XI'AN JIAOTONG UNIVERSITY

语义关系抽取
Relation Extraction (RE)

李辰

2024年10月

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Introduction

● Relationship Extraction

● **Relationship Extraction**: identify mentions of the relations of interest in each sentence of the given documents.

Example: “International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…”

Introduction

- ## Relationship Extraction

- **Relationship Extraction**: identify mentions of the relations of interest in each sentence of the given documents.

Example: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"

Extracted Complex Relation:

Company-Founding

Company	IBM
Location	New York
Date	June 16, 1911
Original-Name	Computing-Tabulating-Recording Co.

Focus on the simpler task of extracting relation **triples**

Founding-year (IBM, 1911)

Founding-location (IBM, New York)

Introduction

● Relationship Extraction

● **Relationship Extraction**: identify mentions of the relations of interest in each sentence of the given documents.

Example: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"

Extracted Complex Relation:

Company-Founding

Company	IBM
Location	New York
Date	June 16, 1911
Original-Name	Computing-Tabulating-Recording Co.

Focus on the simpler task of extracting relation **triples**

Founding-year (IBM, 1911)

Founding-location (IBM, New York)

● **Note:** it is possible to treat relation tagging as a classification problem, classifying each pair as a relation type or NONE.

Introduction

● Main goals of relation extraction

- Fill a predefined "template" from raw text
- Extract who did what to whom and when?
    - Event extraction
- Organize information so that is useful to people
- Put information in a form that allows further inferences by computers
    - Big data
- Create new structured knowledge bases, useful for any app
- Augment current knowledge bases
    - Such as: WordNet, FreeBase, DBPedia, Biomodels
- Support question answering

Introduction

- **Main goals of relation extraction**
  - Fill a predefined "template" from raw text
  - Extract who did what to whom and when?
    - Event extraction
  - Organize information so that is useful to people
  - Put information in a form that allows further inferences by computers
    - Big data
  - Create new structured knowledge bases, useful for any app
  - Augment current knowledge bases
    - Such as: WordNet, FreeBase, DBPedia, Biomodels
  - Support question answering

Some concrete examples

- Extracting earnings, profits, board members, headquarters from company reports
- Searching on the WWW for e-mails for advertising (spamming)
- Learn drug-drug or gene-gene interactions from biomedical text.

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Hand-built patterns

- **Hand-built patterns for relations**
  - Idea: define some extraction patterns

    Y such as X ((, X)* (, and/or) X)

    such Y as X…

    X… or other Y

    X… and other Y

    Y including X…

    Y, especially X…

Hearst, 1992. Automatic Acquisition of Hyponyms.

Hand-built patterns

- **Hearst's lexico-syntactic patterns**
  - Idea: define some extraction patterns

    Y such as X ((, X)* (, and/or) X)

    such Y as X…

    X… or other Y

    X… and other Y

    Y including X…

    Y, especially X…

Example:

Hearst pattern	Example occurrences
X and other Y	…temples, treasuries, and other important civic buildings.
X or other Y	bruises, wounds, broken bones or other injuries...
Y such as X	The bow lute, such as the Bambara ndang...
Such Y as X	...such authors as Herrick, Goldsmith, and Shakespeare.
Y including X	...common-law countries, including Canada and England...
Y, especially X	European countries, especially France, England, and Spain...

Hearst, 1992. Automatic Acquisition of Hyponyms.

Hand-built patterns

- **Pros and Cons**

  - Pros
    - Human patterns tend to be high-precision
    - Can be tailored to specific domains
    - Easy to debug: why a prediction was made, how to fix?

Hand-built patterns

- **Pros and Cons**

➢ Pros
  - Human patterns tend to be high-precision
  - Can be tailored to specific domains
  - Easy to debug: why a prediction was made, how to fix?

➢ Cons
  - Human patterns are often low-recall
  - Requires hand-building patterns for each relation
    - hard to write; hard to maintain
    - there are zillions of them
    - domain-dependent
  - Requires higher accuracy
    - Hearst: 66% accuracy on hyponym extraction

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6.Joint Extraction

Supervised techniques

- **The supervised approach requires:**
  - Defining an inventory of output labels
    - Relation detection: true/false
    - Relation classification: located-in, employee-of, inventor-of, …
  - Collecting labeled training data: MUC, ACE, …
  - Defining a feature representation: words, entity types, …
  - Choosing a classifier: Naïve Bayes, MaxEnt, SVM, …
  - Evaluating the results

Supervised techniques

- **ACE 2008: relations**

  The ACE dataset is a relational classification dataset, which predefines five major categories of relationships: location, institution, membership, whole-part, and person-society.

