NLP 南京大学 自然语言处理研究组 Natural Language Processing Group

南京大学 NANJING UNIVERSITY

文本摘要简介

黄书剑

- **Automatic summarization** is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content.

- In addition to text, images and videos can also be summarized.
  - Text summarization finds the most informative sentences in a document;
  - image summarization finds the most representative images within an image collection;
  - video summarization extracts the most important frames from the video content.

https://en.wikipedia.org/wiki/Automatic_summarization

By Donald G. McNeil Jr.

Published April 18, 2020
Updated April 19, 2020, 2:09 a.m. ET

1478

The coronavirus is spreading from America's biggest cities to its suburbs, and has begun encroaching on the nation's rural regions. The virus is believed to have infected millions of citizens and has killed more than 34,000.

Yet President Trump this week proposed guidelines for reopening the economy and suggested that a swath of the United States would soon resume something resembling normalcy. For weeks now, the administration's view of the crisis and our future has been rosier than that of its own medical advisers, and of scientists generally.

In truth, it is not clear to anyone where this crisis is leading us. More than 20 experts in public health, medicine, epidemiology and history shared their thoughts on the future during in-depth interviews. When can we emerge from our homes? How long, realistically, before we have a treatment or vaccine? How will we keep the virus at bay?

Some felt that American ingenuity, once fully engaged, might well produce advances to ease the burdens. The path forward depends on factors that are certainly difficult but doable, they said: a carefully staggered approach to reopening, widespread testing and surveillance, a treatment that works, adequate resources for health care providers — and eventually an effective vaccine.

The Coronavirus in America: The Year Ahead

There will be no quick return to our previous lives, according to nearly two dozen experts. But there is hope for managing the scourge now and in the long term.

By DONALD G. MCNEIL JR.

https://www.nytimes.com/2020/04/18/health/coronavirus-america-future.html

Automatic Text Summarizer

Best Online Summarizing Tool

Yet President Trump this week proposed guidelines for reopening the economy and suggested that a swath of the United States would soon resume something resembling normalcy. For weeks now, the administration's view of the crisis and our future has been rosier than that of its own medical advisers, and of scientists generally.

In truth, it is not clear to anyone where this crisis is leading us. More than 20 experts in public health, medicine, epidemiology and history shared their thoughts on the future during in-depth interviews. When can we emerge from our homes? How long, realistically, before we have a treatment or vaccine? How will

Clear Summarize

For weeks now, the administrations view of the crisis and our future has been rosier than that of its own medical advisers, and of scientists generally.

More than 20 experts in public health, medicine, epidemiology and history shared their thoughts on the future during in-depth interviews.

https://autosummarizer.com/

Text Compactor

Free Online Automatic Text Summarization Tool

Home | About

Follow these simple steps to create a summary of your text.

Step 1

Type or paste your text into the box.

The coronavirus is spreading from America's biggest cities to its suburbs, and has begun encroaching on the nation's rural regions. The virus is believed to have infected millions of citizens and has killed more than 34,000.

Yet President Trump this week proposed guidelines for reopening the economy and suggested that a swath of the United States would soon resume something resembling normalcy. For weeks now, the administration's view of the crisis and our future has been rosier than that of its own medical advisers, and of scientists generally.

In truth, it is not clear to anyone where this crisis is leading us. More than 20 experts in public health, medicine, epidemiology and history shared their thoughts on the future during in-depth interviews. When can we emerge from our homes? How long, realistically, before we have a treatment or vaccine? How will we keep the virus at bay?

Some felt that American ingenuity, once fully engaged, might well produce advances to ease the burdens. The path forward depends on factors that are certainly difficult but doable, they said: a carefully staggered approach to reopening, widespread testing and surveillance, a treatment that works, adequate resources for health care providers — and eventually an effective vaccine.

Step 2

Drag the slider, or enter a number in the box, to set the percentage of text to keep in the summary.

15 %

Step 3

Read your summarized text. If you would like a different summary, repeat Step 2. When you are happy with the summary, copy and paste the text into a word processor, or text to speech program, or language translation tool

The path forward depends on factors that are certainly difficult but doable, they said: a carefully staggered approach to reopening, widespread testing and surveillance, a treatment that works, adequate resources for health care providers — and eventually an effective vaccine.

https://www.textcompactor.com/

This is a 7 sentence summary of https://www.nytimes.com/2020/04/18/health...

"My optimistic side says the virus will ease off in the summer and a vaccine will arrive like the cavalry," said Dr. William Schaffner, a preventive medicine specialist at Vanderbilt University medical school.

