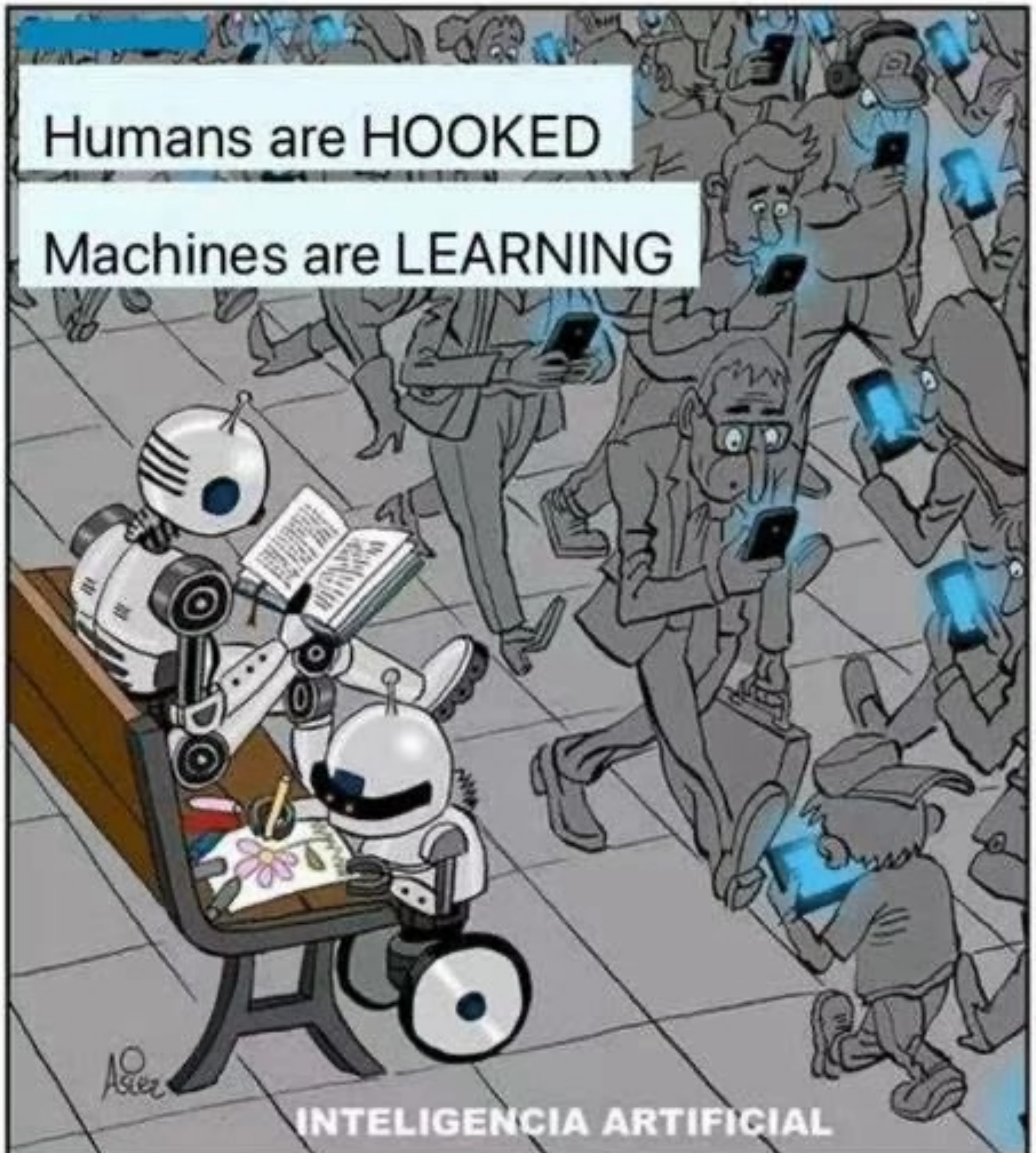
南京大學
NANJING UNIVERSITY

信息抽取
Information Extraction

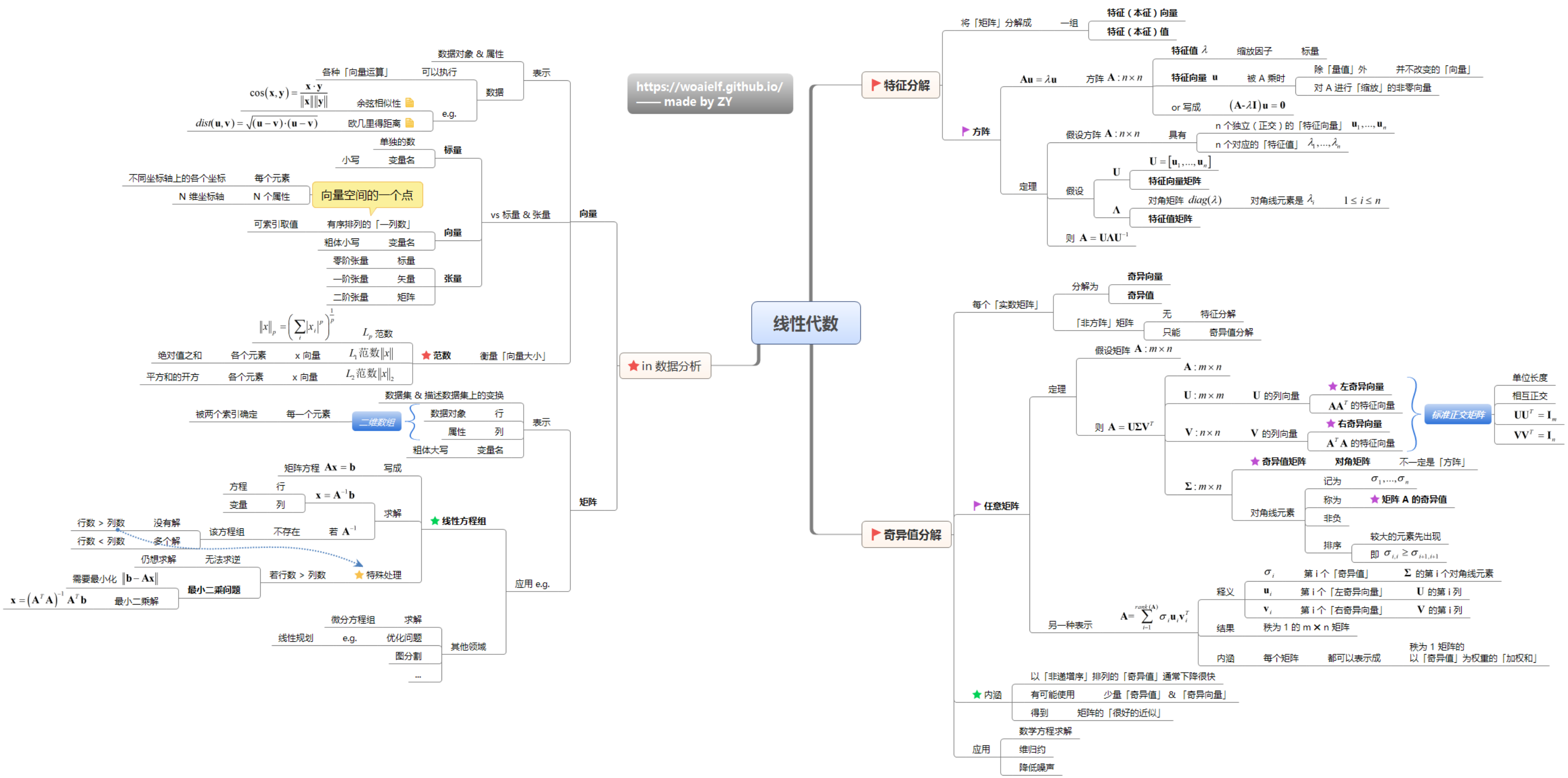
戴新宇

2020-12



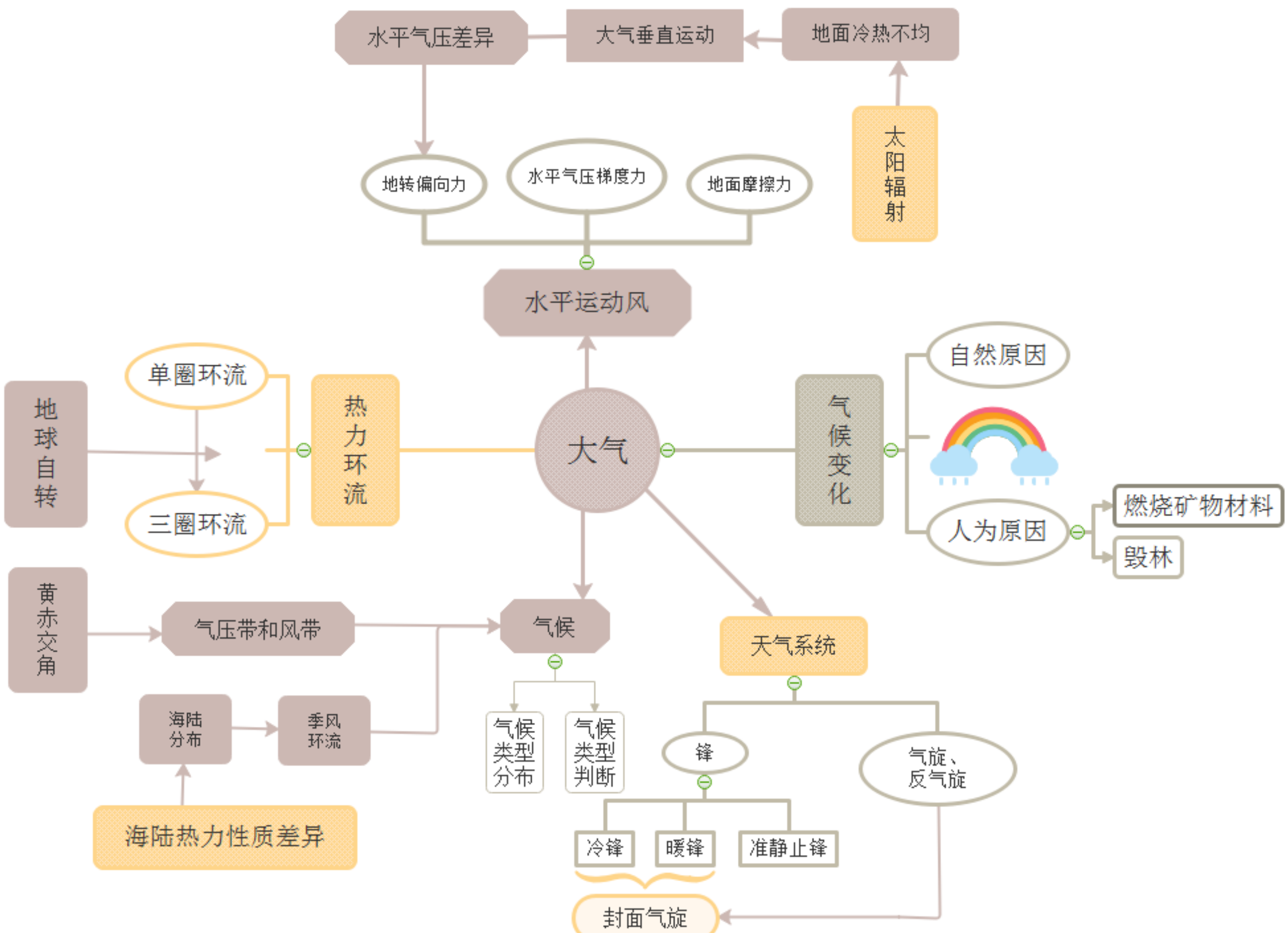
We are hooked

Cheatsheet



https://woaielf.github.io/
—— made by ZY
线性代数
in 数据分析
向量
表示
数据
数据对象 & 属性
各种「向量运算」 可以执行
e.g.
余弦相似性
欧几里得距离
$\cos(\mathbf{x},\mathbf{y})=\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$
$dist(\mathbf{u},\mathbf{v})=\sqrt{(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})}$
vs 标量 & 张量
标量
单独的数
小写 变量名
向量
向量空间的一个点
不同坐标轴上的各个坐标 每个元素
N 维坐标轴 N 个属性
可索引取值 有序排列的「一列数」
粗体小写 变量名
张量
零阶张量 标量
一阶张量 矢量
二阶张量 矩阵
范数
衡量「向量大小」
$\|x\|_p=\left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$
L_p 范数
绝对值之和 各个元素 x 向量 L_1 范数 $\|x\|$
平方和的开方 各个元素 x 向量 L_2 范数 $\|x\|_2$
矩阵
表示
数据集 & 描述数据集上的变换
被两个索引确定 每一个元素 二维数组
数据对象 行
属性 列
粗体大写 变量名
线性方程组
写成 矩阵方程 $\mathbf{Ax}=\mathbf{b}$
方程 行
变量 列
求解 $\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$
若 $\mathbf{A}^{-1}$ 不存在 该方程组
行数 > 列数 没有解
行数 < 列数 多个解
特殊处理
若行数 > 列数 仍想求解 无法求逆
最小二乘问题 需要最小化 $\|\mathbf{b}-\mathbf{Ax}\|$
最小二乘解 $\mathbf{x}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$
应用 e.g.
其他领域
微分方程组 求解
线性规划 e.g. 优化问题
图分割
...
特征分解
将「矩阵」分解成 一组
特征（本征）向量
特征（本征）值
方阵
$\mathbf{Au}=\lambda\mathbf{u}$ 方阵 $\mathbf{A}:n\times n$