Type	Subtype
ART (artifact)	User-Owner-Inventor-Manufacturer
GEN-AFF (General affiliation)	Citizen-Resident-Religion-Ethnicity, Org-Location
METONYMY*	*None*
ORG-AFF (Org-affiliation)	Employment, Founder, Ownership, Student-Alum, Sports-Affiliation, Investor-Shareholder, Membership
PART-WHOLE (part-to-whole)	Artifact, Geographical, Subsidiary
PER-SOC* (person-social)	Business, Family, Lasting-Personal
PHYS* (physical)	Located, Near

Supervised techniques

- **Feature-based**

Feature Types	**Example**
Words: Words of both the mentions and all the words in between	M11 leaders, M21 Venice; B1 of, B2 Italy, B3 ’ s, B4 left-wing, B5 government, B6 were, B7 in
Entity Types: Entity types of both the mentions	E1 PERSON, E2 GPE
Mention Level: Mention types (NAME, NOMINAL or PRONOUN) of both the mentions	M1 NOMINAL, M2 NAME
Overlap: #words separating the two mentions, #other mentions in between, flags indicating whether the two mentions are in the same NP, VP or PP	7 Words Apart, 2 Mentions In Between (Italy & government), Not Same NP, Not Same VP, Not Same PP
Dependency: Words, POS and chunk labels of words on which the mentions are dependent in the dependency tree, #links traversed in dependency tree to go from one mentions to another	M1W were, M1P VBD, M1C VP, M2W in, M2P IN, M2C PP, DepLinks 3
Parse Tree: Path of non-terminals connecting the two mentions in the parse tree, and the path annotated with head words	PERSON-NP-S-VP-PP-GPE, PERSON-NP:leaders-S -VP:were-PP:in-GPE

Table: Various feature types with examples described by Kambhatla [1]

Supervised techniques

- **Feature-based**
  - These approaches require **labelled data** where each pair of entity mentions is labelled with one of the pre-defined relation types.
- **References for feature-based methods**

[1] N. Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In Proceedings of the ACL 2004, 2004.
[2] Zhou GuoDong, Su Jian, Zhang Jie, and Zhang Min. Exploring various knowledge in relation extraction. In Proceedings of the 43rd annual 40 meeting on association for computational linguistics, pages 427–434. Association for Computational Linguistics, 2005.
[3] Jing Jiang. Multi-task transfer learning for weakly-supervised relation extraction. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1012– 1020. Association for Computational Linguistics, 2009.
[4] Dat PT Nguyen, Yutaka Matsuo, and Mitsuru Ishizuka. Relation extraction from wikipedia using subtree mining. In Proceedings of the National Conference on Artificial Intelligence, volume 22, page 1414. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.
[5] Yee Seng Chan and Dan Roth. Exploiting syntactico-semantic structures for relation extraction. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 551–560. Association for Computational Linguistics, 2011.