Even though limited human trials of three candidates - two here and one in China - have already begun, Dr. Fauci has repeatedly said that any effort to make a vaccine will take at least a year to 18 months.

Dr. Paul Offit, a vaccinologist at the Children's Hospital of Philadelphia, noted that the record is four years, for the mumps vaccine.

A new vaccine is usually first tested in fewer than 100 young, healthy volunteers.

As arduous as testing a vaccine is, producing hundreds of millions of doses is even tougher, experts said.

Most American vaccine plants produce only about 5 million to 10 million doses a year, needed largely by the 4 million babies born and 4 million people who reach age 65 annually, said Dr. R. Gordon Douglas Jr., a former president of Merck's vaccine division.

Flu vaccine plants are large, but those that grow the vaccines in chicken eggs are not suitable for modern vaccines, which grow in cell broths, he said.

Reduced By: 56 % **Characters:** 56

SETTINGS **NEW SUMMARY**

SUMMARIZE | ABOUT | API | PARTNER | BOOKMARK WIDGET | CONTACT

REGISTER | **LOGIN**

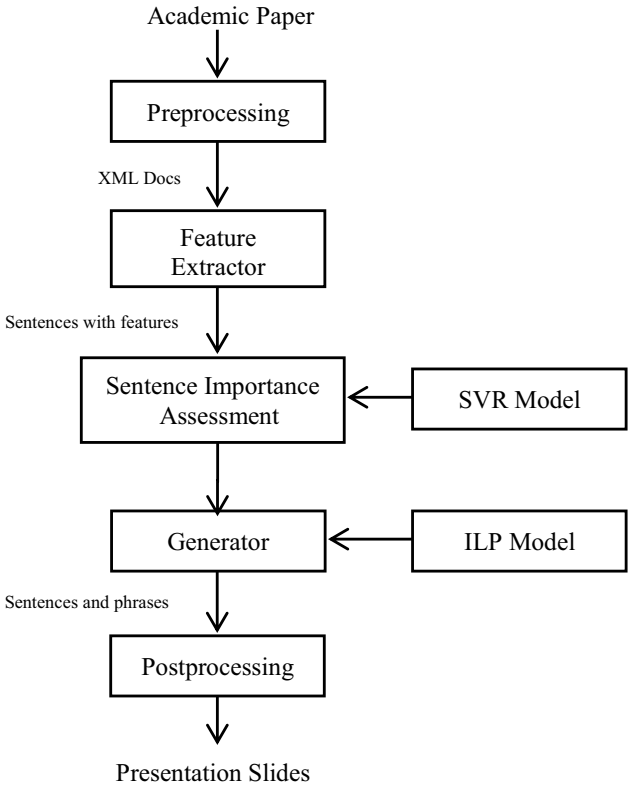
Most American vaccine plar
the
said
arge
derr
rs:

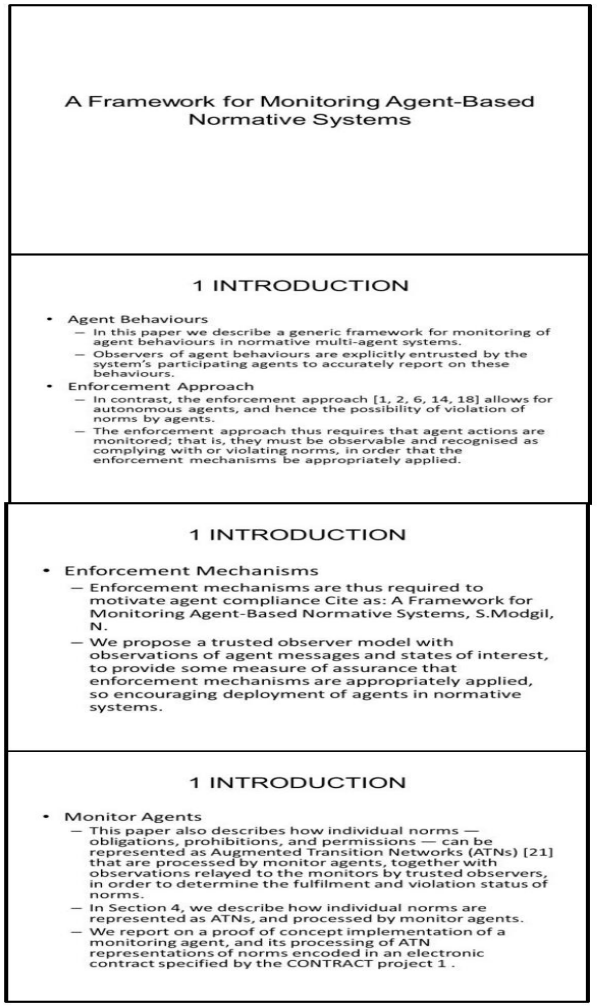
SETTINGS

- [ ] Avoid Questions [?]
- [ ] Avoid Exclamations [?]
- [ ] Avoid Quotations [?]
- [ ] Heat Map [?]
- [ ] Quick Transitions [?]
- [ ] Specify Topic [?]
- [ ] Remember Settings [?]

