Statistical Analysis on NIPS Conference Papers

Liu Ping Guo Feifei

March 27, 2017

Dataset:
NIPS Confercence Papers 1987-2015 dataset.
11463 words $\times$ 5811 papers

Objective:
Do some statistical analysis to gain the development of the conference, field and some specific methods.

Outline:
Clustering analysis of words
Clustering analysis of articles
Correlation analysis of words

Clustering analysis of words

Clustering of words

- select the words with high frequency according to the summation of the word frequency numbers
- select the words with more useful information according to the variance of each word
- sum the number of words in each year to reduce data dimension and meanwhile obtain more accurate information
- do clustering of these words using K-means method and determine the optimal numbers of clusters by sum of the total within sum of square

Clustering result

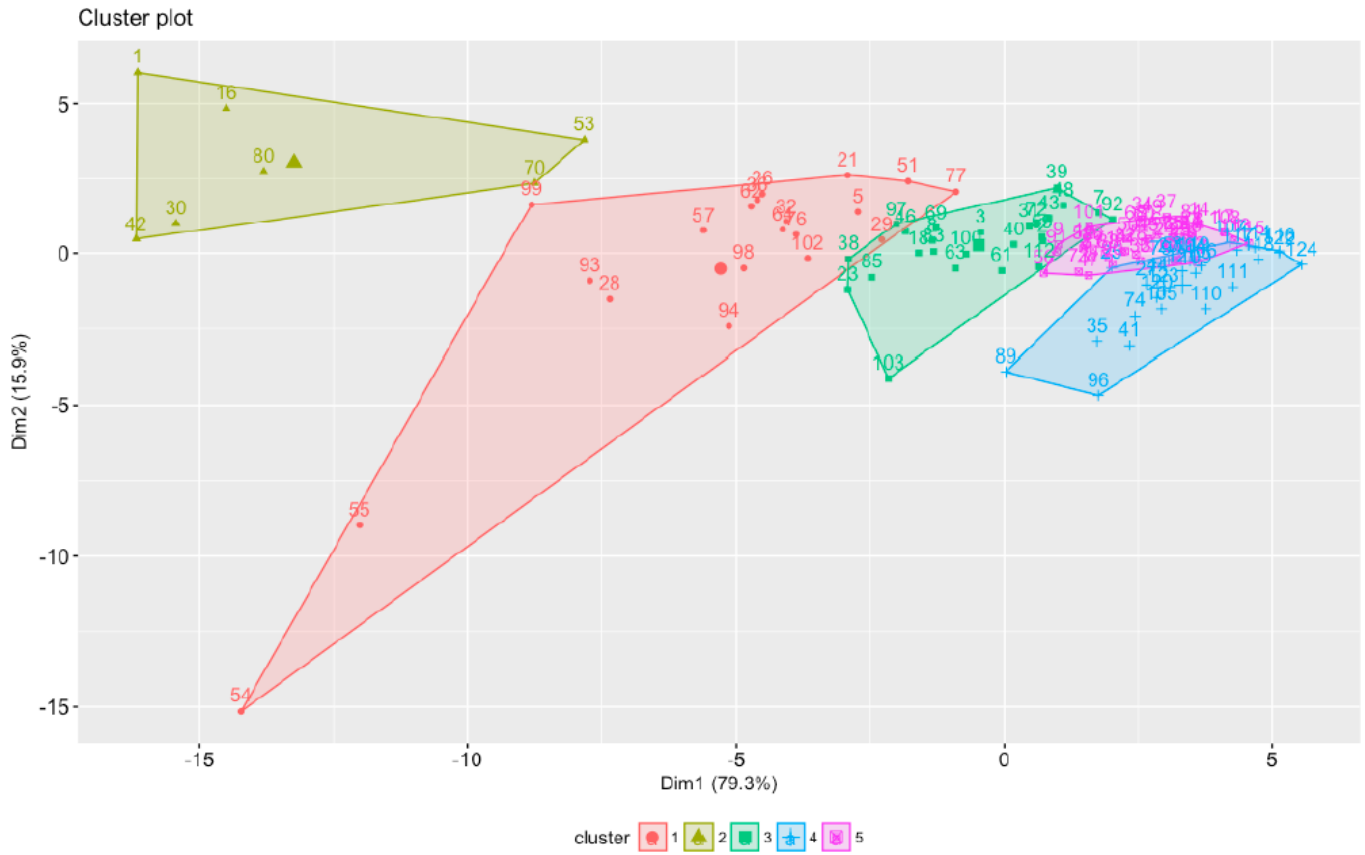


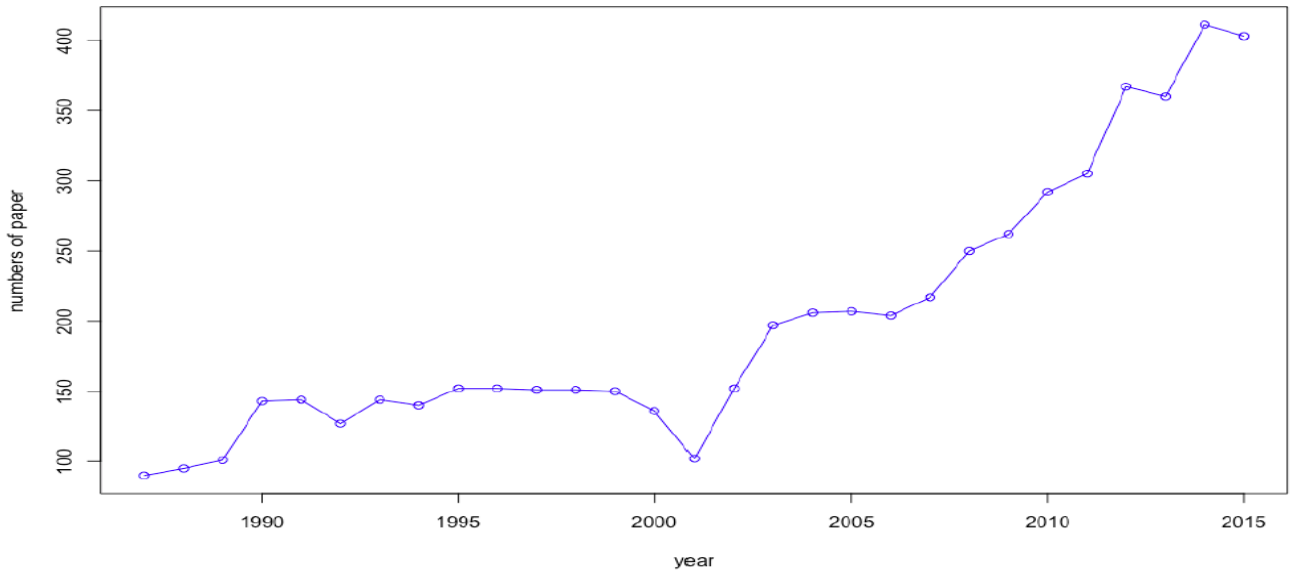
Figure: Clustering result

Clustering result

Category	Word
I	image neural network training
II	algorithm data function learning
III	optimal gaussian kernel random
IV	deep visual layer recognition
V	convex bayesian sparse machine

Table 3.1: Partial exhibition of classification result

Clustering analysis of words



The paper number every year has a steady and large increase from year 2001 to 2015.
It indicates the popularity of the conference and the field recent 15 years.

Cluster result

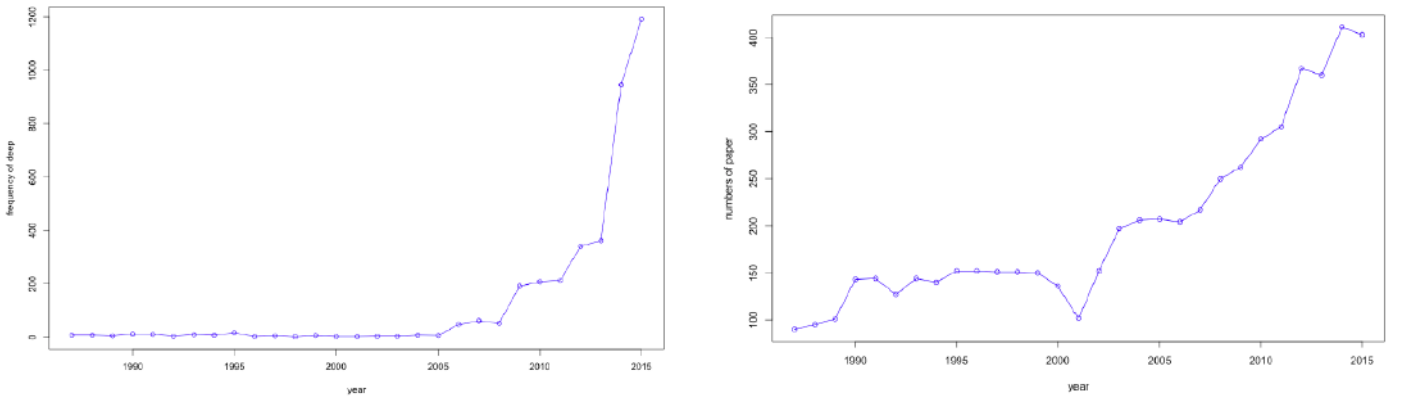


Figure 3.3: The change in number of "deep" and paper number

There is a sharp change in the word "deep".
Research of deep learning explosively increases in year 2014 and 2015.