特征值 λ 缩放因子 标量
特征向量 $\mathbf{u}$ 被 A 乘时
除「量值」外 并不改变的「向量」
对 A 进行「缩放」的非零向量
or 写成 $(\mathbf{A}-\lambda\mathbf{I})\mathbf{u}=\mathbf{0}$
定理
假设方阵 $\mathbf{A}:n\times n$ 具有
n 个独立（正交）的「特征向量」 $\mathbf{u}_1,\dots,\mathbf{u}_n$
n 个对应的「特征值」 $\lambda_1,\dots,\lambda_n$
假设
U $\mathbf{U}=[\mathbf{u}_1,\dots,\mathbf{u}_n]$
特征向量矩阵
Λ 对角矩阵 $diag(\lambda)$ 对角元素是 λ_i $1\le i\le n$
特征值矩阵
则 $\mathbf{A}=\mathbf{U\Lambda U}^{-1}$
奇异值分解
每个「实数矩阵」
分解为
奇异向量
奇异值
「非方阵」矩阵
无 特征分解
只能 奇异值分解
任意矩阵
定理
假设矩阵 $\mathbf{A}:m\times n$
则 $\mathbf{A}=\mathbf{U\Sigma V}^T$
$\mathbf{A}:m\times n$
$\mathbf{U}:m\times m$ U 的列向量
左奇异向量
$\mathbf{AA}^T$ 的特征向量
$\mathbf{V}:n\times n$ V 的列向量
右奇异向量
$\mathbf{A}^T\mathbf{A}$ 的特征向量
标准正交矩阵
单位长度
相互正交
$\mathbf{UU}^T=\mathbf{I}_m$
$\mathbf{VV}^T=\mathbf{I}_n$
$\mathbf{\Sigma}:m\times n$
奇异值矩阵 对角矩阵 不一定是「方阵」
对角线元素
记为 $\sigma_1,\dots,\sigma_n$
称为 矩阵 A 的奇异值
非负
排序 较大的元素先出现
即 $\sigma_{i,i}\ge\sigma_{i+1,i+1}$
另一种表示
$\mathbf{A}=\sum_{i=1}^{rank(\mathbf{A})}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$
释义
σ_i 第 i 个「奇异值」 Σ 的第 i 个对角线元素
$\mathbf{u}_i$ 第 i 个「左奇异向量」 U 的第 i 列
$\mathbf{v}_i$ 第 i 个「右奇异向量」 V 的第 i 列
结果 秩为 1 的 m × n 矩阵
内涵 每个矩阵 都可以表示成 秩为 1 矩阵的 以「奇异值」为权重的「加权和」
内涵
以「非递增序」排列的「奇异值」 通常下降很快
有可能使用 少量「奇异值」 & 「奇异向量」
得到 矩阵的「很好的近似」
应用
数学方程求解
维归约
降低噪声

