

# AI CUP 2019 - 人工智慧論文機器閱讀競賽之論文分類 進度報告1

Created by 黃彥鈞(Weber Huang) on 2019/10/17

Team : TMU-Autobot

AI CUP 2019 Link : [教育部全國大專校院人工智慧競賽\(AI CUP 2019\)-人工智慧論文機器閱讀競賽之論文分類](#)

## Background

如何設計一個系統，能自動閱讀論文摘要後，標註並統整論文裡所涉及的演算法？鑑於當今電腦科學的發展日新月異，演算法的更迭與演進以爆炸式的成長，歸納及統整這些演算法所需的人力將不復以往，而爬梳相關文獻所需的時間也往往讓研究者們深感無力。因此，讓機器自動梳理這些不斷推陳出新的演算法，將會是無可避免的嘗試。即便在人力可負擔的情形下，讓機器自動統整相關演算法，將可以讓研究者騰出時間做更有意義的事。

## Objective

在本系列的競賽中，我們將嘗試以語意分析的技術解決一個令電腦科學研究者頭痛已久的問題：「如何設計一個能自動閱讀論文摘要，標註並統整論文中所發明、使用或用來比較的演算法的系統」。

## Main Mission

從arXiv的電腦科學相關論文摘要，預測出摘要所屬的類別(Theoretical Paper, Engineering Paper, Empirical Paper, Others)。需注意的是摘要可以有多個分類，例如: 摘要可以同時是Theoretical Paper和Engineering Paper。

From the abstract of arXiv computer science related paper, predicting the categories to which the abstract belongs (Theoretical Paper, Engineering Paper, Empirical Paper, Others). It should be noted that the abstract can have multiple categories, for example: The abstract can be both Theoretical Paper and Engineering Paper.

## Data Discription

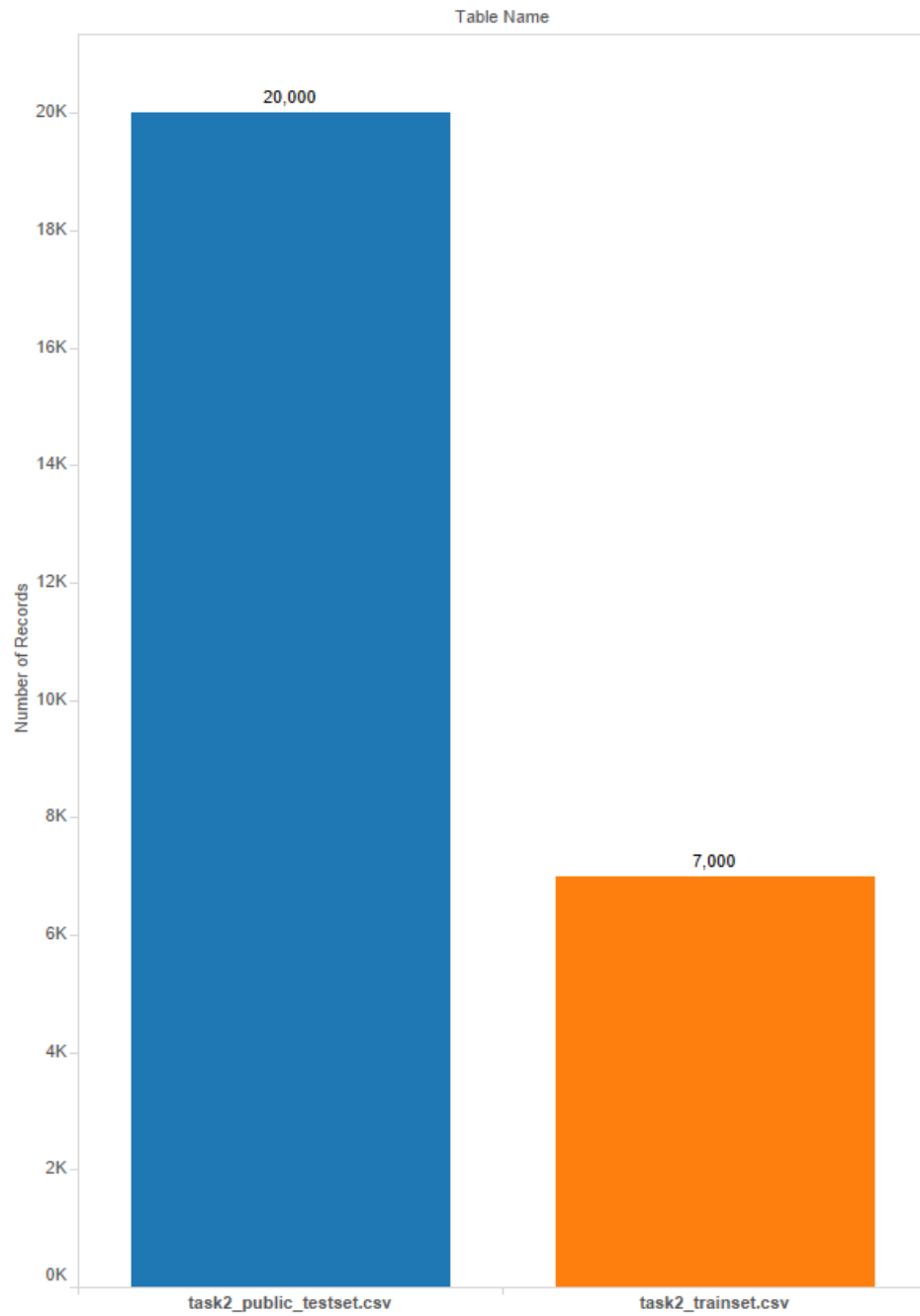
Variables_Name	Variables_Description	Variables_Content
Id	Serial number	D00001
Title	Title of the paper	A Brain-Inspired Trust Management Model to Assure Security in a Cloud based IoT Framework for Neuroscience Applications
Abstract	Abstract of the paper. Each sentence is be seperated by \$\$\$.	<p>Rapid popularity of Internet of Things (IoT) and cloud computing permits neuroscientists to collect multilevel and multichannel brain data to better understand brain functions, diagnose diseases, and devise treatments.</p> <p><i>This paper introduces a Neuro – Fuzzy based Brain – inspired Trust Management Model (TMM) to secure IoT devices and relay nodes, and to ensure data reliability. The proposed TMM utilizes node behavioral trust and data trust estimated using Adaptive Neuro – Fuzzy Inference System and weighted – additive methods respectively to assess the node trustworthiness. In contrast to the existing fuzzy based TMMs, the NS2 simulation results confirm the robustness and accuracy of the proposed TMM in identifying malicious nodes in the communication network.</i></p> <p>With the growing usage of cloud based IoT frameworks in Neuroscience research, integrating the proposed TMM into the existing infrastructure will assure secure and reliable data communication among the E2E devices.</p>
Authors	Authors of the paper. Each name is be seperated by /.	Mahmud/Kaiser/Rahman/Rahman/Shabut/Al-Mamun/Hussain
Categories	Each field is be seperated by /.	cs.CR/cs.AI/q-bio.NC
Created Date	The date which the paper uploaded to <a href="http://www.arxiv.com">www.arxiv.com</a> .	1/11/2018
Task 2 (Training data)	Multilabel of the paper. Each label is be seperated by space.	THEORETICAL