Supervised techniques

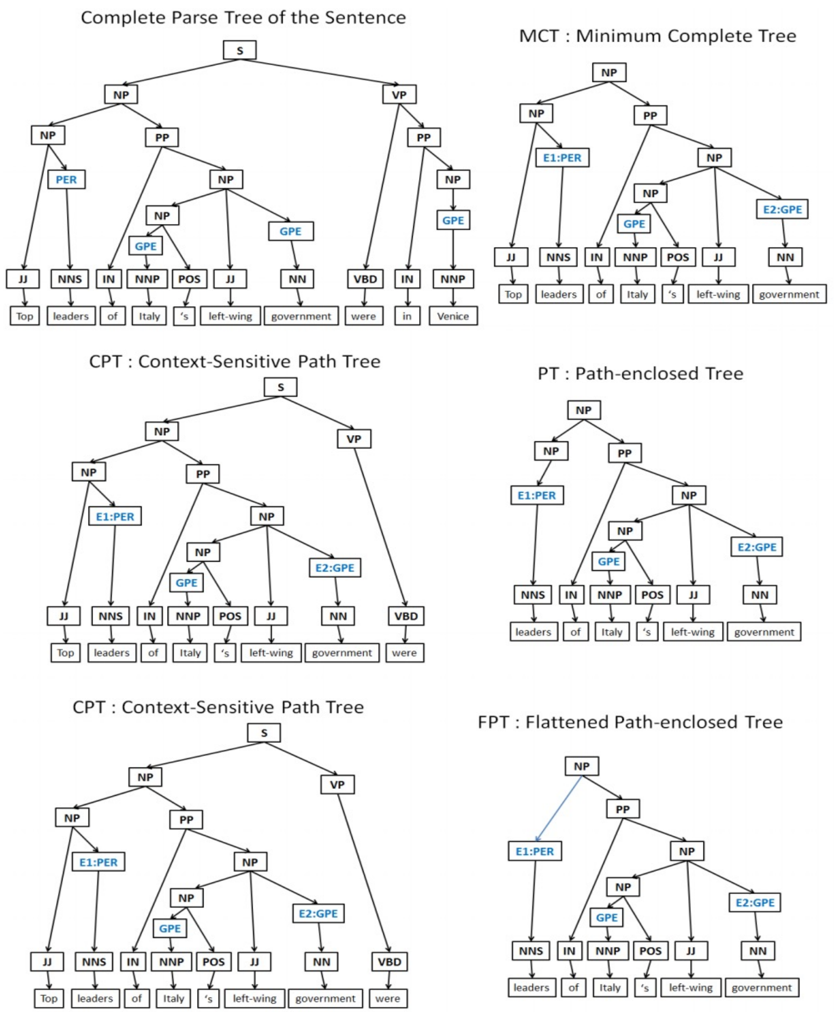
- **Kernel Methods**

  - The main advantage of kernel based methods is that such explicit feature engineering is avoided. In kernel based methods, kernel functions are designed to compute similarities between representations of two relation instances

  - Sequence Kernel
  - Syntactic Tree Kernel
  - Dependency Tree Kernel
  - Dependency Graph Path Kernel
  - Composite Kernels

Supervised techniques

- **Kernel Methods**



Various tree representations described in Zhang et al. [1]

Supervised techniques

- **Comparison between the two types of methods**

Type	Approach	P	R	F
Feature-based	LX, ST and DT based features [2]	0.737	0.694	0.715
	Additional features based on Syntactico-Semantic structures [3]	0.754	0.680	0.715
Kernel-based	Composite kernel combining individual LX, ST and DT kernel [4]	0.692	0.705	0.704
	Composite kernel combining ST and EN kernels [5]	0.761	0.684	0.721
	ST kernel with dynamically determined tree span [6]	0.822	0.702	0.758
	Composite kernel combining ST and DT kernels along with semantic information [7]	0.766	0.670	0.715
	ST kernel where parse tree is augmented with entity features [8]	0.792	0.674	0.728
	ST kernel with dynamically determined tree span [9]	0.830	0.720	0.771

Table: Comparative relation extraction performance of various supervised approaches on ACE 2004 dataset (5-fold cross validation, LX: lexical, ST: syntactic tree, DT: dependency tree)

Supervised techniques

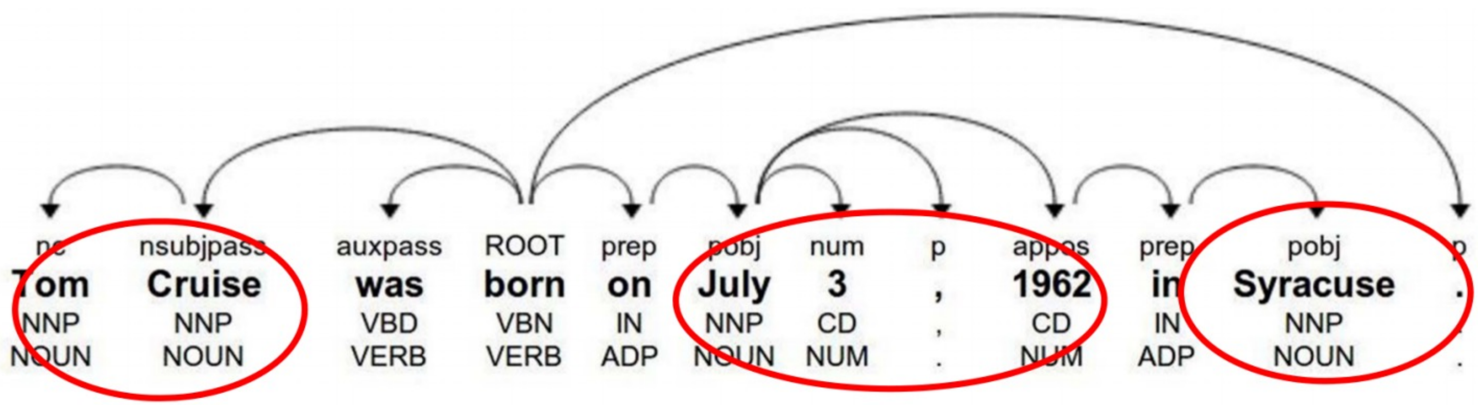
- **How to apply a statistical classifier for relation extraction**

