Similarity Measurement among Technologies using Euclidean Distance

Dae-hyun Jeong 1, Young-il Kwon 2

Abstract—Recently, various technologies for conserving and using energy resources efficiently have been developed worldwide, among which wind energy and fuel cells have been receiving much attention as renewable energy, and LED application field has received attention in the aspect of energy usage. This study aims to build an IPC network among technologies using patent information on wind energy, fuel cells, and LED application field, and to investigate the connectivity structure between technologies with network clustering. As a result, it was concluded that technologies in the fields of wind energy, fuel cells, and LED application were interconnected. When analyzing the yearly similarities between technologies using Euclidean distance, it was estimated that the technological similarities in the fields of LED application and fuel cells have been increasing.

Keywords—Clustering, Network Analysis, Euclidean

I. INTRODUCTION

In response to the climate change due to the global warming, movement toward conservation and efficient use of energy resources has affected the entire process of social and economic activities. If the current level of energy consumption continues, the CO$_2$ level is expected to go up to 1,000 ppm—double the current level-in 2030, and the average temperature will increase by 6°C.

Therefore, major advanced countries in the world have been striving to develop renewable energy that can replace the existing fossil fuels, as well as energy efficient technologies.

Wind energy and fuel cells are renewable energy-related technology fields that generate energy using unpolluted energy sources, whereas LED application technology field utilizes the generated energy efficiently. Many research projects regarding the renewable energy and energy efficiency technology have been under way mainly by major advanced countries. In this study, networks were built among technologies using IPC information among patent information on wind energy, fuel cells, and LED application technology fields.

In addition, the relationships between technologies were analyzed through clustering, and the recent fusion technology trends were investigated through the distances between technologies in the fields of wind energy, fuel cells, and LED application each year.

II. THEORETICAL BACKGROUND

A. Network Analysis

A network represents a certain relationship among multiple persons, organizations, or objects. Those persons, organizations, or objects that form the network are called actors. The configuration of the network arrangement among actors is called a network structure, and the activity of describing and analyzing this structure is called network analysis or social network analysis. Network analysis is based on a theoretical premise that a network structure (a shape of a network surrounding actors) affects the behaviors and attitudes of actors, meaning that the behaviors and attitudes of actors can be specified to a certain extent if a network structure is known.

In the studies of network analysis, some conduct qualitative analysis by only using the concept of network analysis and its theoretical premise, although network analysis usually implies quantitative analysis, which will be described below.

In many cases, network analysis investigates the relationships among all actors in a certain area. If an actor is a person, information about the actor’s relationships with other actors is collected by taking conduct of on all persons in the target area. Therefore, network analysis does not employ random sampling, which is typically used in a poll. Particularly, within a large population, it is very unlikely that there exist any relationships between actors that are randomly sampled (Scott, 1991)[1].

In addition, the social network theory is based on the graph theory, which is a mathematical model representing the relationships between element pairs in each specific set, expressed by the nodes and the links connecting the nodes (KISTEP, 2008:14). The entire structure of the network, the characteristics of the links, and the influences of nodes can be explained by analyzing the shapes of nodes or links. Social network analysis is one of the analysis methods based on this network theory, and it is widely employed in various fields today, such as sociology, anthropology, geology, and medicine [2].

Unlike degree centrality that measures the centrality of a node, betweenness centrality measures centrality using the degree of betweenness between a node in a network and a specific node that is connected to another node. Betweenness of a specific node is calculated by the percentage of the number of specific nodes existing on the actual minimum distance to the

1 Dae-hyun Jeong is with Korea Institute of Science & Technology Information, Hoegi-ro 66, Dong Daemun-gu, Seoul, Korea (phone: +82-02-3299-6238; fax: +82-02-3299-6117; e-mail: ydkoo@kisti.re.kr)

2 Young-il Kwon is with Korea Institute of Science & Technology Information, Hoegi-ro 66, Dong Daemun-gu, Seoul, Korea (phone: +82-02-3299-6031; fax: +82-02-3299-6117; e-mail: ylkwn@kisti.re.kr)
number of minimum distances among all pairs of nodes except
the specific node. In other words, betweenness centrality
measures the degree of betweenness between different nodes.
Therefore, the applicable technology in the technological
network is likely to be a fusion technology in all fields [3].

The objectives of network analysis are to investigate the
relationships between social beings such as persons,
organizations, or countries through a network, and to find out
the shapes and content of the network structure. Network
analysis is a quantitative method for analyzing the interactions
between the nodes using a graph that visualizes the
relationships of the nodes within a system. Particularly, this
method has provided a powerful analysis means to social
network researchers who have been studying social bonding,
connections, and networks through quantitative analysis of the
specific concepts that were adopted in network analysis, such
as density, centrality, and structural equivalence. In the studies
of organization theory and policy network, this method has
been used for social Network Analysis (SNA) or network
theory [4]. There are various fields of research where they use
network analysis: Jin-soo Jung (2009) proposed a new
interpretation of historical events through the network analysis
of people by studying the kings from Goguryeo, Baekje, and
Shilla Dynasties; Cheesung Park (2009) studied organizational
diagnosis methods using social network analysis; Seunggook
Song (2011) discovered that social network analysis of
 technological improvement is affected by the existing level of
technology prior to the execution of the project. Therefore,
the consideration of effective investment methods for national
R&D projects in this study is meaningful because it uses the
social network analysis that is widely used in various fields [5],
[6].

