The most popular topics in IEEE journals in 2014

Satoko Kora
Illinois State University
skora@ilstu.edu

April 15, 2015

1. Introduction

IEEE stands for Institute of Electrical and Electronics Engineers. It is the worlds largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE has a large variety of contributions to the advancement of technology, known as a vast group of IEEE members around 430,000 people over 160 countries, and as a leader of international standards of emerging technologies, and sponsored more than 1600 conferences.

In terms of publications, IEEE owns a third of the worlds technical literature in electrical engineering, computer science, and electronics. Approximately 180 transactions, journals, and magazines are published annucally. IEEE Xplore Digital Library is an online library containing more than 3.5 million documents from IEEE and IEEE journals, transactions, magazines, letters, conference proceedings, and active IEEE standards.

This paper will explore the worlds most highly cited IEEE journals and reveal the most popular topics in academic journals published by IEEE in 2014. For this research, IEEE Xplore Search Gateway service is used to retrieve statistical data. IEEE Xplore Search Gateway is a service providing information about all publications stored in IEEE Xplore Digital Library. Among abundant information about academic journals, I will retrieve controlled index term information which represents a set of keywords and main topics dealt with in each academic journal, and gather data showing how frequent each term is chosen as a topic in academic journals published in 2014. In this way, I would like to reveal the top 5 popular topics in IEEE journals in 2014.

2. Obtain data

In order to retrieve dataset from IEEE Xplore Search Gateway, I used "IEEER" package available on GitHub, developed by kbroman. The package contains methods to make a query to access to the library based on our criteria and handles memory consumption for users when they retrieve a large amount of data from the enormous library. To run the methods, it is required to install the package with devtools package for the first time.

```
##install.packages("devtools")
##library(devtools)
##install_github("ropensci/IEEER")

#load a package to query IEEE Xplore Search Gateway
library(IEEER, quietly=TRUE)
```

The following is sending a request to retrieve data about open-access academic journals published by IEEE in 2014.

```
# make a data request query and get data from IEEE Xplore Search Gateway
# pu=IEEE: published by IEEE
# ctype="Journals": academic journals only
# pys=2014:Start value of publication year
# pye=2014:End value of publication year
# oa=1: Open Access journals only
# rs=1:Sequence number of first record to fetch is 1
# limit: maximun number of reterieved journals must be 2000

ie3data ← suppressMessages(IEEE_search(query = list(pu="IEEE", ctype = "Journals", pys =
    2014, pye = 2014,oa=1,rs=1), limit=2000))
```

The total number of corresponding journals is 1196.

3. Scrub data

After obtaining corresponding information, controlled index terms, which represents keyword and topics dealt within journals, are extracted from the data. Journal information is retrieved in a data frame form, and firstly the column "controlledterms" is extracted.

```
# extract only Controlled Index Terms data in academic journal information
terms←ie3data[,c("controlledterms")]
```

Usually an academic journal includes more than one index term, so multiple words are stored as one string sepated with vertical bar. The following is splitting those strings and making one vector of all retrieved index terms.

```
# split vectors of controlled index terms into an array
index ← unlist(strsplit(terms, "[|]"))
```

In the following, a ranking of top 5 popular topics is created. Table function counts the number of times of each word's appearnce in the vector. The table is sorted by the frequency in descending order and top 5 terms are set into table. No.1 popular topic term is set to top varaiable.

```
# Count frequencies of keywords and
sortedTable ← sort(table(index), TRUE)


#get top 5 keywords
top5←sortedTable[1:5]

# Make a bar chart based on the data
df ← data.frame(top5)
# get No.1 keyword
top←row.names(df)[[1]]
```

According to the processed data table, No.1 popular topic in 2014 is medical image processing, and the number of its appearance in journals in 2014 is 59 times.