- Preprocess raw document
- Run NER
- Run any other auxiliary tools, such as POS tagging, dependency parsing
- For all pairs of entities (within each sentence or paragraph): decide whether they are related or not (binary classification)
- For related entity pairs: classify the relation type

Supervised techniques

- **Typical features for the statistical classifier**

- Context words + POS
- Dependency path between entities
- Named entity tags
- Token/parse-path/entity distance



X was born on DDDD in Y

- DEP: X <*nsubjpass* / born *prep*> on *pobj*> DATE *prep*> in *pobj*> Y
- NER: X = PER, Y = LOC
- POS: X = NOUN, NNP; Y = NOUN, NNP
- Context: born, on, in , "born on"

Supervised techniques

- **Example**

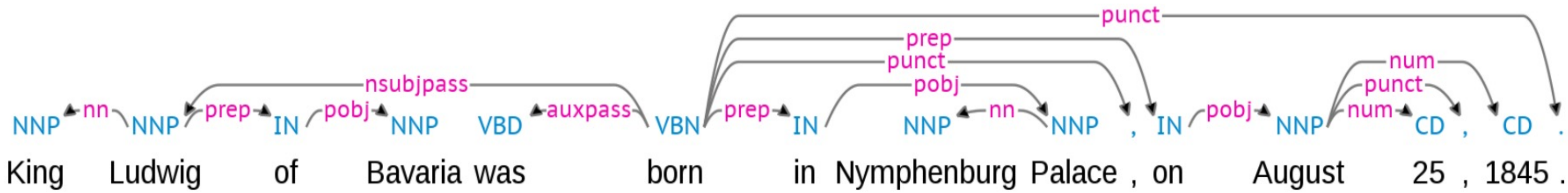
  - Preprocess

    King Ludwig of Bavaria was born in Nymphenburg Palace , on August 25 , 1845.

  - Run NER (PERSON / DATE / LOCATION)

    King Ludwig of Bavaria was born in Nymphenburg Palace , on August 25 , 1845.

  - Run auxiliary tools



  - For all pairs of entities, decide whether they are related or not

    (King Ludwig of Bavaria, Nymphenburg Palace) ? Related.

    (King Ludwig of Bavaria, August 25 , 1845) ? Related.

    (Nymphenburg Palace, August 25 , 1845) ? Unrelated.

  - For related entity pairs, classify the relation type

    Born_in (King Ludwig of Bavaria, Nymphenburg Palace)

    Born_on (King Ludwig of Bavaria, August 25 , 1845)

Supervised techniques

- **Pros and Cons**

➢ Pros

- Can get high accuracies with enough hand-labeled training data, if test similar enough to training

Supervised techniques

- **Pros and Cons**

➢ Pros
- Can get high accuracies with enough hand-labeled training data, if test similar enough to training

➢ Cons
- Annotating a large corpus is expensive
- Supervised models are brittle, don't generalize well to different domains (topics and genres)

Supervised techniques

• References for kernel-based methods

[1] Min Zhang, Jie Zhang, Jian Su, and Guodong Zhou. A composite kernel to extract relations between entities with both flat and structured features. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pages 825–832. Association for Computational Linguistics, 2006.

[2] Jing Jiang and ChengXiang Zhai. A systematic exploration of the feature space for relation extraction. In HLT-NAACL, pages 113–120, 2007.

[3] Yee Seng Chan and Dan Roth. Exploiting syntactico-semantic structures for relation extraction. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 551–560. Association for Computational Linguistics, 2011.

[4] Shubin Zhao and Ralph Grishman. Extracting relations with integrated information using kernel methods. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pages 419–426. Association for Computational Linguistics, 2005.

[5] Min Zhang, Jie Zhang, Jian Su, and Guodong Zhou. A composite kernel to extract relations between entities with both flat and structured features. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pages 825–832. Association for Computational Linguistics, 2006.

[6] GuoDong ZHOU, Min Zhang, Dong Hong Ji, and Qiaoming Zhu. Tree kernel-based relation extraction with context-sensitive structured parse tree information. EMNLP-CoNLL 2007, page 728, 2007.

[7] Truc-Vien T Nguyen, Alessandro Moschitti, and Giuseppe Riccardi. Convolution kernels on constituent, dependency and sequential structures for relation extraction. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3, pages 1378–1387. Association for Computational Linguistics, 2009.