B. Cluster Analysis

Cluster analysis is an exploratory analysis method to help
understand the structure of all the data by grouping similar
individual members into different groups and investigating the
characteristics of each group. The ultimate purpose of this
analysis is to effectively divide the entire set of data by using
clusters solely by depending on data. Clustering should be done
in a way that individual members of a cluster will have
distinctive characteristics according to their own cluster.
Cluster analysis can find a meaningful data structure without
having prior information on the inner structure of given data,
and it can be applied to most types of data if the distance
between observation units is defined according to a data type.
However, the interpretation of cluster analysis is difficult
because the results largely depend on the distance that
represents similarity between the observation units, and there
are no purposes given in advance. Cluster analysis is roughly
divided into hierarchical and nonhierarchical cluster analyses.
The Single Linkage Method, Complete Linkage Method, and
Average Linkage Method are used for hierarchical cluster
analysis, whereas the K-Means is widely used for
nonhierarchical cluster analysis. Since there is no special scope
of application for the cluster study, comparative studies on
clustering are actively under way [7].

In addition, betweenness centrality is frequently used for
network analysis. Cluster analysis using betweenness centrality
repeatedly eliminates betweenness centrality values between
nodes in the applicable network structure until a cluster is
formed by the changes in the network structure, as other
betweenness centrality values change. Since this clustering
depends on the relationship between the nodes, it is different
from network classification, which simply uses a cutoff value
[8].

C. Distance factor

Distance factors indicate dissimilarity. The Knowledge
Matrix (developed by KISTI), which was employed in this
study, uses three types of distance factors: squared Euclidean
distance, Euclidean distance, and Minkowski distance, among
which Euclidean distance was used for this study. Since the
range of distance factors is not between 0 and 1, it was
standardized first before calculating the similarities between
distance factors.

III. EXPERIMENTAL METHODS

For network implementation and clustering for LED
application, fuel cells, and wind energy, the total of 50,000
pieces of data were sampled through keyword search using
WINTELIPS provided by WIPS Co., Ltd. The entire network
was implemented through IPC information that was multiply
classified by patent. The Netminer 3.0 developed by Cyram
was used for implementing network, and the Knowledge
Matrix developed by KISTI was used for co-occurrence matrix
to implement the Netminer 3.0.

IV. RESULTS

A. Cluster Analysis

Network analysis was conducted on the entire data using IPC
information among the patent data that concern wind energy,
fuel cells, and LED application fields. Here, the network was
clustered using the betweenness Clustering method. As a result,
total three clusters were revealed as shown in the figure. IPC
information on each cluster indicates that clustering was done
according to wind energy, fuel cells, and LED application
fields.

Betweenness centrality in each cluster showed that in Cluster
A, H02J-007/00 (circuit device for charging or depolarizing
storage battery, or power supply from storage battery to load)
has the highest betweenness centrality, whereas H01B-013/00
(device or method specifically used for manufacturing
conductors or cables) in Cluster B and H01L-031/042 (that
includes the panel or arrangement of photoelectric cells) in
Cluster C have high betweenness centrality. When simplifying
the network structure by a cutoff value higher than 30, both fuel
cells and LED application technology are connected to wind
energy as shown in the figure. Main technologies connecting
each cluster are H02J-007/00, F21S-009/03(lightin...
that can be charged through light), and F21W-131/103 (distance or road technology with lighting) in the fields of LED application and wind energy. Wind energy and fuel cells were found to be connected through H01M-008/06 (combination between the means of reagent manufacturing and residue treatment and fuel cells) technology.

**TABLE I**

<table>
<thead>
<tr>
<th>Group</th>
<th>IPC</th>
<th>Betweenness centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>H02J-007/00</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>F03D-009/00</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>H02N-006/00</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>H02J-003/38</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>F03D-009/02</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>H01B-013/00</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>H01M-008/06</td>
<td>0.027</td>
</tr>
<tr>
<td>Group B</td>
<td>H01M-008/00</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>H01M-008/04</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>H01B-001/06</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>H01L-031/042</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>H05B-037/02</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>G09F-009/00</td>
<td>0.042</td>
</tr>
<tr>
<td>Group C</td>
<td>H01L-033/00</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>F21Y-101/02</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>F21S-009/03</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>F21W-131/103</td>
<td>0.003</td>
</tr>
</tbody>
</table>

**B. Regression analysis**

**Euclidean Distance**

Regression analysis on yearly Euclidean distance in the fields of LED application and fuel cells indicated—shown in the figure—that the technological similarities in the fields of LED application and fuel cells have been rapidly increasing by 0.0246 per year after 2006. And the applicable increase rate of the $R^2$ value of 0.93 and the P-Value of 0.001 showed reliability in the 99% confidence interval.
Fig 4. Yearly Euclidean distance in the fields of LED application and wind energy

The results for the yearly Euclidean distance in the fields of LED application and wind power indicated the decreasing technological similarities between two technology fields, as shown in the figure. And the R2 value of 0.845 and the P-Value of 0.001 showed reliability in the 99% confidence interval.

V. CONCLUSION

In this study, technologies were clustered by patent IPC information on the fields of LED application, wind energy, and fuel cells, and betweenness centrality was analyzed in each IPC network. Also, the yearly distances among fields were analyzed with Euclidean distance factors. As a result, it was concluded that each field has established a technologically independent cluster, although all the fields are interconnected. Technologies with high betweenness centrality in the LED application technology field also act as connectors to other technology fields, whereas fuel cells and wind energy have a relatively stronger inner connection. This is because more attempts for fusion technology have been made in the energy usage field than in the energy generation field. The yearly distances between technologies showed that the distances between technologies in the fields of LED application and fuel cells have drastically increased, indicating that LED application technologies using fuel cells are actively under development. However, the content analyzed in this study is only a part of the entire fields of energy source and usage, and the analysis is needed for the additional relationships with the storage field such as the secondary battery.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea Grant founded by the Korean Government (MEST) (NRF-C1AAA002-2012-0001006).

REFERENCES


