# predicting the future influential researchers in big scholarly network

汇报人：顾清鉴 5150309103328

# contents

# 01

# INTRODUCTION

# PURPOSE

predict academic influential researchers in the future by using the weighted model of factors such as author, publication venue, title, abstract, and publication year, etc.

# MAIN CONTENT

representative evaluation: H-index

- drawbacks:

  ① co-authors

  ② Not suitable for comparing different subjects

# MAIN CONTENT

① theoretical method

② machine learning

# 02

# METHODS

# CONTRIBUTION WEIGHT

improvement for H-index

based on the author's signature rank, the position weight:

$$W_i > W_j \ , \ \forall i < j$$

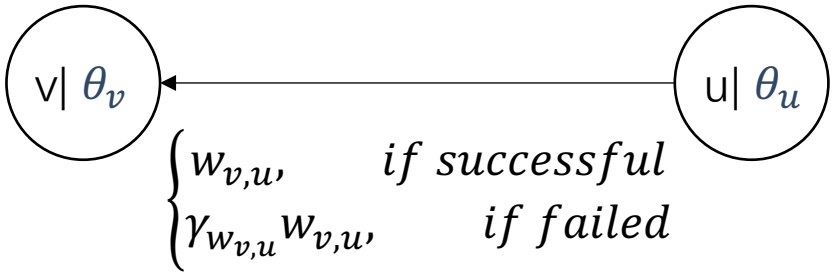
assume there are k co-authors, the contribution weight $W_j = \frac{2(k-j+1)}{k(k+1)}$, then

$$H_w = WH$$

# DIFFUSION MODEL

Each node v has an information acceptance threshold $\theta_v$. Each node v is affected by all its active neighbor nodes. The impact value is represented by weight w.

Assume node u diffuses node v with influence weight $w_{v,u}$. If successful, node v is converted to active node; If failed, consider attenuation coefficients $\gamma$. After several steps, the effect of node u on node v is not then $w_{v,u}$, but $\gamma w_{v,u}$.



# MACHINE LEARNING- extract features

## Non-numeric values

- Title
- Abstract

## Numeric values

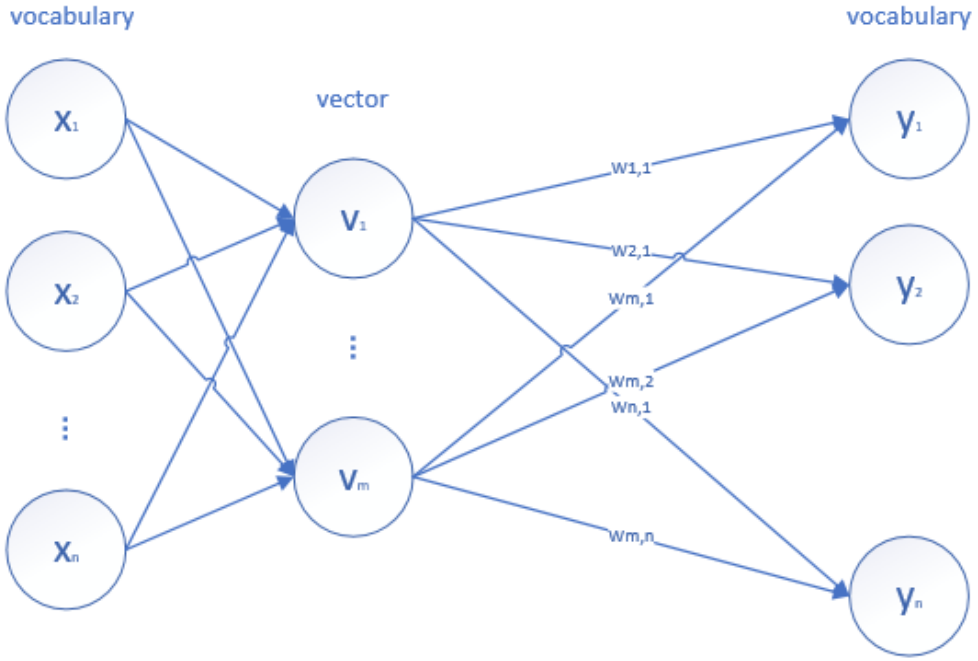
- Author citation
- Publication venue citation
- Publication years

# NON-NUMERIC VALUES

preprocessing: extract features

① non-numeric value: word → vector

- Word Segmentation / Stemming and Morpheme Reduction
- Construct a dictionary, count word frequency



