Technology opportunity identification combining SAO semantic analysis and link prediction

Jia Li 1107210296@qq.com **School of Management and Economics, Beijing Institute of Technology**

Xuefeng Wang wxf5122@bit.edu.cn **School of Management and Economics, Beijing Institute of Technology**

Yujia Hou hyj946@126.com **School of Management and Economics, Beijing Institute of Technology**

Jiayun Wang wangjiayun1223@163.com **School of Management and Economics, Beijing Institute of Technology**

Yuqin Liu liuyuqin2004@126.com **Beijing Institute of Graphic Communication**

Keywords: SAO(subject-action-object) network; link prediction; technology opportunity identification; 3D printing

Technology opportunity identification has been regarded as a crucial process for companies due to the success of many entrepreneurs who have identified and exploited these opportunities(Lee and Kim et al., 2014). Most of the researchers focus on finding emerging technologies based on Keyword-based analysis(KWA), which is a widely-used content-based method to identify technology opportunities. However, it is only based on the co-occurrence among the keywords and can't express their relationship. That is to say, it cannot represent how a technology is used and how it interacts with other technologies. Thus, Subject-Action-Object(SAO) analysis is generally used to solve the shortcoming of the keyword-based method. But previous studies simply analyze the SAO network characteristics using degree analysis, centrality analysis or other SNA (social network analysis) indices(Choi and Yoon et al., 2011; Yoon and Choi et al., 2011), having no predictive effect on future links among nodes, or combine S and O that are not connected by manual observation, time-consuming, laborious and lacking of prediction accuracy, which is the problem we are to solve.

We build a SAO network by using SAO structures extracted from the abstracts of the patent documents and predict links between S and O having no links at present by link prediction as Figure 1 shows.

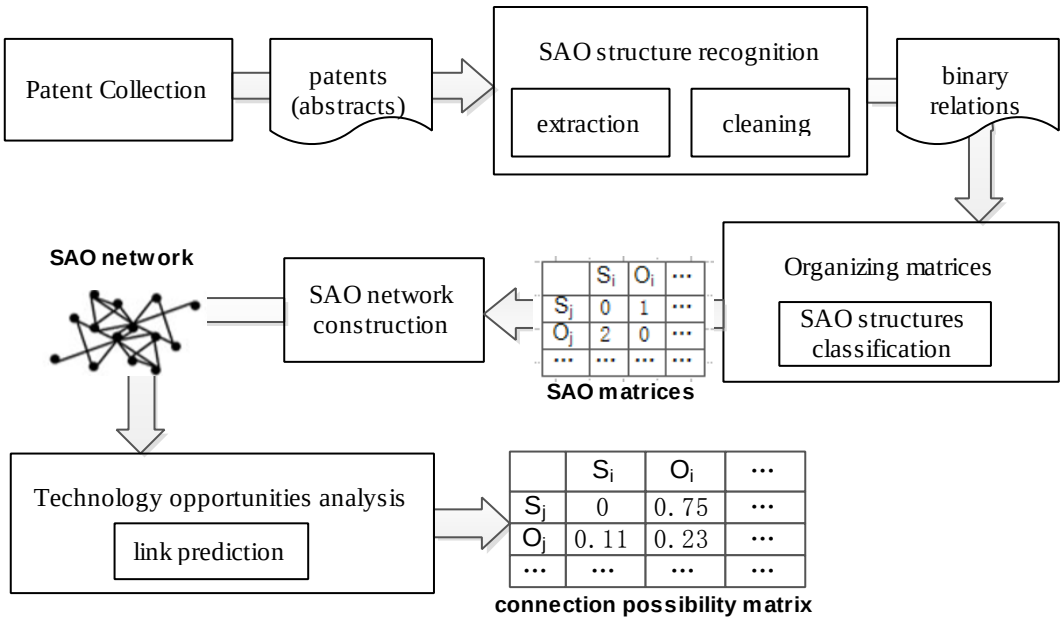

Figure 1 research procedure

The specific steps are as follows:

1. SAO structure recognition.

This paper extracts the SAO structures using Stanford Open IE from the patent abstracts and clean these structures that have a lot of redundant and meaningless data by means of setting rules, the stop-words list and thesauri. Then, we classify the SAO structures based on the kind of action. There are kinds of relations in SAO structures, but we only care about the kinds "technology-realize-function" and "technology-contain-technology". So, this paper induces information of "action" and extract the two types of structures. The process of SAO structure recognition is shown in Figure 2.

