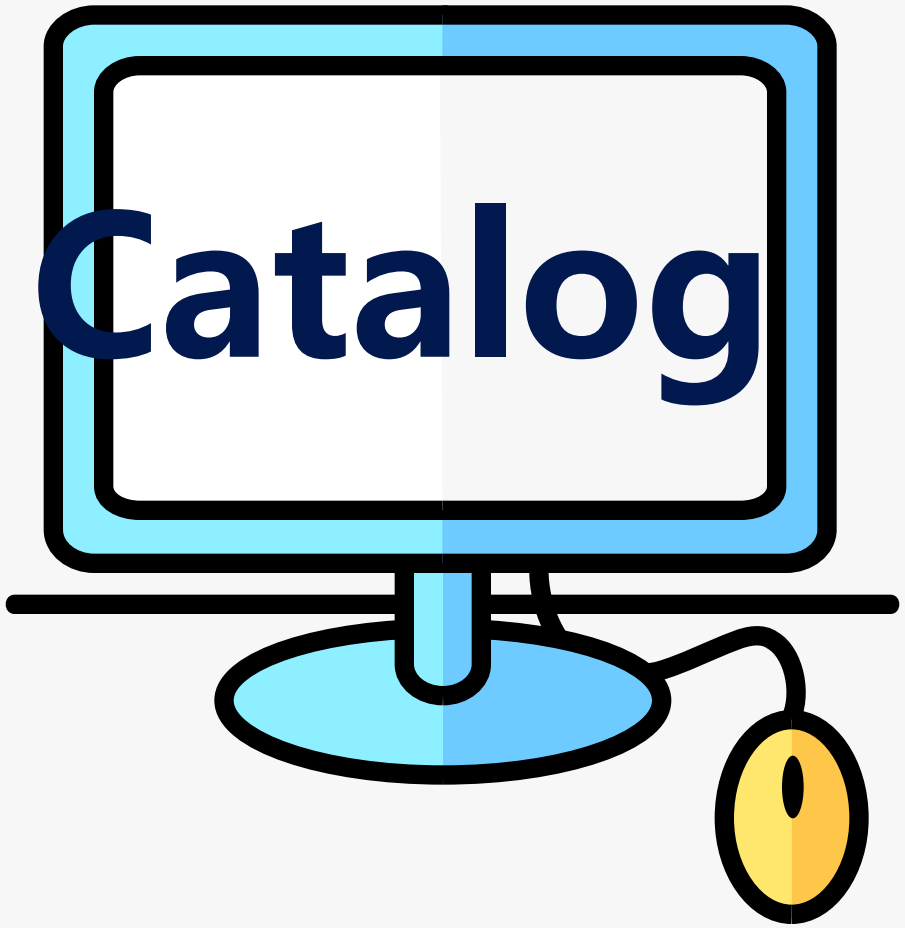
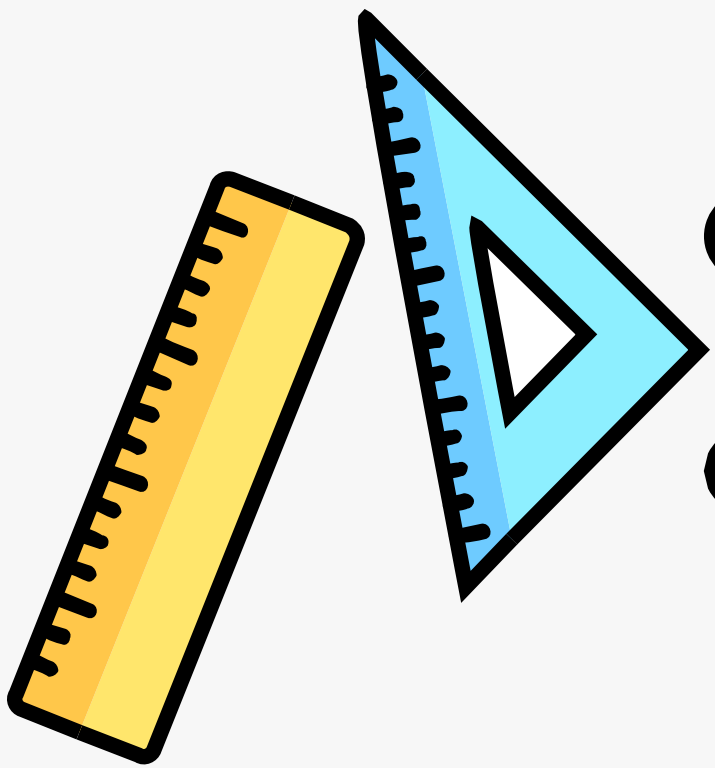
Project Report

Who is more likely to gain a large number of citations?