[8] Longhua Qian, Guodong Zhou, Qiaomin Zhu, and Peide Qian. Relation extraction using convolution tree kernel expanded with entity features. In Proceedings of the 21st Pacific Asian Conference on Language, Information and Computation (PACLIC-21), pages 415–421, 2007.

[9] Deepak Ravichandran and Eduard Hovy. Learning surface text patterns for a question answering system. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pages 41–47. Association for Computational Linguistics, 2002.

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Semi-supervised techniques

● Motivation

It is cost, effort and time intensive task to generate labelled data for relation extraction. Major motivation of semi-supervised techniques is:

- To reduce the manual efforts required to create labelled data;
- Exploit the unlabelled data which is generally easily available without investing much efforts.

Semi-supervised techniques

● Motivation

It is cost, effort and time intensive task to generate labelled data for relation extraction. Major motivation of semi-supervised techniques is:

- To reduce the manual efforts required to create labelled data;
- Exploit the unlabelled data which is generally easily available without investing much efforts.

● Approaches

- Bootstrapping Approaches

  Require a large unlabelled corpus and a few seed instances of the relation type of interest. [1,2]
- Active learning

  The key idea behind active learning is that the learning algorithm is allowed to ask for true labels of some selected unlabelled instances. [3,4]
- Label Propagation Method

  A graph based semi-supervised method [5]
- Others

  i.e.multi-task transfer learning [6]

Semi-supervised techniques

- **Typical instance: bootstrapping approach**

Name: Dual Iterative Pattern Relation Expansion (DIPRE) [1]:

- Given a good set of patterns, a good set of tuples (entity pairs following a certain relation type) can be found;
- Given a good set of tuples, a good set of patterns can be learned.

Semi-supervised techniques

● Typical instance: bootstrapping approach

Name: Dual Iterative Pattern Relation Expansion (DIPRE) [1]:

- Given a good set of patterns, a good set of tuples (entity pairs following a certain relation type) can be found;
- Given a good set of tuples, a good set of patterns can be learned.

● Overview of DIPRE

Input: Seed set S of tuples, i.e. entity pairs known to be related with certain relation type R

Output: Set S grown over multiple iterations

1. Find all occurrences of the tuples from the seed set S on the Web
2. Learn patterns from these occurrences
3. Search the web using these patterns and find new tuples and add to the set S
4. Go to step 1 and iterate till there are no new tuples to be added

Semi-supervised techniques

- **Example: DIPRE—Extract<author,book> pairs**
  - Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

  - Find instances on the web:

    The Comedy of Errors, by William Shakespeare, was
    The Comedy of Errors, by William Shakespeare, is
    The Comedy of Errors, one of William Shakespeare's earliest attempts
    The Comedy of Errors, one of William Shakespeare's most

  - Extract patterns (group by middle, take longest common prefix/suffix)

    ?x , by ?y ,   ?x , one of ?y ' s

  - Iterate, finding new seeds that match the pattern

Semi-supervised techniques

- **Pros and Cons**

  - Pros
    - More relations can be discovered
    - Less human effort.

  - Cons
    - The set of patterns becomes more error-prone with each iteration
    - Need to find best practices for generalizing the context around tuple occurrences when generating new patterns
    - Extending to new relation types requires additional seed patterns, to be written manually

Semi-supervised techniques

- **References**

[1] Sergey Brin. Extracting patterns and relations from the world wide web. In The World Wide Web and Databases, pages 172–183. Springer, 1999.
[2] Zhu Zhang. Weakly-supervised relation classification for information extraction. In Proceedings of the thirteenth ACM international conference on Information and knowledge management, pages 581–588. ACM, 2004.
[3] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
[4] Ang Sun and Ralph Grishman. Active learning for relation type extension with local and global data views. In Proceedings of the 21st ACM international conference on Information and knowledge management, pages 1105–1112. ACM, 2012.
[5] Zhu Xiaojin and Ghahramani Zoubin. Learning from labeled and unlabeled data with label propagation. In CMU CALD tech report CMU CALD-02-107, 2002.
[6] Jing Jiang. Multi-task transfer learning for weakly-supervised relation extraction. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1012– 1020. Association for Computational Linguistics, 2009.

Unsupervised techniques

- Do **not** require any **labelled data**

- **References**

[1] Takaaki Hasegawa, Satoshi Sekine, and Ralph Grishman. Discovering relations among named entities from large corpora. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, page 415. Association for Computational Linguistics, 2004.
[2] Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. Unsupervised feature selection for relation extraction. In Proceedings of IJCNLP, 2005.