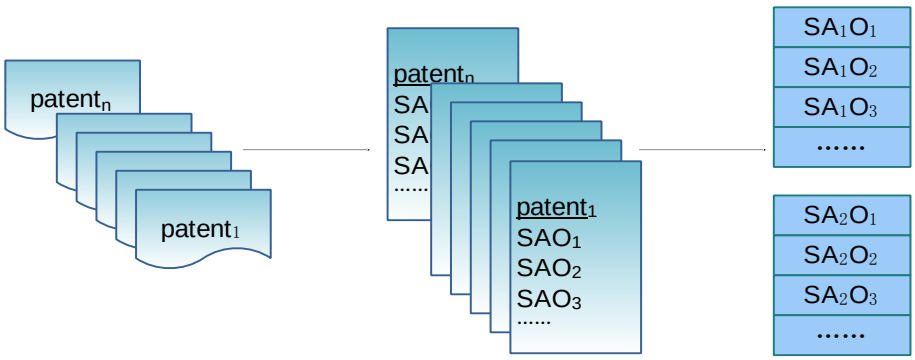

Figure 2 Process of SAO structures identification

2. SAO network construction.

We construct two sub-networks from the whole network according to the two SAO structures produced in the last step. As is shown in Figure 3, the process to construct SAO network consists of two steps. First, we generate a S-O matrix on the basis of SAO structures. The matrix only contains S and O because the relation of them is the same in one matrix. We simplify the network for the convenience of link prediction in the back steps. Second, we build a S-O network based on the S-O matrix. The network is directed and weighted. The weight

represents the number of relationship between S_i and O_i and the direction of relationship is row-to column.

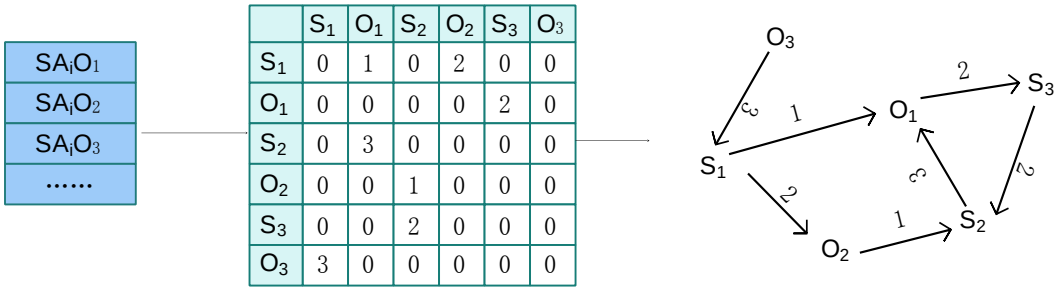

Figure 3 Process of constructing SAO network.

3. Technology opportunities analysis

This paper adopts the method of link prediction to measure the connection possibility of two nodes having no links at the moment. We regard the network as unauthorized and undirected and make a preliminary prediction using Common Neighbors(CN), Adamic-Adar(AA) and Resource Allocation(RA)(Lue and Zhou, 2010).

This paper uses a patent dataset of 3D printing which builds objects by layer-by-layer printing based on digital model files as a case study. The case study to measure the connection possibility between two nodes indicates the feasibility of our method and accuracy of link prediction.

References:

Choi, S. and J. Yoon, et al. (2011). "SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells." SCIENTOMETRICS **88** (3): 863-883.

Lee, Y. and S. Y. Kim, et al. (2014). "Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis." SCIENTOMETRICS **100** (1): 227-244.

Lue, L. and T. Zhou (2010). "Link prediction in weighted networks: The role of weak ties." EPL **89** (180011).

Yoon, J. and S. Choi, et al. (2011). "Invention property-function network analysis of patents: a case of silicon-based thin film solar cells." SCIENTOMETRICS **86** (3): 687-703.