思维导图


水平气压差异
大气垂直运动
地面冷热不均
太阳辐射
地转偏向力
水平气压梯度力
地面摩擦力
水平运动风
单圈环流
三圈环流
地球自转
热力环流
大气
气候变化
自然原因
人为原因
燃烧矿物材料
毁林
黄赤交角
气压带和风带
气候
海陆分布
季风环流
海陆热力性质差异
气候类型分布
气候类型判断
天气系统
锋
气旋、反气旋
冷锋
暖锋
准静止锋
封面气旋

信息抽取

- 信息抽取（Information Extraction，简称IE，又译信息截取技术）主要是从大量文字资料中自动抽取特定消息（Particular Information），以作为数据库访问（Database Access）之用的技术。
- 非结构化文本 → 结构化数据 → 知识库（知识

海量文本

唐朝中央的三省中书、门下和尚书，分别负责决策、审议和执行。三省的长官都是宰相，相权分散。

→

结构化知识

ISA(中书省，唐代中央机构)
ISA(门下省，唐代中央机构)
ISA(尚书省，唐代中央机构)
职能(中书省，决策)
职能(门下省，审议)
职能(尚书省，执行)
长官(中书省，宰相)

Why is IE useful?

Christopher Manning

The Full Task of Information Extraction

As a family of techniques:

Information Extraction =
segmentation + classification + association + clustering

For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Now Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“

Richard Stallman, founder of the Free Software Foundation, countered saying…

Microsoft Corporation
CEO
Bill Gates

Gates
Microsoft

Bill Veghte
Microsoft
VP

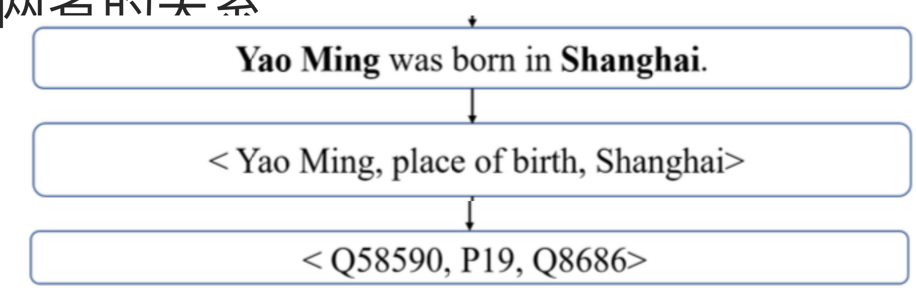
Richard Stallman
founder
Free Software Foundation

NAME	TITLE	ORGANIZATION
Bill Gates	CEO	Microsoft
Bill Veghte	VP	Microsoft
Richard Stallman	founder	Free Soft..

Slide by Andrew McCallum. Used with permission.

Pipeline

- 实体识别
  - 人名、地名、机构名…
  - 药物名、蛋白质名、基因名…
  - 领域专有名词
- 关系抽取
  - 给定头实体和尾实体，识别两者的关系



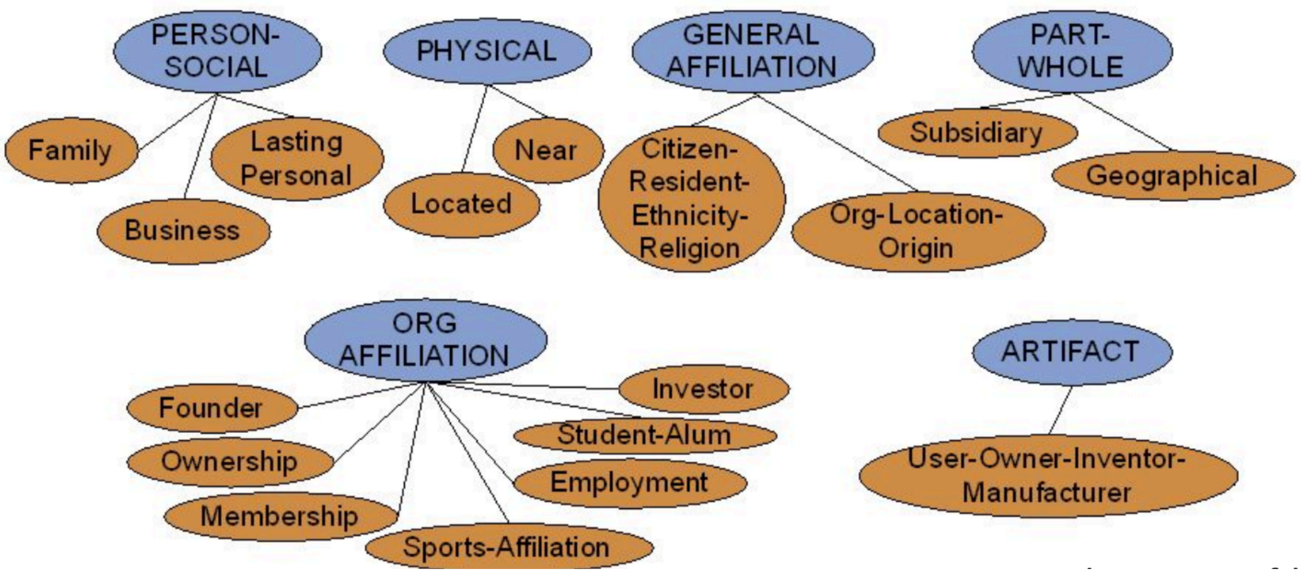
- 知识库填充
  - 实体消歧，实体链接等
- 事件抽取
  - *Who did what to whom when?*
  - …..