4. Explore data

In this section, I will focus on metadata of the data that represents the number of appearance of all terms used in the journals, stored in the variable "sortedtable" addressed above.

```
class(sortedTable)
```

```
[1] "array"
```

```
str(sortedTable)
```

```
 int [1:2326(1d)] 59 42 42 40 40 35 34 33 33 32 ...
 - attr(*, "dimnames")=List of 1
  ..$ index: chr [1:2326] "medical image processing" "learning (artificial intelligence)" "
      silicon" "elemental semiconductors" ...
```

```
summary(sortedTable)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   2.000   3.151   3.000  59.000
```

class() method shows the class name of the object. In this case, array is this object type.

str() displays the internal structure of the object, and the data clarifies that sortedTable contains the frequency that each word appears in the journals, and the frequency numbers are treated as a numeric object. In addition, each number is associated with a controlled term index.

summary() is a generic function used to produce result summaries of specified objects. the summary of sortedTable is the summary of the number of the words' appearance. The top keyword, medical image processing is appeared 59 times, and the number is the maximum number of the summary. From Median and Mean value, we can observe that each term is appeared 3 times on average within published journals in 2014, thus we can say on average approximately 3 journals have dealt with the same topic.

5. Results

The following is the complete dataset of the scrubbed data. The data represents top 5 most frequently appeared index terms within academic journals published by IEEE in 2014.

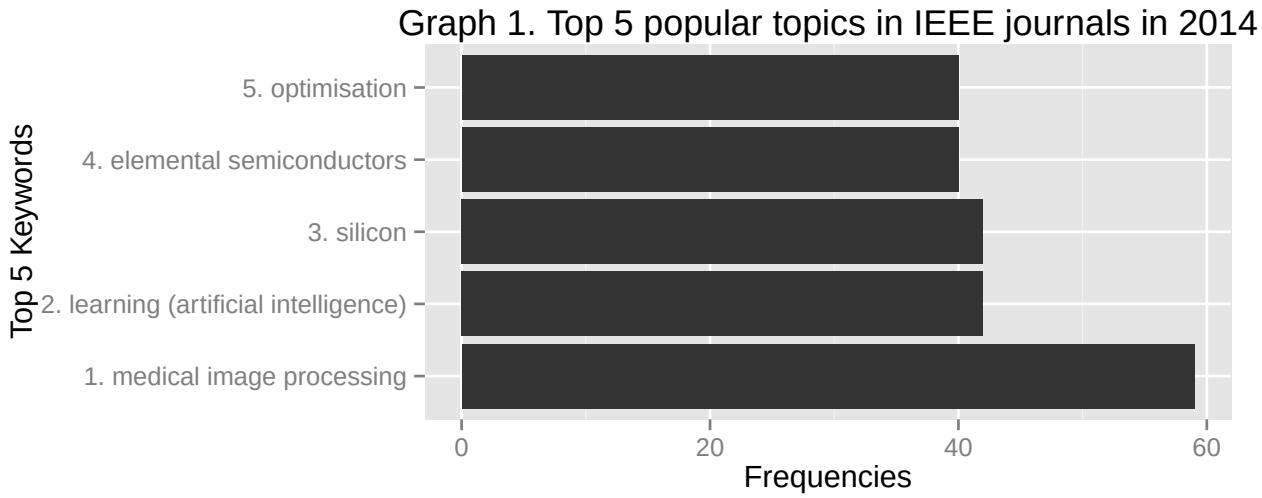
```
library(xtable)

## Printing a table using xtable ##
caption <- "Top 5 keywords in IEEE publications in 2014"
c1<-c("1.","2.","3.","4.","5.")
row.names(df)<-paste(c1, row.names(df), sep=" ")
colnames(df) <- c("Frequencies")
print(xtable(df, caption=caption),
      caption.placement="top", include.rownames=TRUE)
```

Table 1: Top 5 keywords in IEEE publications in 2014

	Frequencies
1. medical image processing	59
2. learning (artificial intelligence)	42
3. silicon	42
4. elemental semiconductors	40
5. optimisation	40

The following figure is a bar chart of the ranking table above.

```
# load a package to make a graph
library("ggplot2",quietly=TRUE)


g <- ggplot(df, aes(x=row.names(df), y=Frequencies))
g + xlab("Top 5 Keywords") + geom_bar(stat="identity") + coord_flip() + labs(title = "Graph
    1. Top 5 popular topics in IEEE journals in 2014")
```



Based on these results, we can assume that these terms represent that they are considered as topics today's reserachers focus on and where a lot of investments and contributions are involved in these studies.