513030946
5孙元璞





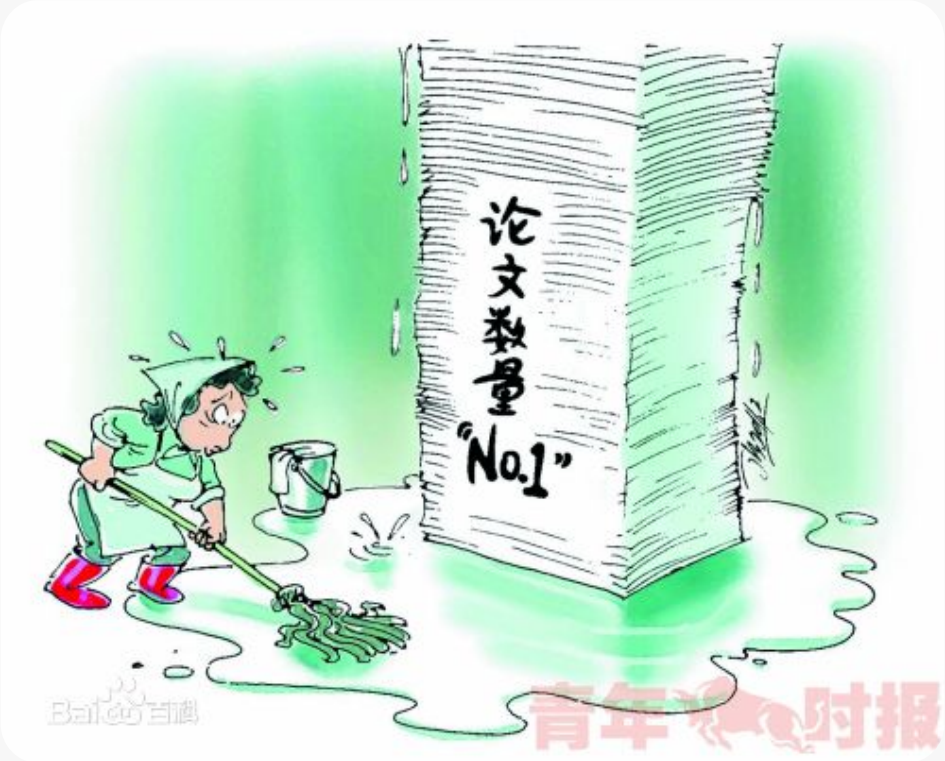
Background&Meaning

Background

The cited rate

Reflect the value of a papr to some extent.

Need to be increased instead of the paper number.



Meaning of this project

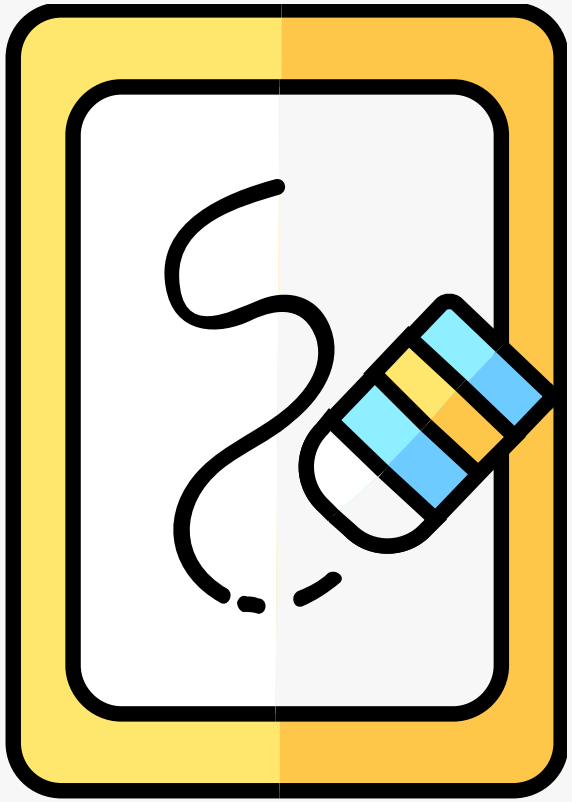
-for searching

high cited rate≠high quality

to find the more useful one

-for writing

taking example by their writting skills



Idea&Method

Idea&Method

Analyze the motivation

Summarize the peculiarities

Find the key component

Correlation analysis

Predictive Model

Analyze the motivation

- Provides a new direction
- Supplements information
- For comparison
- Use of formula
- Practice of method
- New method of proof