实体识别

- 规则方法
- 统计学习方法：
  - 任务描述
  - 任务形式化
  - 模型
  - 特征
  - 评价
  - ……

封闭域关系抽取

- 有限的实体类型，有限的关系类型
  - ACE，17种关系
  - UMLS：134种实体类型，54种关系类型



Doppler echocardiography can be used to diagnose left anterior descending artery stenosis in patients with type 2 diabetes

↓

Echocardiography, Doppler DIAGNOSES Acquired stenosis

封闭域关系抽取

- 有监督统计学习方法：
  - 任务描述及任务形式化
  - 训练数据
  - 模型
  - 特征
  - 评价
  - ……

开放域关系抽取

- Freebase、DBpedia、Yago
  - 包含数以千计的关系类型，百万千万级的实体
- 样本不平衡性和稀疏性
- 语言表达多样性和有限性

开放域关系抽取

- 基于模版的方法 （Berland and Charniak）

Sentence Fragment	Pattern
... building's basement …	whole NN[-PL] 's POS part NN[-PL]
…basement of building…	parts NN-PL of PREP wholes NN-PL
…basement in building…	parts NN-PL in PREP wholes NN-PL
…basement in the big building…	part NN in PREP {the|a} DET mods [JJ|NN]* whole NN
…basements of a building…	part NN[-PL] of PREP {the|a} DET mods [JJ|NN]* whole NN

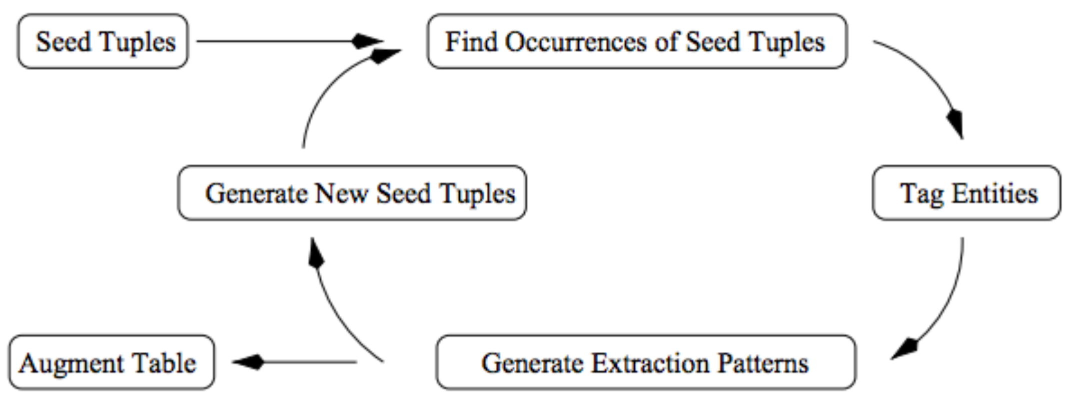
For each pattern:

1. Find occurrences of the pattern
2. Filter those ending with *-ing*, *-ness*, *-ity*
3. Applied a likelihood metric.

First two are reliable patterns.
The rest are noisy in practice. (~ 55% accuracy)

开放域关系抽取

- Snowball: Improved Bootstrapping
  [Agichten and Gravano, 2000]



基于远程监督方法的关系分类

- Task of Relation Classification
- Distant Supervision for Relation Classification
- Challenges: Noisy data
  - Suppress Noise
  - Remove Noise
  - Rectify Noise
- Summary

Task of Relation Classification

- Aim to predict relations of the target entity pair given a plain text.
- Play an important role in downstream tasks such as knowledge base completion, question answering, and so on.

e.g.

Sentence : ***Obama*** was born in Honolulu, Hawaii, ***USA*** as he has always said.

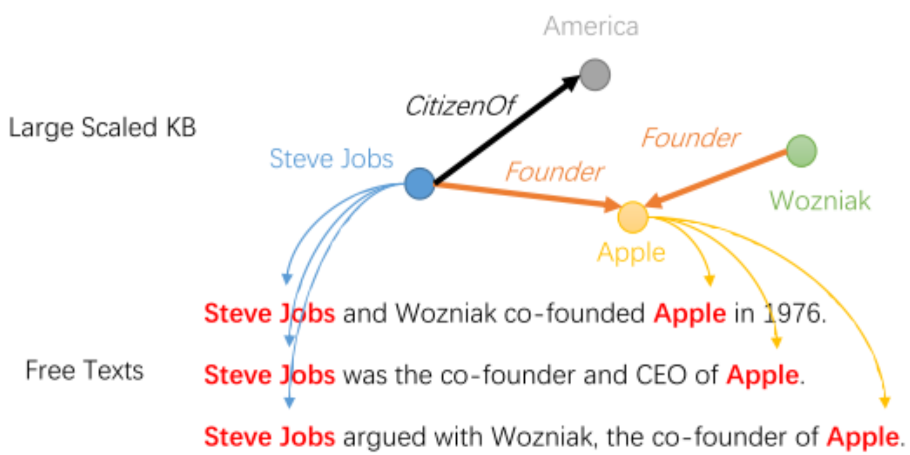
Relation : born-in

Fact triplet : <Obama, born-in, USA>

- Be treated as a traditional supervised learning problem.
- So, it heavily relies on large scale annotated data which is time and labor consuming.

Distant Supervision for Relation Classification

- Goal: Automatically generate a large scale of annotated data.
- Assumption: if two entities have a relation in KB, all sentences containing these two entities will express the relation.



Distant supervision for relation extraction without labeled data. (ACL 2009)

Challenges: Noise

- The assumption is too strong and introduces noisy labeling data.
  - False Positives: not each sentence containing two entities mention the same relation in KB.
  - False Negatives: two entities are mislabeled as no relation (NA) due to the absence of relational fact in KB, even though they express the target relation.

Sentence	**DS Label**	**Gold Label**
S1: ***Steve Jobs*** says ***Apple*** is a curator, nothing more.	Founder	NA
S2: I followed the work of ***Steve Jobs*** even after he was booted from ***Apple*** in 1985.	Founder	EmployedBy
S3: The richest youngster, according to Hurun, is ***Ma Huateng***, the 39-year-old co-founder of Internet giant ***Tencent***.	NA	Founder

Challenges: Noise

- The assumption is too strong and introduces noisy labeling data.
  - False Positives
  - False Negatives
- Solution：
  - Suppress Noise: de-emphasize the false positive sentences.
  - Remove Noise: remove the false positive sentences.
  - Rectify Noise: not only rectify the wrong label of false positive sentences but also false negative ones.