- **medical image processing** : Medical image processing is the technique and process of creating visual representations of the internal body. Produced images through the process are used for clinical analysis to identify diseases and perform medical examinations. There is a variety of methods to perform medical image processing, such as radiography, magnetic resonance imaging(MRI), nuclear medicine, ultrasound...etc.

- **learning (artificial intelligence)** : Artifical intelligenece is the intelligence exhibited by machines and software. It is an academic field of study whose ultimate goal is to create intelligence. There are different approaches for the research of AI, and learning is the study of algorithms that enables AI improve its intelligenece through experience.

- **silicon** : Silicon is a chemical element with symbol Ai and atomic number 14. Silicon is recognized as the eighth most common element in the universe by mass, and widely distributed in sands and dusts, and

over 90% of Earth's crust consists of silicate minerals. For commercial use, silicon is included in crays, silica sands and stones used for the industrial building construction. In addition, silicon is widely used in integrated circuits, which is a necessary component for computers. Because of the large penetration of computers in the world, we are highly dependent on this element and it is a huge impact on today's world.

- **elemental semiconductors** : According to the definition of Wikipedia, "A semiconductor material has an electrical conductivity value falling between that of a conductor, such as copper, and an insulator, such as glass.. Semiconductors are the foundation of modern electronics. Semiconducting materials exist in two types - elemental materials and compound materials." The elemental materials is focused on this topic, and they belong to IV group of the periodic table, and silicon is one of the most important elements in the group. Therefore, we can assume that the study of this topic is often associated with silicon, the term addressed above.

- **optimization** : Optimization is a abstract word but used in different fields of technial studies. In technology and industril fields, this term is used in the meaning of improving efficiency, quality, and desirability. In Mathematics, mathematical optimization signifies the theory of and computation of extrema or stationary points of functions.

The next graph focuses on the most frequently appeared topic, medical image processing, and represents top 5 countries that published the topic related journals the most.

As the way top 5 keywords are extracted, the following program has access to IEEE Xplore Search Gateway. This time, an additional search keyword is added: controlled index term must contain "medical image processing."