Clustering analysis of words

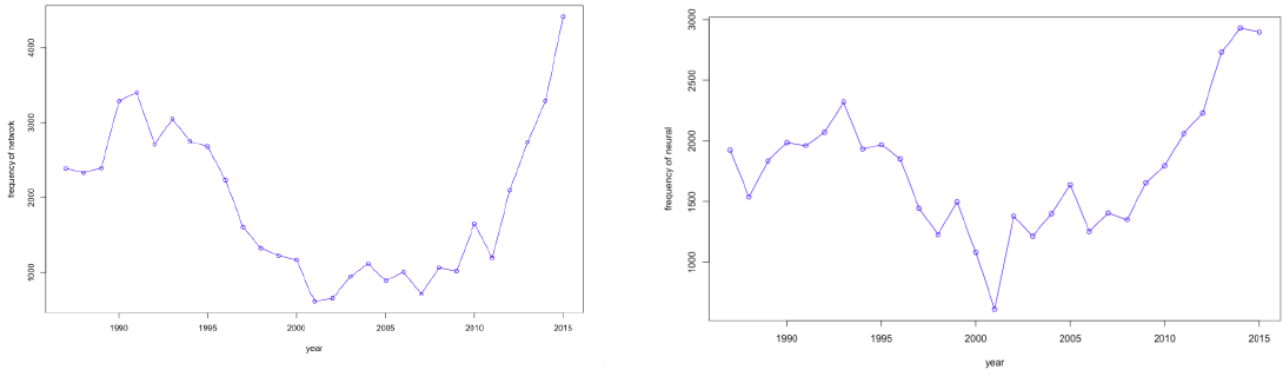


Figure 3.4: The change in number of "network" and "neural"