CLOSE

https://smmry.com/





A Framework for Monitoring Agent-Based Normative Systems

1 INTRODUCTION

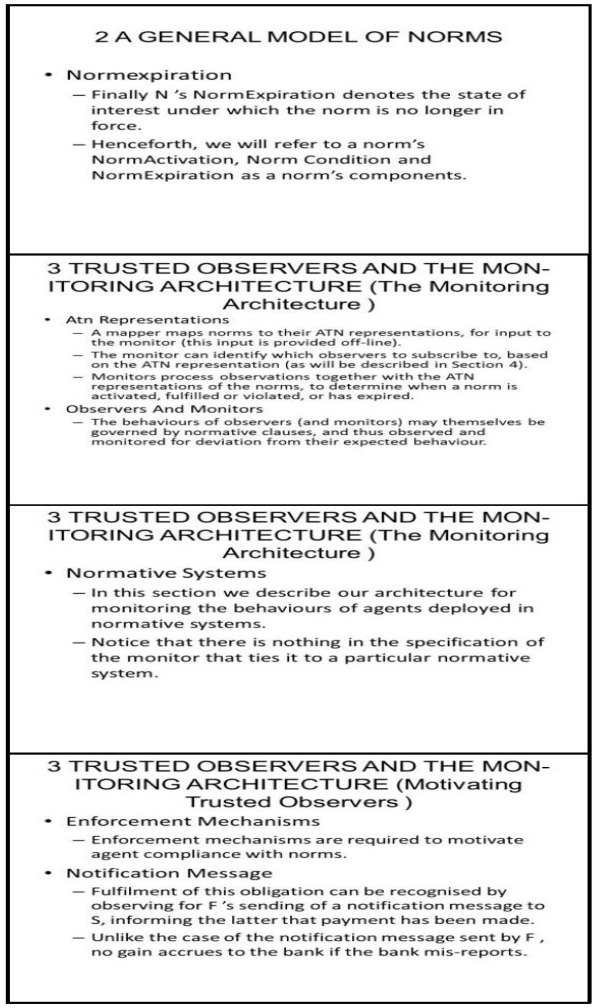
- Agent Behaviours
  - In this paper we describe a generic framework for monitoring of agent behaviours in normative multi-agent systems.
  - Observers of agent behaviours are explicitly entrusted by the system's participating agents to accurately report on these behaviours.
- Enforcement Approach
  - In contrast, the enforcement approach [1, 2, 6, 14, 18] allows for autonomous agents, and hence the possibility of violation of norms by agents.
  - The enforcement approach thus requires that agent actions are monitored; that is, they must be observable and recognised as complying with or violating norms, in order that the enforcement mechanisms be appropriately applied.

1 INTRODUCTION

- Enforcement Mechanisms
  - Enforcement mechanisms are thus required to motivate agent compliance Cite as: A Framework for Monitoring Agent-Based Normative Systems, S.Modgil, N.
  - We propose a trusted observer model with observations of agent messages and states of interest, to provide some measure of assurance that enforcement mechanisms are appropriately applied, so encouraging deployment of agents in normative systems.

1 INTRODUCTION

- Monitor Agents
  - This paper also describes how individual norms — obligations, prohibitions, and permissions — can be represented as Augmented Transition Networks (ATNs) [21] that are processed by monitor agents, together with observations relayed to the monitors by trusted observers, in order to determine the fulfilment and violation status of norms.
  - In Section 4, we describe how individual norms are represented as ATNs, and processed by monitor agents.
  - We report on a proof of concept implementation of a monitoring agent, and its processing of ATN representations of norms encoded in an electronic contract specified by the CONTRACT project 1 .



2 A GENERAL MODEL OF NORMS

- Normexpiration
  - Finally N 's NormExpiration denotes the state of interest under which the norm is no longer in force.
  - Henceforth, we will refer to a norm's NormActivation, Norm Condition and NormExpiration as a norm's components.