Unsupervised techniques

- Do **not** require any **labelled data**
- **Clustering based approaches**

The approach can be described in following steps [1]:

- The named entities in the text corpora are tagged (use the named entity recognition tagger)
- Co-occurring named entity pairs are formed and their contexts are recorded
- Context similarities among the pairs identified in the step 2, are computed
- Using the similarity values computed in previous step, the pairs are clustered
- As each of these clusters represent one relation, a label is automatically assigned to each cluster describing the relation type represented by it

- **References**

[1] Takaaki Hasegawa, Satoshi Sekine, and Ralph Grishman. Discovering relations among named entities from large corpora. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, page 415. Association for Computational Linguistics, 2004.

[2] Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. Unsupervised feature selection for relation extraction. In Proceedings of IJCNLP, 2005.

Unsupervised techniques

- **Pros and Cons**

➢ Pros
- Patterns tend to be high-precision
- Can be tailored to specific domains

➢ Cons
- Human patterns are often low-recall
- A lot of work to think of all possible patterns
- We'd like better accuracy

Distant Supervision

- **Do not require labelled data**

- Relations from an existing **knowledge base** can be employed for corpus annotation

Distant Supervision

- **Do not require labelled data**

- Relations from an existing **knowledge base** can be employed for corpus annotation

- Automatically create corpus annotation by labeling all cooccurrences of entity pairs that are related according to the knowledge base
  assuming that sentences that contain a related pair are expressing the type(s) of relationship(s) that these entities have in the knowledge base

- Ideally, evaluate on a small amount of gold-standard data

Distant Supervision

- **Example:**

① For each relation	Born-In
② For each tuple in big database	<Edwin Hubble, Marshfield> <Albert Einstein, Ulm>
③ Find sentences in large corpus with both entities	Hubble was born in Marshfield Einstein, born (1879), Ulm Hubble's birthplace in Marshfield
① Extract frequent features(parse, words, etc)	PER was born in LOC PER, born (XXXX), LOC PER's birthplace in LOC
② Train supervised classifier using these patterns	P(born-in | f1,f2,f3,…,f70000)

Distant Supervision

- **Pros and Cons**

➢ Pros

- Less manual effort
- Scalability: can use large amounts of unlabeled data

➢ Cons

- Noisy automatic annotation of the training corpus: sentences with entity cooccurrences might not express the seed relation
- No explicit negative samples for classifier training

Distant Supervision

- **References**

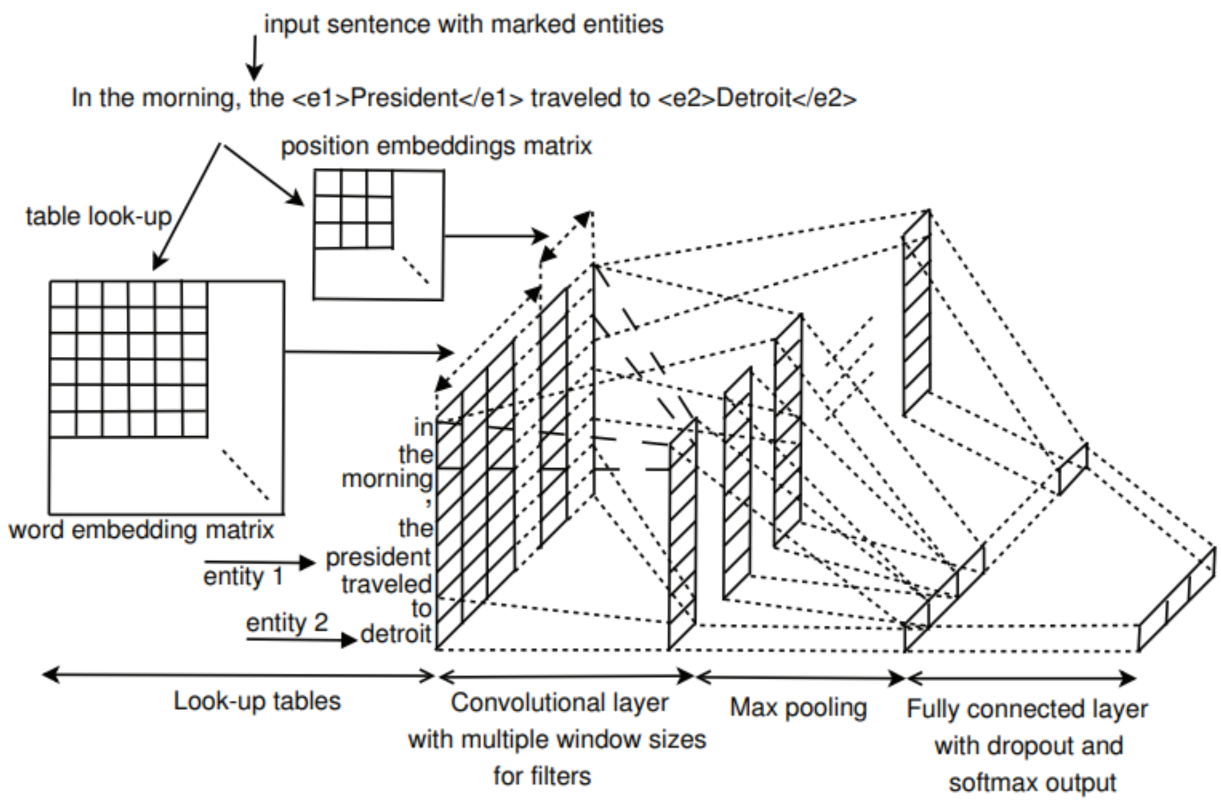
[1] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 43 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1003–1011. Association for Computational Linguistics, 2009.
[2] Mark Craven and Johan Kumlien. Constructing biological knowledge bases by extracting information from text sources. In ISMB, volume 1999, pages 77–86, 1999.
[3] Razvan Bunescu and Raymond Mooney. Learning to extract relations from the web using minimal supervision. In Annual meeting-association for Computational Linguistics, volume 45, page 576, 2007.
[4] Dat PT Nguyen, Yutaka Matsuo, and Mitsuru Ishizuka. Exploiting syntactic and semantic information for relation extraction from wikipedia. In IJCAI Workshop on Text-Mining & Link-Analysis (TextLink 2007), 2007.
[5] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pages 1247–1250. ACM, 2008.