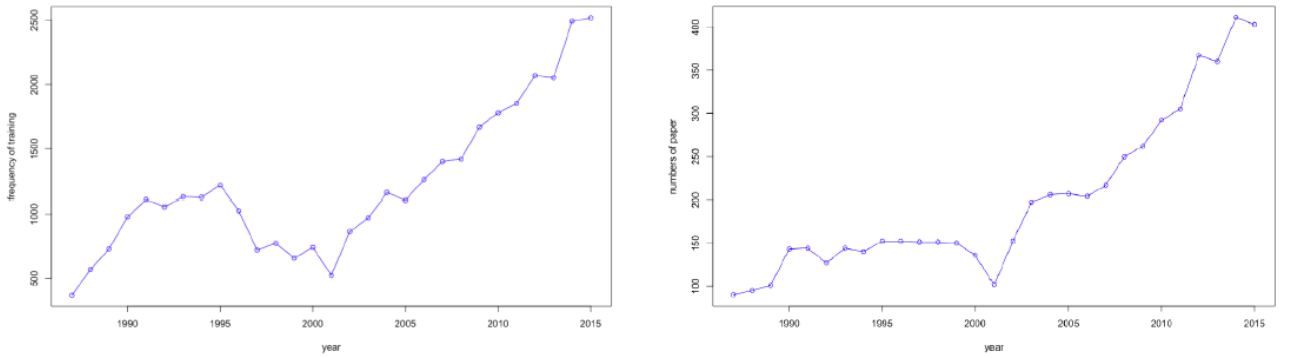


Figure 3.5: The change in number of "training" and paper number

Clustering analysis of words

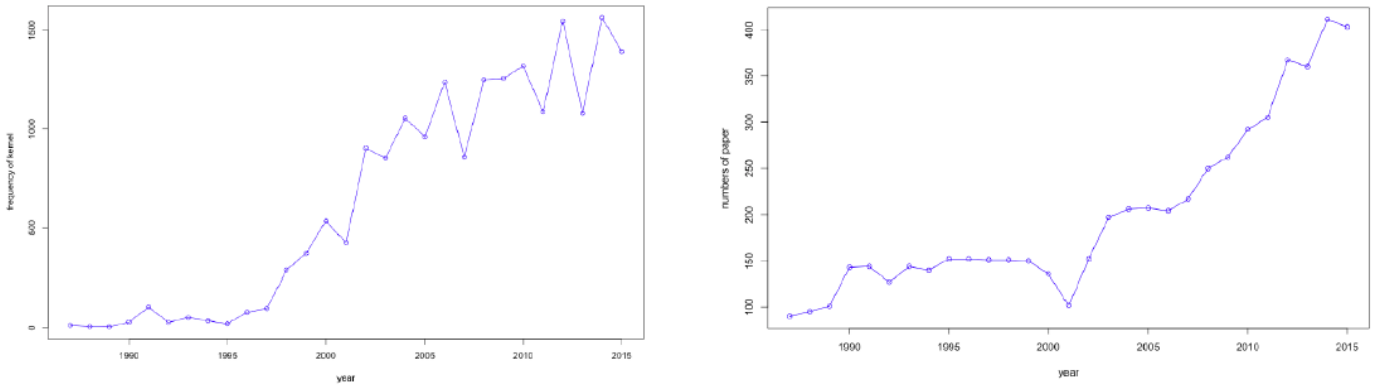


Figure 3.6: The change in number of "kernel" and paper number

Kernel may represent the kernel method and it start to increase in year 1997, while the paper number doesn't change or somehow decreases.

Around year 1997, there must be some breakthrough in the kernel method or in applying it into neuroscience and machine learning.

Clustering analysis of articles

Besides machine learning and neuroscience, other fields represented at NIPS include cognitive science, psychology, computer vision, statistical linguistics, and information theory[Wikipedia].

This gives us inspiration to classify all the articles into several categories to figure out some information.

Clustering analysis of articles

Field	Word
Neuroscience	neutron neutrons
Machine learning	machine learning
Cognitive science	cognition cognitive
Computer vision	vision visual
Information theory	entropy entropies

Table 2.1: Words indicate specific fields.

By searching information of the different field we determine some specific words to cluster the articles.

Clustering analysis of articles

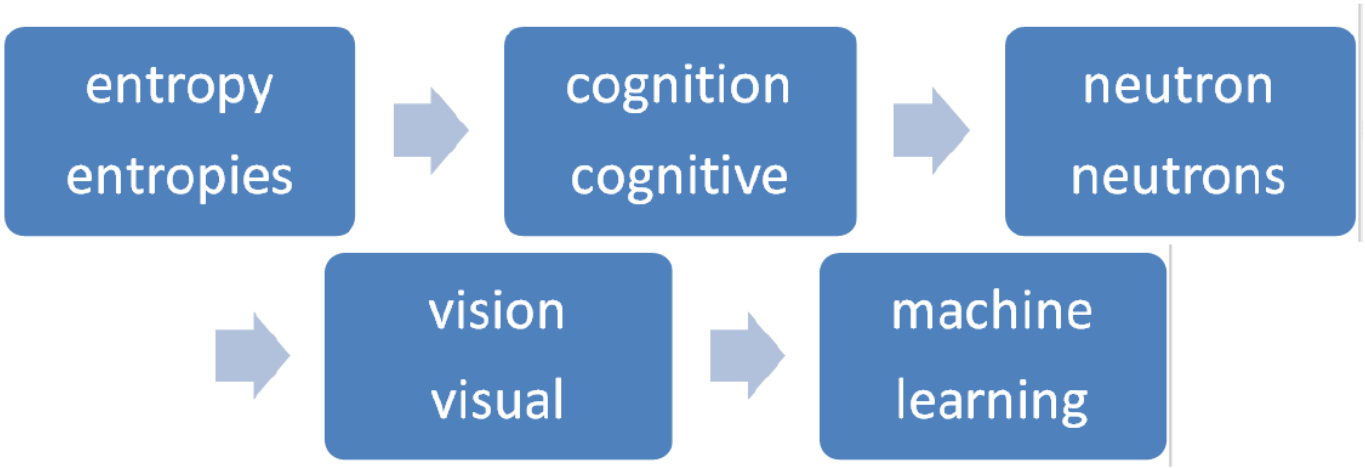


Figure 2.1: Priority rule of words

Then we determine a priority rule to justify the article. According to the rule we have a classification of 500 artilces as a training set. Then we use SVM to classify all 5811 papers.

Clustering analysis of articles

Field	quantity
Neuroscience	2158
Machine learning	2791
Cognitive science	244
Computer vision	111
Information theory	103
Undetermined	404
Total	5811

Table 2.2: Whole article quantities 1987-2015.