- Authors, Categories, Created Date may have missing data.

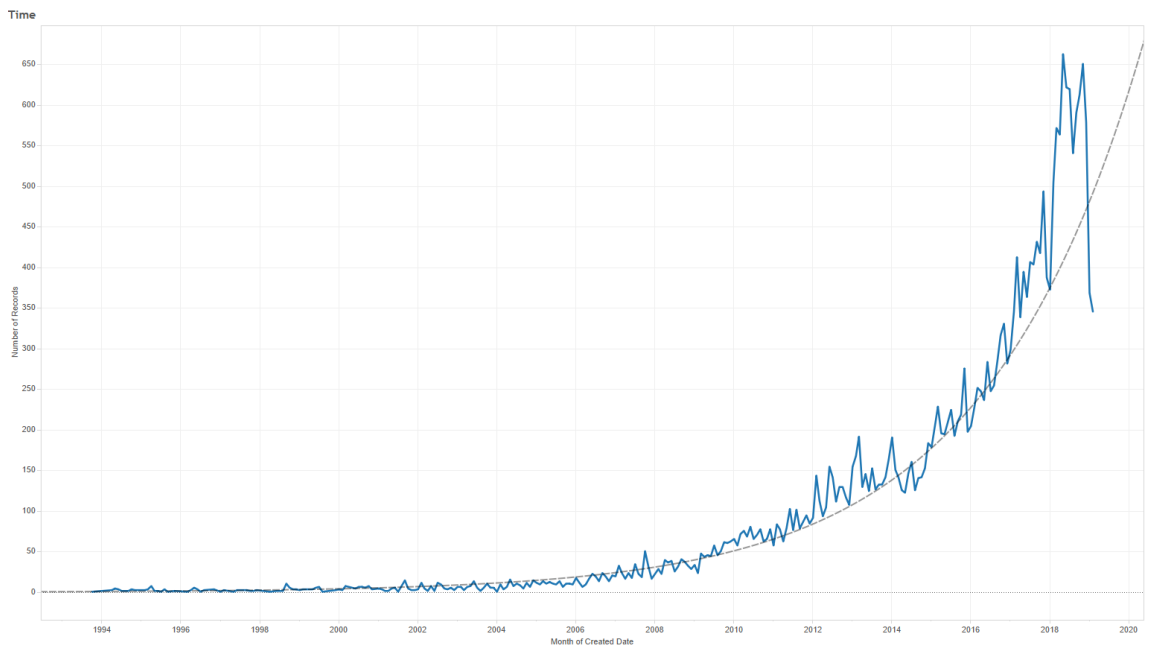
## What we planning to do

- Data preprocessing and Features Engineering Step
  - Visual analytics - Using Python and Tableau