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Deep Learning Method

- **CNN-based Methods**
  - **Word Representation**: the word embeddings and the position embeddings are concatenated.
  - **Convolution**: multiple window sizes filters bring more structured information to the model.



Nguyen T H , Grishman R . Relation extraction: Perspective from convolutional neural networks[C]// Workshop on Vector Space Modeling for Natural Language Processing. 2015.

Deep Learning Method

CNN-based Methods

- **Experiments:** the model proposed in this paper achieves the best performance without any external resources.

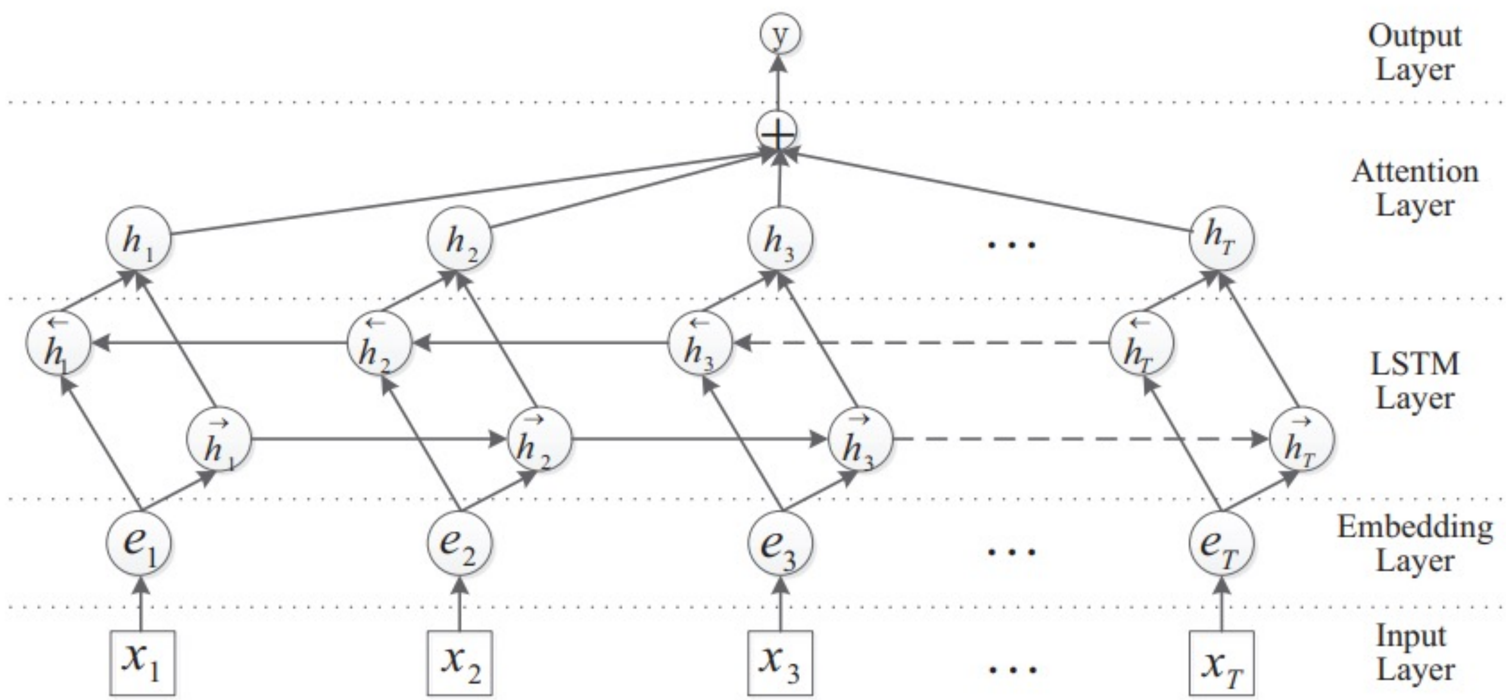
Classifier	Feature Sets	F
SVM	POS, WordNet, morphological features, thesauri, Google n-grams	77.6
MaxEnt	POS, WordNet, morphological features, noun compound system, thesauri, Google n-grams	77.6
SVM	POS, WordNet, prefixes and other morphological features, dependency parse, Levin classes, PropBank, FrameNet, NomLex-Plus, Google n-grams, paraphrases, TextRunner	82.2
RNN	-	74.8
RNN	POS, name tagging, WordNet	77.6
MVRNN	-	79.1
MVRNN	POS, name tagging, WordNet	82.4
O-CNN	-	78.9
O-CNN	WordNet	82.7
FCM	-	80.6
FCM	dependency parse, name tagging	83.0
Our CNN	-	**82.8**