Challenges: Noise

- The assumption is too strong and introduces noisy labeling problem.
  - False Positives
  - False Negatives
- Solution：

Sentence	**DS Label**	**Gold Label**	**restrain**	**Remove**	**Rectify**
S0: Cook reminded the audience that ***Apple*** was founded by ***Steve Jobs***, the son of a Syrian immigrant.	Founder	Founder	0.7	remain	Founder
S1: ***Steve Jobs*** says ***Apple*** is a curator, nothing more.	Founder	NA	0.1	remove	NA
S2: I followed the work of ***Steve Jobs*** even after he was booted from ***Apple*** in 1985.	Founder	EmployedBy	0.2	remove	EmployedBy
S3: The richest youngster, according to Hurun, is ***Ma Huateng***, the 39-year-old co-founder of Internet giant ***Tencent***.	NA	Founder	-	-	Founder

Suppress Noise

- Only focus on false positive

Sentence	DS Label	Gold Label	restrain	Remove	Rectify
S0: Cook reminded the audience that ***Apple*** was founded by ***Steve Jobs***, the son of a Syrian immigrant.	Founder	Founder	0.7	remain	Founder
S1: ***Steve Jobs*** says ***Apple*** is a curator, nothing more.	Founder	NA	0.1	remove	NA
S2: I followed the work of ***Steve Jobs*** even after he was booted from ***Apple*** in 1985.	Founder	EmployedBy	0.2	remove	EmployedBy
S3: The richest youngster, according to Hurun, is ***Ma Huateng***, the 39-year-old co-founder of Internet giant ***Tencent***.	NA	Founder	-	-	Founder

Suppress Noise

- [Riedel et al., 2010] proposed the At-Least-One assumption to relax the strong DS assumption and firstly modeled DS relation classification as a multi-instance classification problem.

- At-Least-One assumption: at least one sentence that mentions these two entities will express their relation.

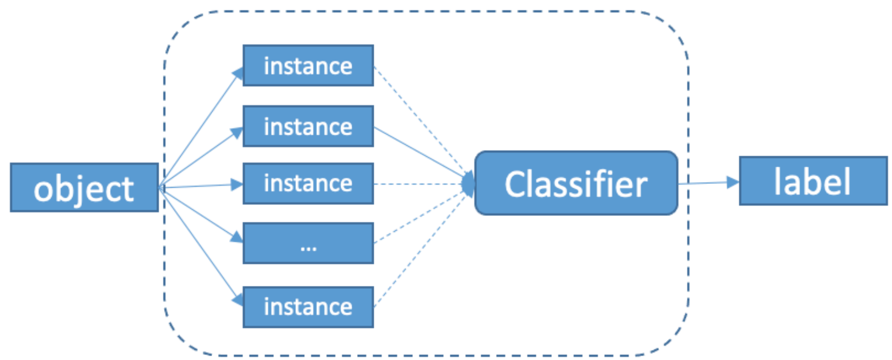
- Multi-instance learning: all sentences mentioning the same entity pair are put into one bag and train model among the bags.

Modeling relations and their mentions without labeled text. (ECML 2010)

Suppressing Noise

- Multi-instance learning
  - A bag is a set of instances.
  - The labels are only assigned to *bags of instances*. No labels for instances.
  - In the binary case, a bag is labeled positive if *at least* one instance in that bag is positive, and the bag is labeled negative if *all* the instances in it are negative.

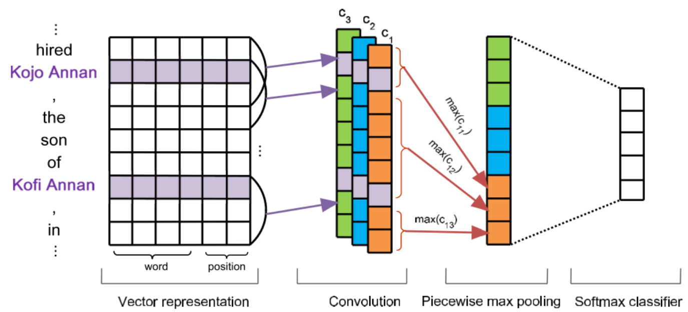
Classic Supervised Learning



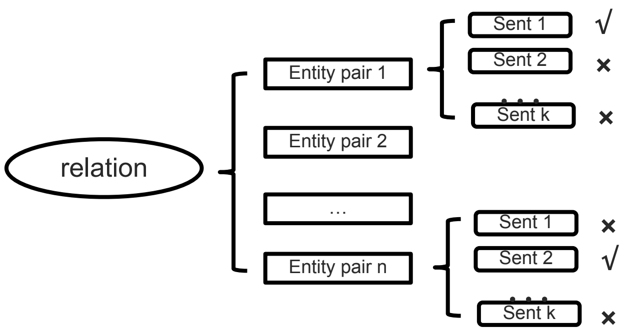
Multi-instance Learning

Suppressing Noise

- PCNN+ONE
  - Sentence encoder: PCNN, which can capture important information between two entities.
  - Select the most reliable sentence (which obtains highest score on target relation) for each entity pair in training and prediction.
  - Neglect other sentences containing rich information.



Sentence Encoder



Distant supervision for relation extraction via piecewise convolutional neural networks. (EMNLP 2015)

Suppressing Noise

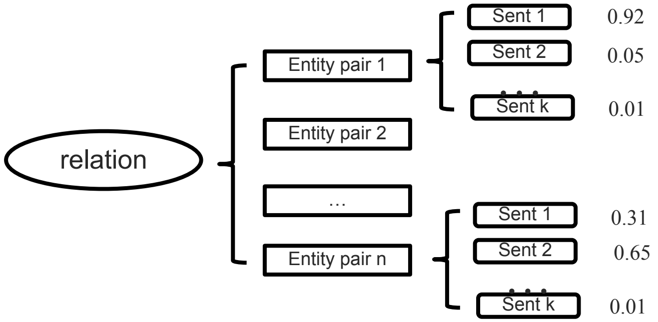
- PCNN+ATT
  - Sentence encoder: PCNN
  - In order to make full use of all informative sentences, this work proposes attention mechanism to select reliable sentences and de-emphasize unreliable ones.
  - The attention weight measure how related with the sentence and target relation.

$$s = \sum_i \alpha_i x_i$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)}$$

$$e_i = x_i A r$$

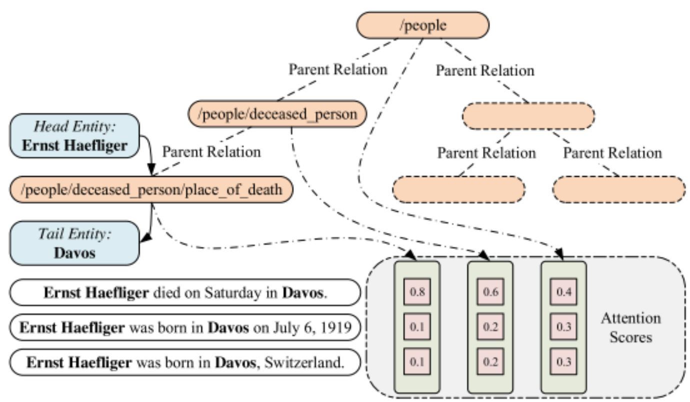
The vector of entity pair $\boldsymbol{s}$ is a weighed sum of these sentence vector $\boldsymbol{x_i}$. The weight $\boldsymbol{\alpha_i}$ is calculated by $\boldsymbol{e_i}$ which scores how well the input sentence $\boldsymbol{x_i}$ and the predict relation $\boldsymbol{s}$ matches.