## proportion

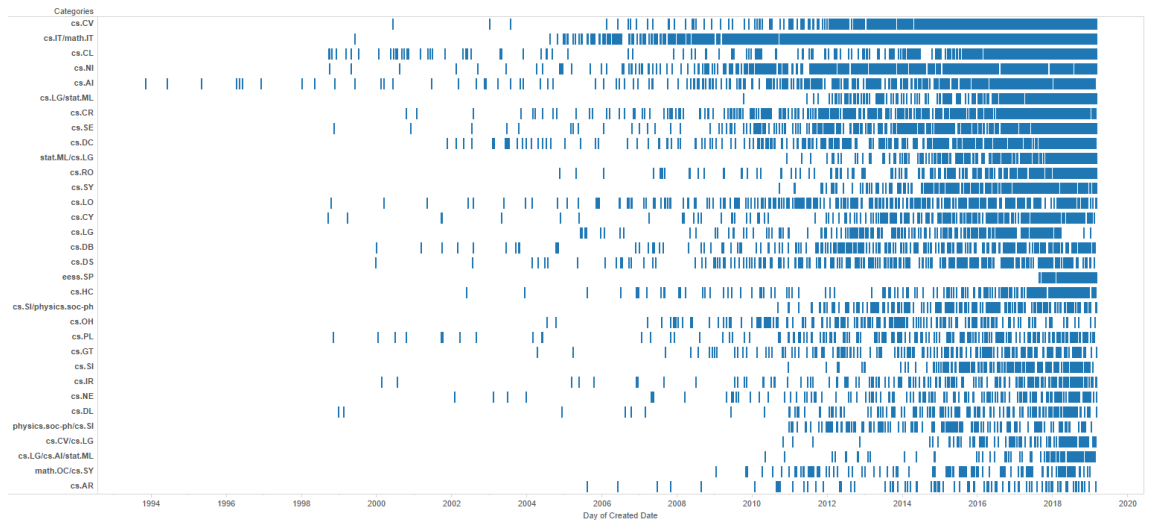


Sum of Number of Records for each Table Name. Color shows details about Table Name. The marks are labeled by sum of Number of Records.



The trend of sum of Number of Records for Created Date Month.

Time\_Categories >100



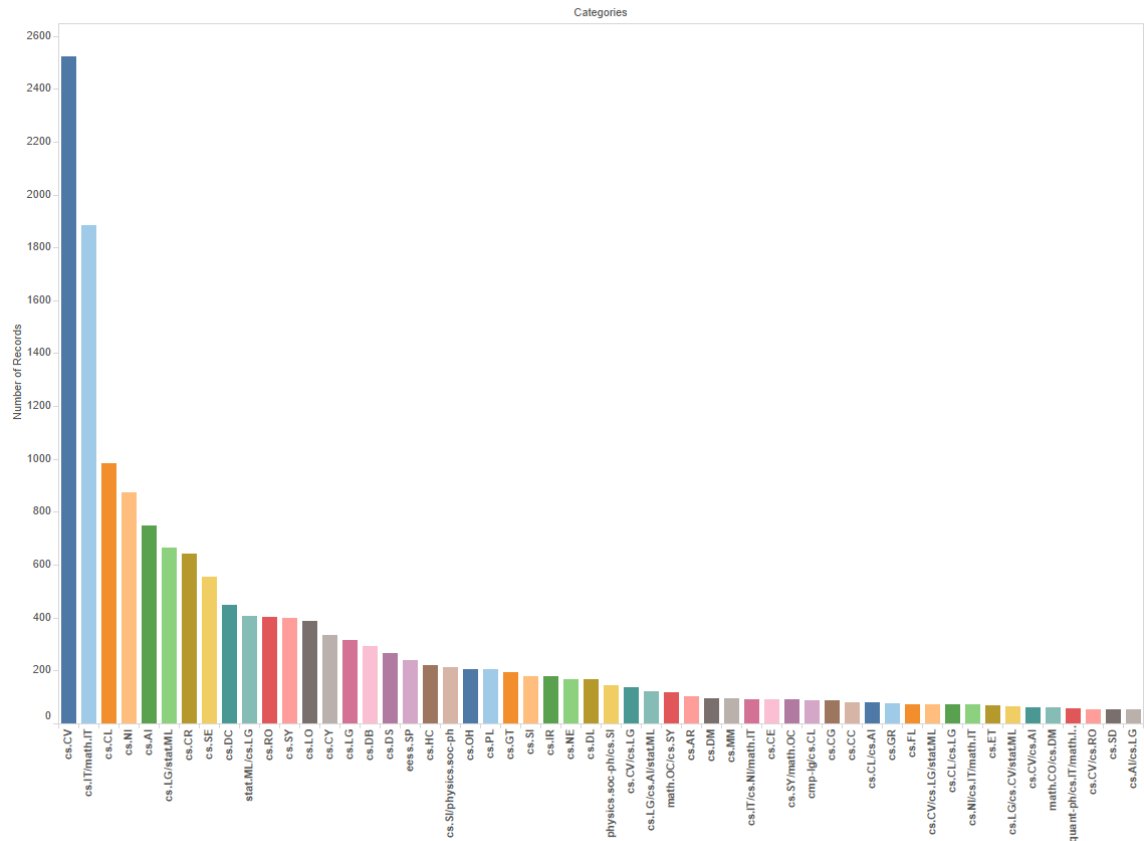
Created Date Day for each Categories. The view is filtered on Categories, which keeps 32 of 4,393 members.