Clustering analysis of articles

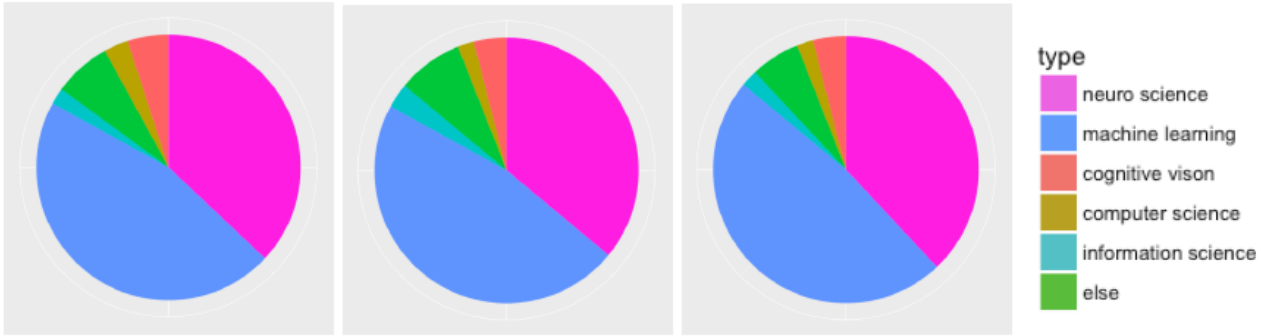


Figure 2.2: distribution of papers 1987-1995 1996-2005 2006-2015

The proportion of each topic hasn't changed too much when considering the total number over 10 years.

Clustering analysis of articles

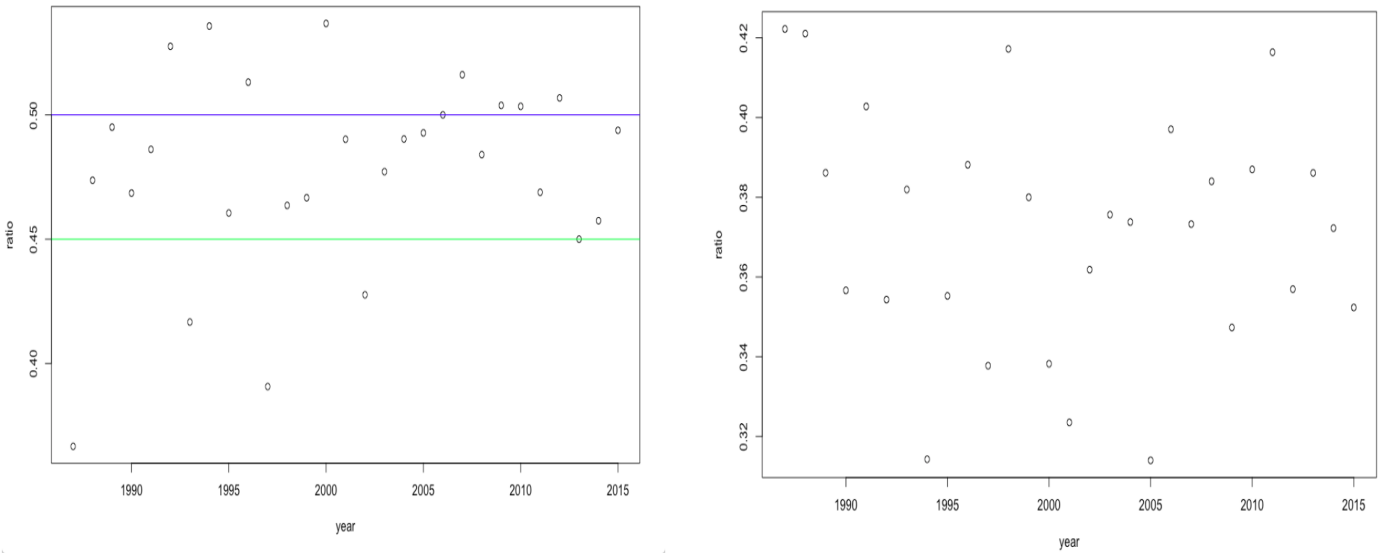


Figure: Proportion of machine learning and neuroscience every year

We find that the conference hosts will make some trade-off in selecting papers from different topic to make the proportion changed slowly. Because we can see that the difference of proportion of both machine learning and neuroscience in neighbored year is usually less than 0.1.

Correlation between words

Word	Correlation
likelihood	0.7299
prior	0.7243
regression	0.6169
stochastic	0.6047
estimation	0.6033

Table 4.1: Correlation with "Bayesian"

Word	Correlation
learning	0.6148
solution	0.6138
performance	0.6050
value	0.6048
linear	0.6038
bound	0.593

Table 4.2: Correlation with "optimal"

We do the correlation analysis to some words.
This result may provide some guides for literature retrieval.

Thanks!