Neural relation extraction with selective attention over instances. (ACL 2016)

Suppressing Noise

- PCNN+HATT
  - Sentence encoder: PCNN
  - Relation hierarchies provide rich correlated information among relations.
  - Compute scores for each sentence on the each layer of the hierarchies, which provides coarse-to-fine granularity for identifying valid sentences.

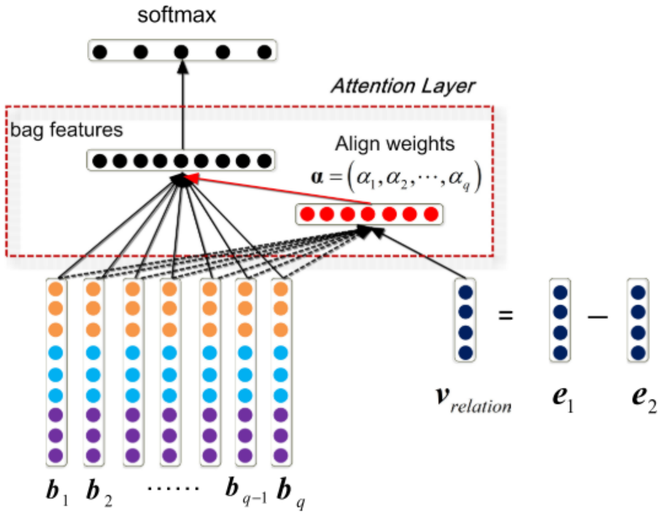


Hierarchical relation extraction with coarse-to-fine grained attention. (EMNLP 2018)

Suppressing Noise – external knowledge

- APCNN
  - Sentence encoder: PCNN
  - Entity descriptions can provide helpful background knowledge for learning entity embedding.
  - Get the relation embedding with TransE
  - Attention mechanism is also used to weight

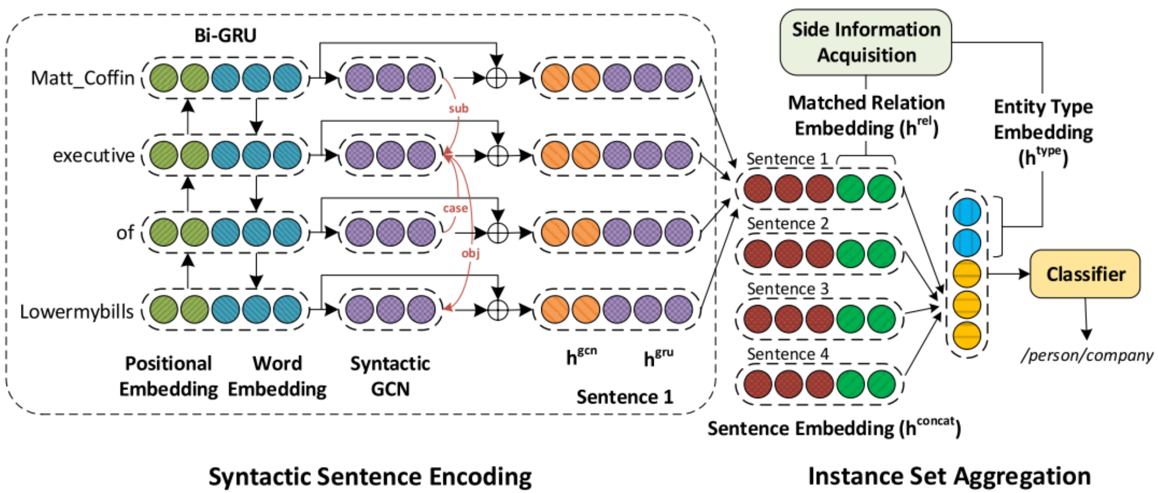
each sentence by relation embedding.



Distant supervision for relation extraction with sentence-level attention and entity descriptions. (AAAI 2017)

Suppressing Noise – external knowledge

- RESIDE
  - Sentence encoder: BiGRU + Syntatic GCN, capturing long-range dependencies
  - Utilize entity type and relation alias information obtained from KB to enhance sentence and bag representation.



RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information. (EMNLP 2018)

Suppress Noise

Unable to deal with the bags in which all sentences are not describing a relation and handle the sentence-level prediction.

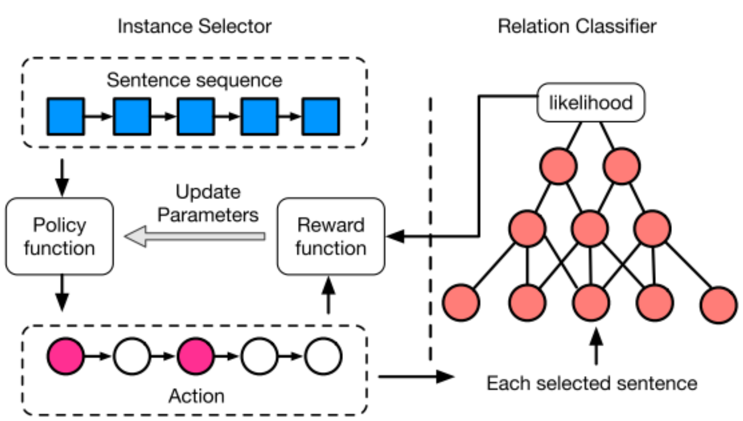
Sentence	**DS Label**	**Gold Label**	**restrain**	**Remove**	**Rectify**
S0: Cook reminded the audience that ***Apple*** was founded by ***Steve Jobs***, the son of a Syrian immigrant.	Founder	Founder	0.7	remain	Founder
S1: ***Steve Jobs*** says ***Apple*** is a curator, nothing more.	Founder	NA	0.1	remove	NA
S2: I followed the work of ***Steve Jobs*** even after he was booted from ***Apple*** in 1985.	Founder	EmployedBy	0.2	remove	EmployedBy

Outline

- Task of Relation Classification
- Distant Supervision for Relation Classification
- Challenges: Noise
  - Suppressing Noise
  - Removing Noise
  - Rectifying Noise
- Summary

Removing Noise

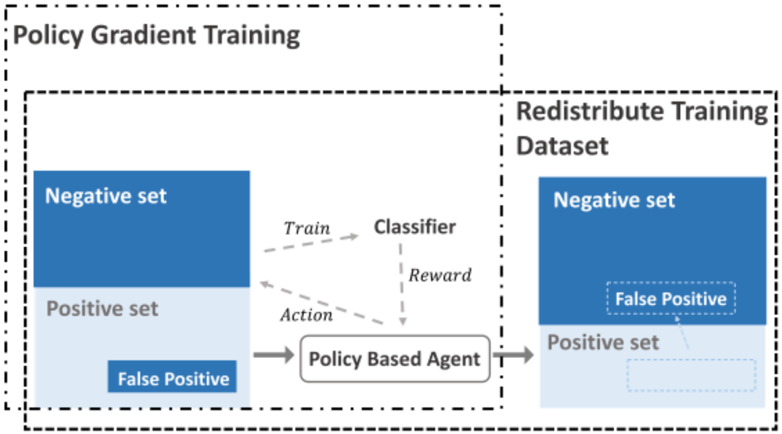
- Reinforcement Learning
  - Distinguish and remove wrong labeled sentences from a sentence bag and train the relation classifier with cleaned dataset.
  - As there is no direct signal indicate whether the sentence is mislabeled or not, this work proposes a reinforcement learning method to solve this problem.