Author



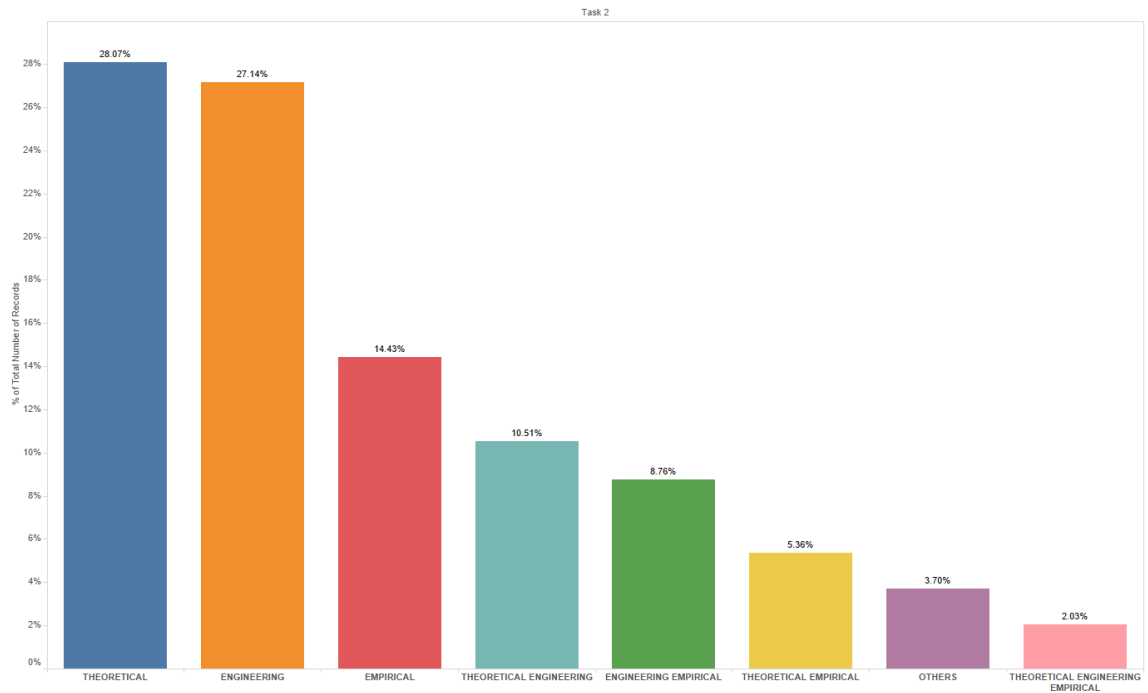
Authors. Color shows details about Authors. Size shows sum of Number of Records. The view is filtered on Authors, which keeps 91 of 25,361 members.

Categories



Sum of Number of Records for each Categories. Color shows details about Categories. The view is filtered on Categories, which keeps 54 of 4,393 members.

labels



% of Total Number of Records for each Task 2. Color shows details about Task 2. The marks are labeled by % of Total Number of Records. The view is filtered on Task 2, which excludes Null.

- Data processing
  - Splitting columns - Authors, Abstract, Categories
  - Find out the difference among categories from created date
  - Processing the missing values
  - Text mining - Abstract, Title
- Features extracting
- Machine Learning Step
  - Trying different algorithms and keeping tuning our model
  - Updating our output model and refreshing the board (maybe per week)
- Personal goal
  - Enhancing the ability of machine learning of Natural Language Processing
  - Learning machine learning skills on Python.