas (Terninko, Zusman, & Zlotin, 1998). Based on thousands of patents collected, he summarized 40 fundamental solutions to these problems and named them the 40 Inventive Principles (IPs) (Tate & Domb, 1997b). As a basic tool of TRIZ, the 40 IPs were always introduced as the first concept to the new comers to TRIZ. With the increasingly popular application of TRIZ during recent decades, experts from various technical and non-technical fields have found and validated the wide application of IPs in their fields, with numerous application examples summarized and published (Dourson, 2004; Hipple, 2005; Mann & Cathain, 2001; Retseptor, 2003). However, some criticisms have also been levied upon the original list of IPs summarized by Altshuller: they are too abstract and sometimes overlapped with each other (Mann, 2002). In recent research, the relationship among IPs has been analyzed. In 1998, Williams analyzed the symmetry and asymmetry of IPs and summarized several groups of Principles that are opposite to each other (Williams & Domb, 1998). In 2002, Mann proposed a 5×3 matrix to group most Principles into five different strategies by the space–time-interface entities (Mann, 2002). However, their analysis and combination of the IPs simply make the IPs easier to remember but do not help our purpose of facilitating automatic patent classification.

Patent classification is critical to patent retrieval, which is of benefit for inventors, R&D units and others who need to search previous patents (Fall, Torcsvari, Fievet, & Karetka, 2004). Manually classifying patents takes a huge amount of manpower and time. In addition, the rapid increase of the number of patent applications has also created the need to build automatic patent classification expert systems (Loh, He, & Shen, 2006). In recent decades,

* Corresponding author. Tel.: +65 9790 8529.
E-mail addresses: hecong@nus.edu.sg (C. He), mpelht@nus.edu.sg (H.T. Loh).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.10.015

a number of researchers have reported their works on automatic patent classification. Larkey created an automatic system to classify US patents into US Patent Codes (Larkey, 1998, 1999). Fall, Torcsvari, Benzineb, and Karetka (2003) published their experiments of automatic classification in the International Patent Classification. Gey, Buckland, Chen, and Larson (2001) attributed patent codes in the US and IPC taxonomies . These previous works, however, were based on the classification schemes associated with the application fields involved by patents. This type of classification systems is helpful to search for prior arts but inadequate for the TRIZ users who are in need of patents identified and clustered by the Contradiction addressed and the solutions (IPs) to the Contradiction (Loh et al., 2006). Therefore, a new expert system of this purpose is needed: a TRIZ based automatic patent classification system. We proposed earlier this type of system and presented our experiment based on a small data set of 150 patent documents (Loh et al., 2006). In that paper, however, only 6 out of 40 IPs were taken into account and the experiment was based on two simplifications: single-label classification and class balance, which do not reflect real life. To further the research of automatic patent classification according to IPs, this paper analyzes all 40 IPs, dividing them into Obscure IPs and Distinct IPs according to the text information used to describe them. Then the similarity among Distinct IPs is analyzed. According to the similarity and some other relationships among them, the original 40 IPs were grouped into 22 new classes. Automatic patent classification according to the new classes is then performed and evaluated.

The rest of this paper is organized as follow: in Section 2, we analyze the 40 IPs and define their similarity. In Section 3, 674 patent documents associated with 40 IPs are automatically classified to 22 new classes, with two issues of multi-label classification and class imbalance addressed. Then Section 4 presents and analyzes the experiment results. Section 5 gives a brief conclusion and proposes possible future research.

2. Analysis of IPs

2.1. Obscure Principles vs. Distinct Principles

As mentioned earlier, analysis of Principles in this paper is based on the text information used in the descriptions of inventions. During our research, we found that 33 out of 40 IPs were identified by obvious descriptive information and defined them as "Distinct Principles" in this paper (He & Loh, 2005). For instance, if an invention uses Principle

Table 1
Obscure Principles

Obscure Principles[a]	Analysis
2 Extraction	If #[b] 2, extraction, is used, usually the descriptions only describe the extracted objects or systems, without specifying the relationship of extraction by obvious text information. e.g. • Use the sound of a barking dog as a burglar alarm (The sound of the dog is extracted from a physical dog, thus we could get the benefit of the alarm without the difficulty of feeding, exercising, training, etc. But there is no clear text that describes this)
3 Local Quality	The general idea of #3 is to enable local parts of a system to perform locally different or optimized functions. However, little distinct text information for this is contained in the instance descriptions. e.g. • Freezer compartment in refrigerator
12 Equipotentiality	This Principle emphasizes on changing the conditions to achieve a certain effect. However the example description usually describes the condition after being changed, with little obvious text information about the relationship of changing
13 Inversion	The general idea of #13 is to turn an object or action upside down. However, the example descriptions seldom specify the "upside down" relationship between two objects or actions. They instead directly describe the object or action after being changed. e.g. • Cool the inner part (instead of heating the outer part)
17 Moving to a new dimension	It is hard to cluster the examples using #17. To understand #17 needs sophisticated semantic analysis. And it is nearly impossible to see from the examples text description itself that this Principle is being applied. e.g. • Five-axis cutting tool • Infrared computer mouse • Cassette with 6 CD
24 Mediator	Documents involving #24 usually do not specify that an object is used as an intermediary. They instead describe the usage or functions of the object. (no obvious text information) Furthermore, the intermediaries used in different inventions are not the same or similar. (no common or similar text information shared). e.g. • Play a guitar with a plectrum • Joining papers with a clip
33 Homogeneity	Make objects interact with a primary object of the same material, or a material similar in behavior. When #33 is used, usually it only specifies the result of using the same material while text information like "same" or "identical" does not appear. e.g. • Full-strength rod design (making the heads of pipes the same material as the wire rope to improve durability)

[a] Please refer to (40 Invention principles with examples; Terninko et al., 1998) for detailed descriptions of each Principle.
[b] "#" represents the Principle number.