3 TRUSTED OBSERVERS AND THE MONITORING ARCHITECTURE (The Monitoring Architecture)

- Atn Representations
  - A mapper maps norms to their ATN representations, for input to the monitor (this input is provided off-line).
  - The monitor can identify which observers to subscribe to, based on the ATN representation (as will be described in Section 4).
  - Monitors process observations together with the ATN representations of the norms, to determine when a norm is activated, fulfilled or violated, or has expired.
- Observers And Monitors
  - The behaviours of observers (and monitors) may themselves be governed by normative clauses, and thus observed and monitored for deviation from their expected behaviour.

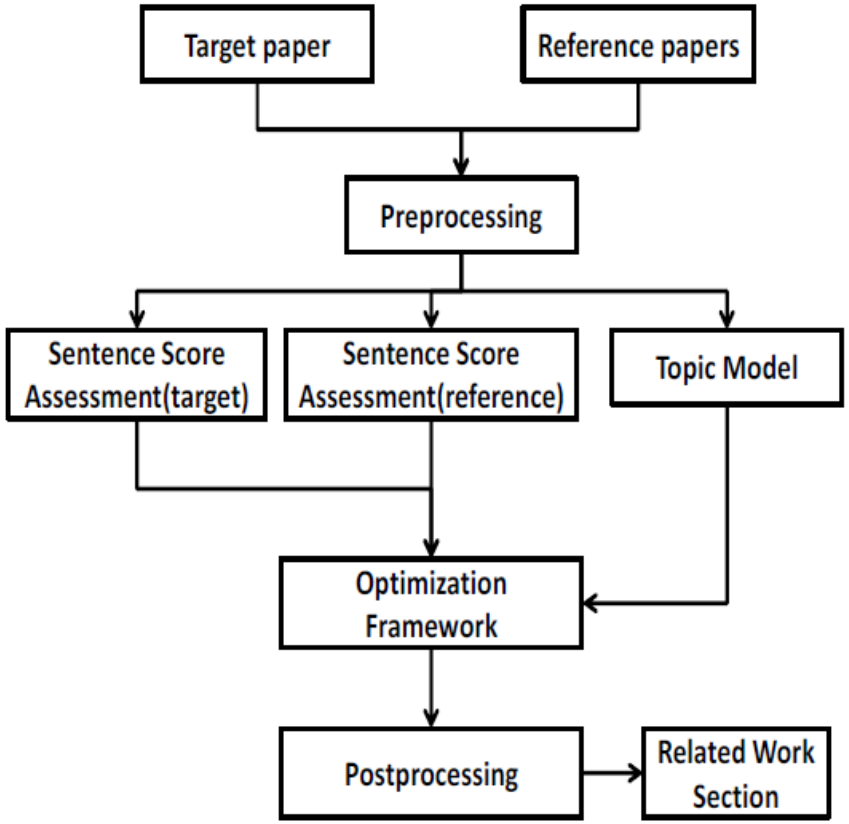
3 TRUSTED OBSERVERS AND THE MONITORING ARCHITECTURE (The Monitoring Architecture)

- Normative Systems
  - In this section we describe our architecture for monitoring the behaviours of agents deployed in normative systems.
  - Notice that there is nothing in the specification of the monitor that ties it to a particular normative system.

3 TRUSTED OBSERVERS AND THE MONITORING ARCHITECTURE (Motivating Trusted Observers)

- Enforcement Mechanisms
  - Enforcement mechanisms are required to motivate agent compliance with norms.
- Notification Message
  - Fulfilment of this obligation can be recognised by observing for F 's sending of a notification message to S, informing the latter that payment has been made.
  - Unlike the case of the notification message sent by F , no gain accrues to the bank if the bank mis-reports.

PPSGen: Learning-Based Presentation Slides Generation for Academic Papers. Hu and Wan. IJCAI2015



There has been a substantial amount of research on automatic taxonomy induction. As we mentioned earlier, two main approaches are pattern-based and clustering-based.

Pattern-based approaches are the main trend for automatic taxonomy induction. ...
Pattern-based approaches started from and still pay a great deal of attention to the most common is-a relations. ...

Clustering-based approaches usually represent word contexts as vectors and cluster words based on similarities of the vectors (Brown et al., 1992; Lin, 1998). ...

Many clustering-based approaches face the challenge of appropriately labeling non-leaf clusters. ... In this paper, we take an incremental clustering approach,... The advantage of the incremental approach is that it eliminates the trouble of inventing cluster labels and concentrates on placing terms in the correct positions in a taxonomy hierarchy.
The work by Snow et al. (2006) is the most similar to ours ... Moreover, our approach employs heterogeneous features from a wide range; while their approach only used syntactic dependency.