Reinforcement Learning for Relation Extraction from Noisy Data.(AAAI 2018)

Removing Noise

- Reinforcement Learning
  - Learn an agent by reinforcement learning to recognize false positives **for each relation type** and **redistribute training dataset.**
  - Reward is reflected by the performance change of the relation classifier while reward of previous work is calculated from the prediction likelihood.



Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning.(ACL 2018)

Removing Noise

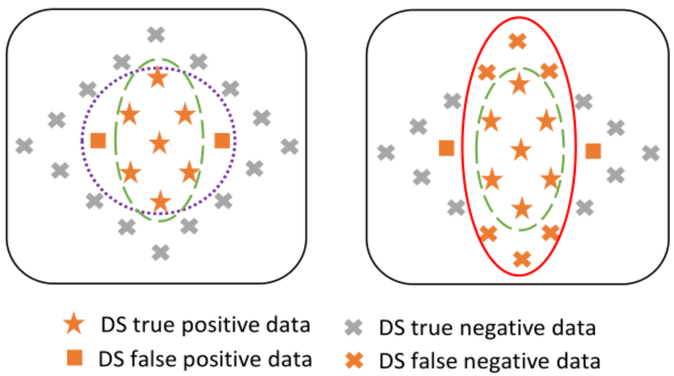
Only address the false positives while the false negatives remain unsolved.

Sentence	DS Label	Gold Label	restraine	Remove	Rectify
S0: Cook reminded the audience that ***Apple*** was founded by ***Steve Jobs***, the son of a Syrian immigrant.	Founder	Founder	0.7	remain	Founder
S1: ***Steve Jobs*** says ***Apple*** is a curator, nothing more.	Founder	NA	0.1	remove	NA
S2: I followed the work of ***Steve Jobs*** even after he was booted from ***Apple*** in 1985.	Founder	EmployedBy	0.2	remove	EmployedBy
S3: The richest youngster, according to Hurun, is ***Ma Huateng***, the 39-year-old co-founder of Internet giant ***Tencent***.	NA	Founder	-	-	Founder

Outline

- Task of Relation Classification
- Distant Supervision for Relation Classification
- Challenges: Noise
  - Suppress Noise
  - Remove Noise
  - Rectify Noise
- Summary

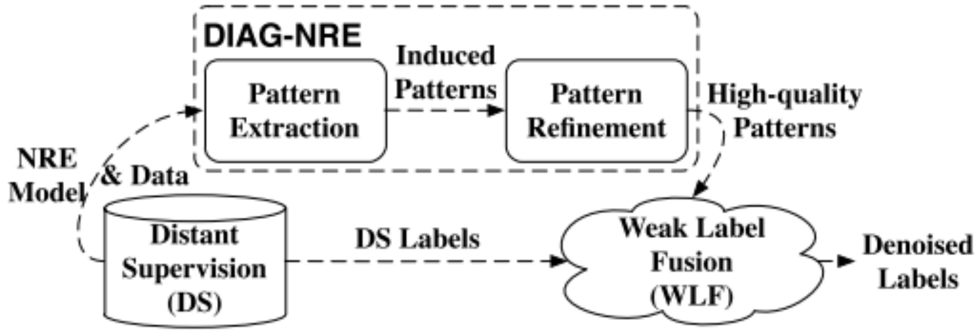
Rectify Noise



- Previous work only suppress or remove the false positives and obtain a suboptimal decision boundary.
- False negatives express similar semantic information with positive data and provide evidence for the target relation.
- Thus, correcting the label of noisy sentences help to obtain the optimal decision boundary.
- For false positives: Wrong label → True label or NA
- For false negatives: NA → True Label

Rectify Noise

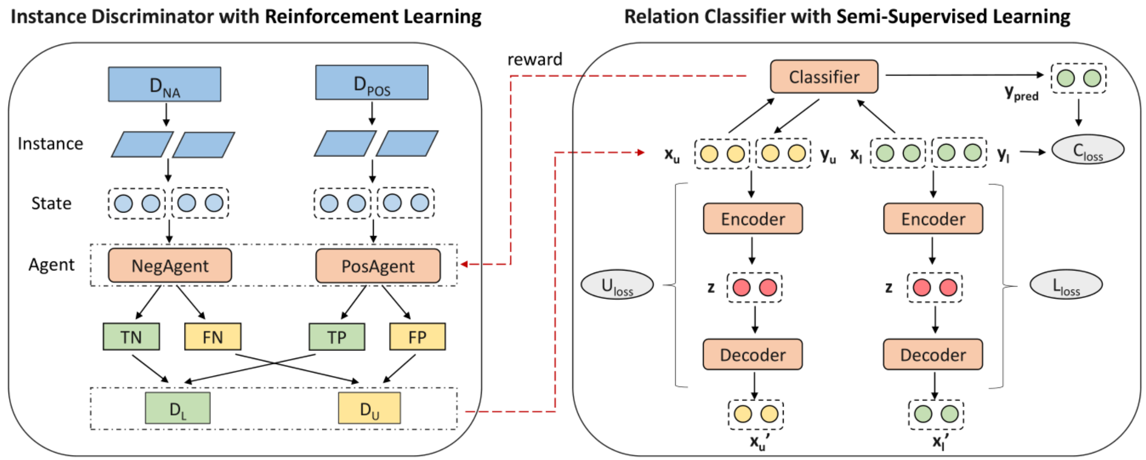
- DIAG-NRE
  - Extract and refine potential patterns from NRE models by employing reinforcement learning and human annotations.
  - Feed them into Weak-Label-Fusion stage to get denoised labels and retrain a model.