Summarize the peculiarities

- Innovative topic
- Innovative researching method
- Annual Review
- Published on high influenced periodical
- The researcher has his own website of research group

Key Components

- Title
- Abstract
- Key words
- Co-authors
- Quotation
- Published time

Correlation analysis

Chart analysis

Correlation index analysis

Covariance analysis

Comentropy&Mutual information analysis

H-index

h- means high citations

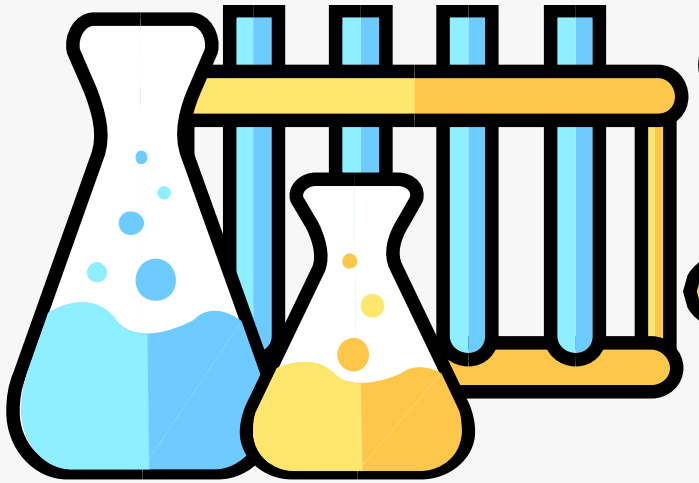
h-index means a researcher has at most h papers that are cited at list h times

Rank	Name	h-index	Field
1	Whitesides, G. M.	155	Organic
2	Karplus, M.	139	Theoretical
3	Corey, E. J.#	138	Organic
4	Heeger, A. J.#	128	Organic
5	Huber, R.#	122	Bio
6	Wüthrich, K.#	120	Bio
7	Bax, A.	118	Bio
8	Schleyer, P. v.	117	Organic
9	Lehn, J. M. #	114	Organic
10	Bard, A. J.	113	Analytical
10	Gratzel, M.	113	Physical
10	Hoffmann, R.#	113	Theoretical
13	Schreiber, S. L.	112	Bio
14	Scheraga, H. A.	111	Bio
15	Fersht, A. R.	105	Bio
15	Frechet, J. M.	105	Inorganic
15	Truhlar, D. G.	105	Theoretical
18	Marks, T. J.	104	Inorganic
18	Trost, B. M.	104	Organic
20	Gray, H. B.	103	Inorganic

Predictive Model

Statistical Learning Theory has provided a very effective framework for classification and regression tasks involving features. Support Vector Machines (SVM) are directly derived from this framework and they work by solving a constrained quadratic problem where the convex objective function for minimization is given by the combination of a loss function with a regularization term (the norm of the weights). There are two main categories for support vector machines: support vector classification (SVC) and support vector regression (SVR). SVM is a learning system using a high dimensional feature space. It yields prediction functions that are expanded on a subset of support vectors.

The model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that is close to the model prediction. Support Vector Regression is the most common application form of SVMs.



EXPERIMENTS&EVALUATION

EXPERIMENTS

Randomly choose 1000 papers from the MAG

Train the SVM model

Predict 8 papers from 8 different researchers

EVALUATION

Name	Yong Yu	Jiawei Han	Kai Li	Yuanyuan Zhou	Dina Katabi	Garth Gibson	Michael I Jordan	Tom Mitchell
Paper count	66	283	131	39	79	51	186	73
H-index	53	159	80	56	64	63	146	76
Citation Count(2017)	1801	13025	4124	1071	2347	1170	14146	2941
Predict Count	1635	14353	4527	779	2420	937	16323	3466

More Features to Improve the Predict Model

Topic Rank

Diversity

Productivity

Sociality

Thanks for listening