Two different topics

Comparison with the author's work

Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach. Hu and Wan. EMNLP 2014

数据集

- DUC
  - Document Understanding Conferences, 2001-2007
  - https://www-nlpir.nist.gov/projects/duc/data.html
- TAC
  - Text Analysis Conference, 2008-2011
  - https://tac.nist.gov/data/
- CNN/DailyMail (2015)
  - Original for QA https://cs.nyu.edu/~kcho/DMQA/
- LCSTS (2015)
  - A Large-Scale Chinese Short Text Summarization Dataset
  - http://icrc.hitsz.edu.cn/Article/show/139.html
- Gigaword (documents with headline)

- DUC
  - Documents
  - Summaries, results, etc.
    - manually created summaries
    - automatically created baseline summaries
    - submitted summaries created by the participating groups' systems
    - tables with the evaluation results
    - additional supporting data and software

https://www-nlpir.nist.gov/projects/duc/data.html

- LCSTS

【江西高考被曝替考 有关考生已被警方控制】人民日报记者吴齐强消息，江西高考被曝光替考，7日中午江西省教育厅发布消息称，接到有人组织替考的举报后，江西省教育厅、江西省教育考试院立即部署南昌市教育考试院，联合南昌市警方开展调查核实，有关考生已被警方控制。调查进展情况将及时向社会公布。

Short Text: 水利部水资源司司长陈明忠今日在新闻发布会上透露，根据刚刚完成的水资源管理制度的考核，有部分省接近了红线的指标，有部分省超过红线的指标。在一些超过红线的地方，将对一些取用水项目进行区域的限批，严格地进行水资源论证和取水许可的批准。

Mingzhong Chen, the Chief Secretary of the Water Devision of the Ministry of Water Resources, revealed today at a press conference, according to the just-completed assessment of water resources management system, some provinces are closed to the red line indicator, some provinces are over the red line indicator. In some places over the red line, It will enforce regional approval restrictions on some water projects, implement strictly water resources assessment and the approval of water licensing.

Summarization: 部分省超过年度用水红线指标 取水项目将被限批

Some provinces exceeds the red line indicator of annual water using, some water project will be. limited approved

Human Score: 5

Part I	2,400,591	
Part II	Number of Pairs	10,666
	Human Score 1	942
	Human Score 2	1,039
	Human Score 3	2,019
	Human Score 4	3,128
	Human Score 5	3,538
Part III	Number of Pairs	1,106
	Human Score 1	165
	Human Score 2	216
	Human Score 3	227
	Human Score 4	301
	Human Score 5	197

http://icrc.hitsz.edu.cn/Article/show/139.html

评价方法

- ROUGE (Lin 2004, Lin and Och, 2004)
  - Recall-Oriented Understudy for Gisting Evaluation
  - ROUGE-N: Overlap of N-grams between the system and reference summaries (ROUGE-1 ROUGE-2)

$$\text{recall} = \frac{N_{hit}}{N_{Hs}} \qquad \text{precision} = \frac{N_{hit}}{N_{As}}$$

  - ROUGE-L, ROUGE-W, ROUGE-S , ROUGE-SU
    - Longest Common Subsequence , weighted , skip-ngram , unigram , etc.
  - ROUGE 2.0 (Kavita Ganesan 2018)
    - Synonyms , Topic , etc.

*BiLingual Evaluation Understudy BLEU Papineni et al. 2002

主要方法

- 抽取式摘要 Extractive Summarization

  - 从文章中抽取出部分句子组成一个句子的集合
  - 优势：
    - 相对简单
    - 句子本身正确、流畅
    - 与文档内容上有关联
- 生成式摘要 Abstractive Summarization

  - 根据文章内容生成新的句子的集合
  - 优势：
    - 更加灵活
    - 具备抽象和概括能力
    - 适用于特殊场景（对话、小说）

扩展研究方向

- **单文档摘要 v.s. 多文档摘要**
- **跨语言摘要**
- **异质文本摘要**
  - 新闻、社交媒体、学术文献、知识图谱等
- **特殊摘要需求**
  - 增量式、演化式
  - 观点摘要 v.s. 比较式摘要
- **多模态摘要**
  - 对图像、视频等进行摘要
  - 结合图像、视频等进行摘要

参考文献

- Lin, Chin-Yew. 2004. ROUGE: a Package for Automatic Evaluation of Summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain, July 25 - 26, 2004.
- Lin, Chin-Yew and E.H. Hovy 2003. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In Proceedings of 2003 Language Technology Conference (HLT-NAACL 2003), Edmonton, Canada, May 27 - June 1, 2003.
- Lin, Chin-Yew and Franz Josef Och. 2004. Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), Barcelona, Spain, July 21 - 26, 2004.
- Kavita Ganesan. 2018. ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks. https://arxiv.org/abs/1803.01937