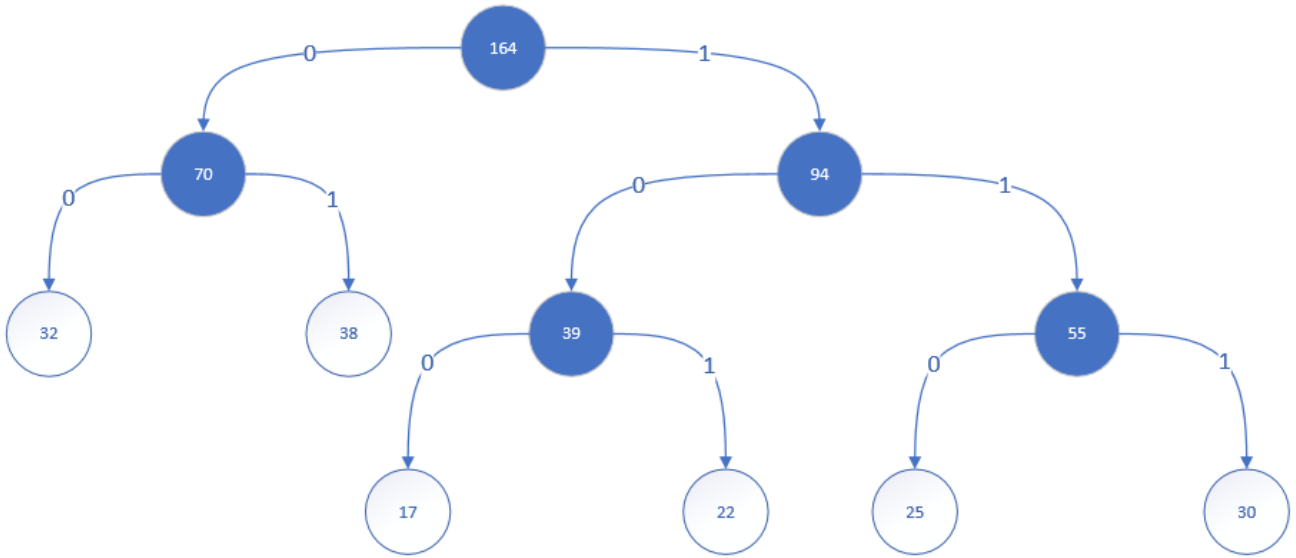
# NON-NUMERIC VALUES

preprocessing: extract features

① non-numeric value: word → vector

- Construct a tree structure

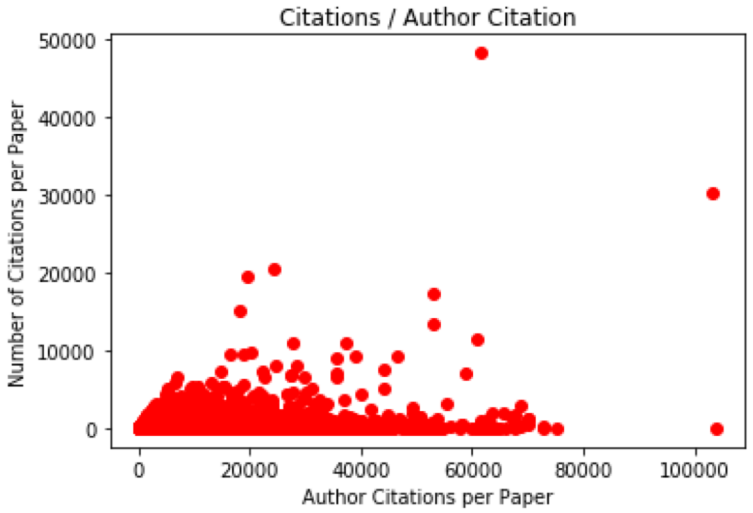
(Huffman tree)

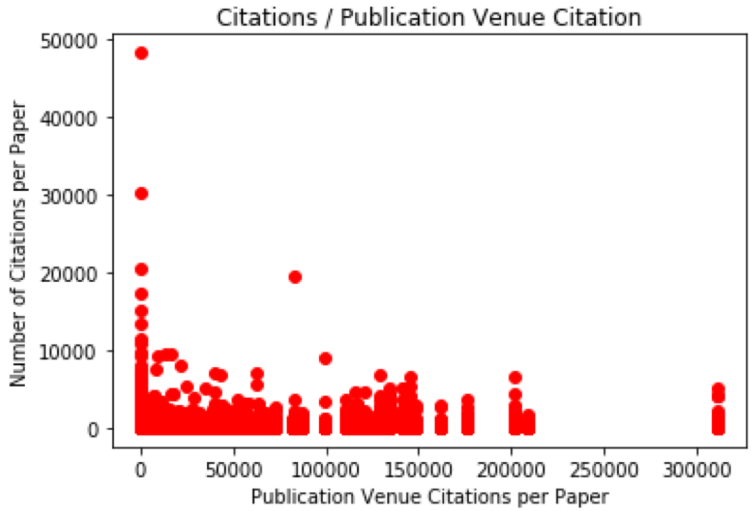


# 03

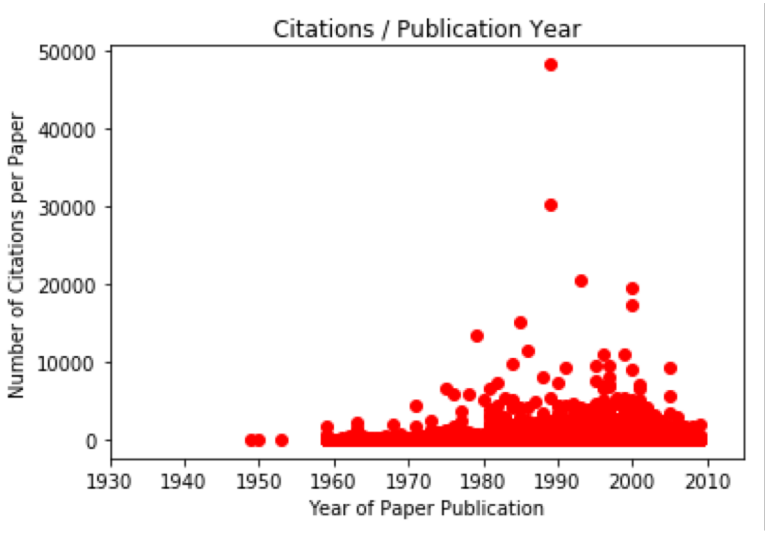
# RESULT & DISCUSSION

# RESULT & DISCUSSION




Dataset: about 500,000 articles



comparison of different algorithm:

Table 1: Table of MMAE & MMSE

| algorithm | minimum mean absolute error | minimum mean square error |
|---|---|---|
| linear regression | 18.96 | 136.80 |
| non-linear regression | 18.48 | 136.32 |
| SVM | 17.74 | 138.42 |
| regression tree | 20.27 | 147.54 |

# Thank you!