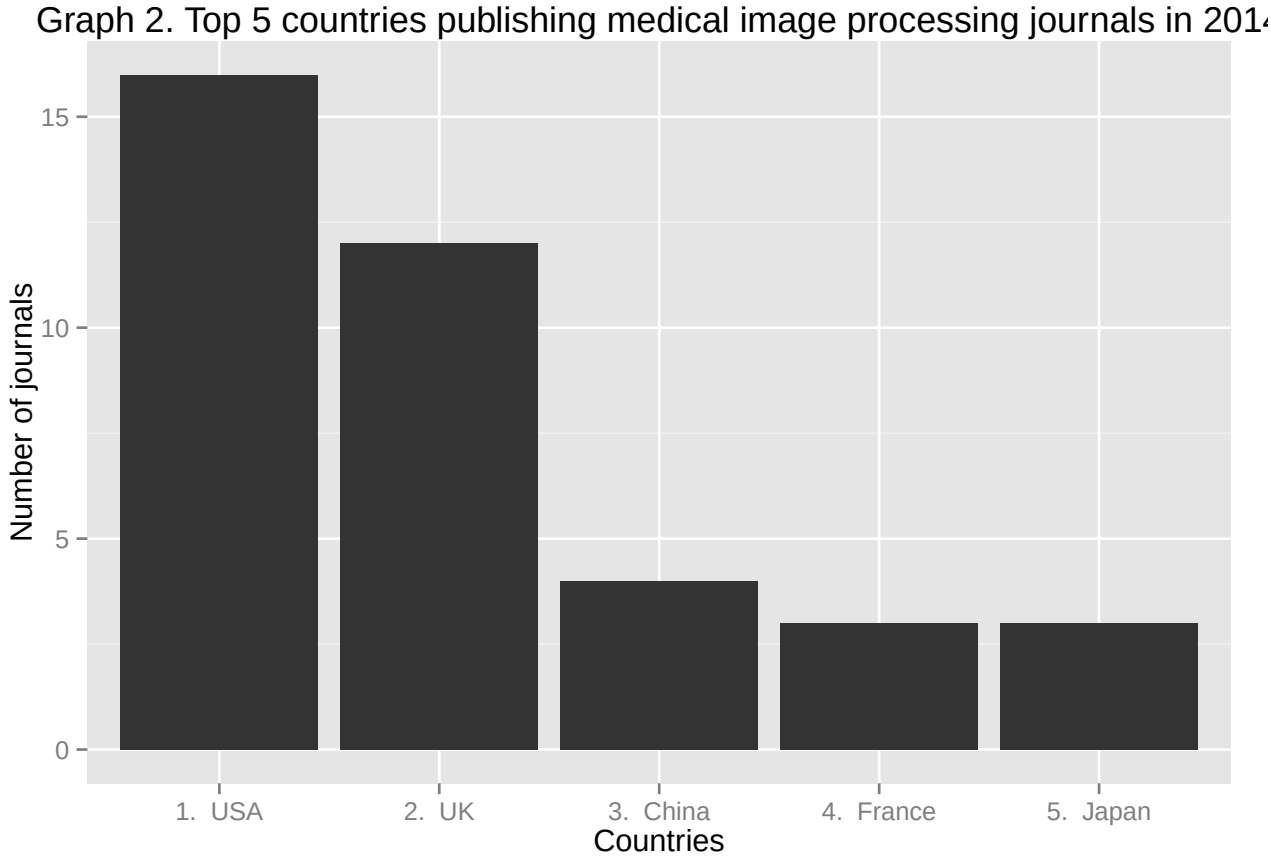
```
# make a data request query and get data from IEEE Xplore Search Gateway
# pu=IEEE: published by IEEE
# ctype="Journals": academic journals only
# pys=2014:Start value of publication year
# pye=2014:End value of publication year
# oa=1: Open Access journals only
# rs=1:Sequence number of first record to fetch is 1
# limit: maximun number of reterieved journals must be 2000
##### cntrlterms= medical image processing ← No.1 keyword
artilcles ← suppressMessages(IEEE_search(query = list(pu="IEEE", ctype = "Journals", pys =
    2014, pye = 2014,oa=1,rs=1, cntrlterms="medical image processing"), limit=300))

# although the term is specified in the previous query, the query executes LIKE search of the
    term # so additional journals (e.g. Biomedical image processing), so it is required to
    grep the retried # data using the exact  keyword again
artilcles.gr ← artilcles[grep("medical image processing", artilcles$controlledterms),]

# Extract affiliation information
affs←artilcles.gr$affiliations

# Split all words into pieces by comma separator
arrAffs ← strsplit(affs, "[,]")

# Retrieve the last element of each vector
# since the last element is the name of the coutry
countries ← c(1:length(arrAffs))
j ←1
for(i in 1:length(arrAffs)){
  countries[i] ← arrAffs[[i]][length(arrAffs[[i]])]
}


# Count frequencies of country names and get top 5 countries which publishes journals the most
sortedTable ← sort(table(countries), TRUE)[1:5]

# convert table into data frame
df2 ← data.frame(sortedTable)


# Add rank label
colnames(df2) ← c("Num")
c1←c("1.","2.","3.","4.","5.")
cols←paste(c1, row.names(df2), sep=" ")
```

```
## Printing a table using xtable ##
caption ← "Top 5 countries publishing medical image processing journals in 2014"
print(xtable(df2, caption=caption),
      caption.placement="top", include.rownames=TRUE)
```

Table 2: Top 5 countries publishing medical image processing journals in 2014

	Num
USA	16
UK	12
China	4
France	3
Japan	3

```
# Make a bar chart
g ← ggplot(df2, aes(x=cols, y=Num))
g + xlab("Countries") + ylab("Number of journals") + geom_bar(stat="identity")+ labs(title = "
    Graph 2. Top 5 countries publishing medical image processing journals in 2014")
```



The results shows No.1 country which published the largest number of journals related with medical image processing is 1. medical image processing, and the number of journals is 59.