DIAG-NRE: A Neural Pattern Diagnosis Framework for Distantly Supervised Neural Relation Extraction. (ACL 2019)

Rectifying Noise

- RCEND



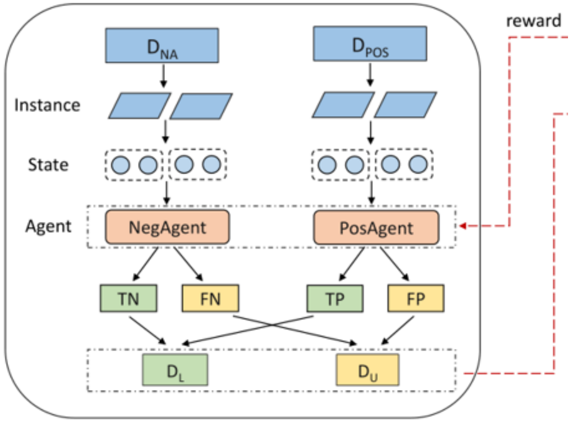
  - Instance discriminator aims to recognize false-positive and false-negative instances from DS dataset.
  - Correctly labeled data are split into labeled data while wrong labeled data are split into unlabeled data.
  - Train a relation classifier with semi-supervised learning utilizing the above data.

Exploiting Noisy Data in Distant Supervision Relation Classification. (NAACL 2019)

Rectifying Noise

RCEND - Instance Discriminator

- Regard discriminator as the agent in reinforcement learning。
- Interact with the environment that consists of a noisy dataset and relation classifier
- Take action to split dataset into correctly and wrong labeled dataset.
- Receive reward to evaluate the redistribution.

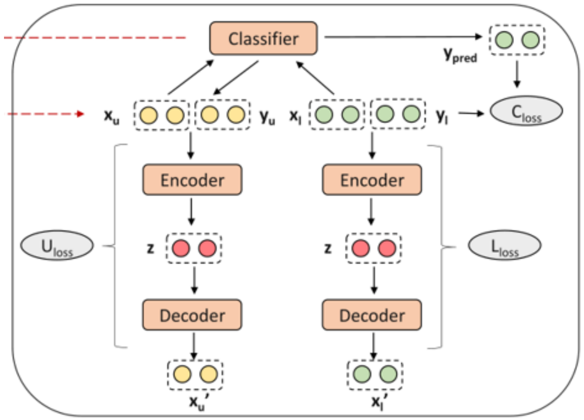


$$r = \lambda\left(\frac{1}{|L|}\sum_{x \in L} p_\phi(y|x) - \frac{1}{|U|}\sum_{x \in U} p_\phi(y|x)\right)$$

Exploiting Noisy Data in Distant Supervision Relation Classification. (NAACL 2019)

Rectifying Noise

- RCEND - Relation Classifier
  - Encoder: encode data x and label y into a latent variable z.
  - Decoder: generate x given z and categorical label y.
  - Classifier: predict the corresponding label y of input x.



Exploiting Noisy Data in Distant Supervision Relation Classification. (NAACL 2019)

Rectifying Noise

RCEND - Experiments

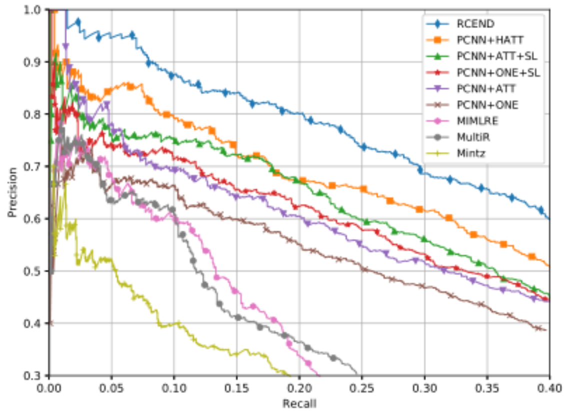


Figure 3: Precision-recall curves of our model and baselines.

P@N	100	200	300	Mean
PCNN+ONE	72.3	69.7	64.1	68.7
PCNN+ATT	76.2	73.1	67.4	72.2
PCNN+ONE+SL	84.0	81.0	74.0	79.7
PCNN+ATT+SL	87.0	84.5	77.0	82.8
PCNN+HATT	88.0	79.5	75.3	80.9
RCEND	**95.0**	**87.5**	**84.4**	**88.9**

Table 3: Top-N precision (P@N) of our model and baselines

- RCEND outperforms feature based models and MIL scheme based models and achieves the best performance.
- RCEND remain high precision when recall increase. It indicates RCEND is more robust and advantageous.

Exploiting Noisy Data in Distant Supervision Relation Classification. (NAACL 2019)

Rectifying Noise

RCEND - Experiments

P@N	100	200	300	Mean
RCEND	**95.0**	**87.5**	**84.4**	**88.9**
RCEND w/o Semi	90.0	84.6	79.1	84.6
RCEND(P)	87.1	82.1	80.1	83.3
RCEND(N)	89.1	85.1	81.1	85.1

Table 4: Top-N precision (P@N) of our model with different settings.

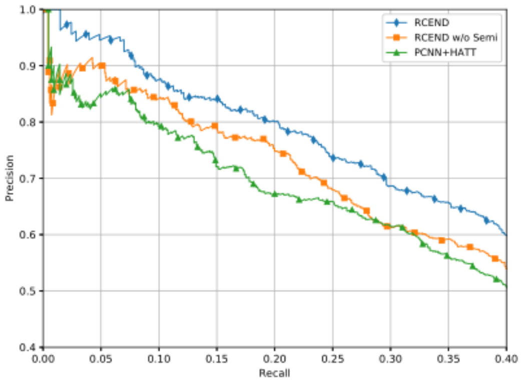


Figure 4: Precision-recall curves of our model with different settings.

- Rectify Noise>Remove Noise>Suppress Noise
- Removing noisy data can solve the problem existing in MIL scheme based models.
- Leveraging information contained in incorrectly labeled data by correcting the noisy label implicitly help to improve the generalization performance.

Rectify Noise

RCEND - Experiments

Type	Sentence	Predict	DS
FN	C1: [**Oliver O'Grady**] is now a silver-haired , twinkly-eyed resident of [**Ireland**] , where Ms. Berg often films him in parks ...	Nationality	NA
	C2: ... said [**John Allison**] , editor of [**Opera**] magazine , based in London.	EmployedBy	NA
	C3: [**Jean-Pierre Bacri**] is a famous writer , who is too self-centered to care about his lonely , overweight , 20-year-old daughter , [**Lolita Marilou Berry**]...	ChildrenOf	NA
FP	C4: they wanted to interview [**Bill Cosby**] after they met with a former Temple University employee who has accused him of groping her in his home in suburban [**Philadelphia**]	LivedIn	BornIn
	C5: "Without the fog , [**London**] wouldn't be a beautiful city."the French painter Claude Monet wrote to his wife , Alice , during one of his long visits to [**England**] from France.	NA	Capital
	C6: MTV Goes to Africa MTV opened its first local music channel in Africa this week , a step touted by the singer [**Lebo Mathosa**] , above , at an MTV event in [**Johannesburg**].	NA	DieIn

- The sentences which are regarded as unlabeled data during training are finally predicted correctly by our relation classifier.
- It shows our model indeed captures the valid information of noisy data.

Summary

- Distant supervision automatically generate large of annotated data but inevitably introduce noisy label.
- Multi-instance learning paradigm reduces the impact of noisy data to some extent and directly removing the noisy data further improves the performance of distant supervised RE.
- Rectify the noisy data can adequately exploit the valid information contained in training data and achieve a better model.
- In addition, incorporate with external knowledge to enhance model is also a promising choice.