Nguyen T H , Grishman R . Relation extraction: Perspective from convolutional neural networks[C]// Workshop on Vector Space Modeling for Natural Language Processing. 2015.

Deep Learning Method

- **RNN-based Methods**

  - Come up with the attention mechanism in BiLSTM, which automatically highlight the important features.



Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, Xu B (2016) Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol 2, pp 207–212

Deep Learning Method

- **RNN-based Methods**

  - Experiments:

Model	Feature Set	F1
SVM (Rink and Harabagiu, 2010)	POS, prefixes, morphological, WordNet, dependency parse, Levin classed, ProBank, FramNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
CNN (Zeng et al., 2014)	WV (Turian et al., 2010) (dim=50)	69.7
	+ PF + WordNet	82.7
RNN (Zhang and Wang, 2015)	WV (Turian et al., 2010) (dim=50) + PI	80.0
	WV (Mikolov et al., 2013) (dim=300) + PI	82.5
SDP-LSTM (Yan et al., 2015)	WV (pretrained by word2vec) (dim=200), syntactic parse	82.4
	+ POS + WordNet + grammar relation embeddings	83.7
BLSTM (Zhang et al., 2015)	WV (Pennington et al., 2014) (dim=100)	82.7
	+ PF + POS + NER + WNSYN + DEP	84.3
BLSTM	WV (Turian et al., 2010) (dim=50) + PI	80.7
Att-BLSTM	WV (Turian et al., 2010) (dim=50) + PI	82.5
BLSTM	WV (Pennington et al., 2014) (dim=100) + PI	82.7
Att-BLSTM	WV (Pennington et al., 2014) (dim=100) + PI	84.0

Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, Xu B (2016) Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol 2, pp 207–212

Outline

1. Introduction to Relation Extraction
2. Hand-built patterns
3. Supervised Machine Learning
4. Semi and Unsupervised Learning
5. Deep Learning Methods
6. Joint Extraction

Joint Extraction

- pipeline techniques

The pipeline techniques try to find all possible entity pairs available in the text in the first phase and then attempt to classify the relations of the pairs in the second phase. However, there are a lot of drawbacks accompanying this pipeline method that we describe in the following points:

Joint Extraction

- pipeline techniques

The pipeline techniques try to find all possible entity pairs available in the text in the first phase and then attempt to classify the relations of the pairs in the second phase. However, there are a lot of drawbacks accompanying this pipeline method that we describe in the following points:

- Error in the NER subtask can propagate to the RE subtask. For example, if the system misidentifies a named entity, the approach misassigns it to a class of relation.

Joint Extraction

- pipeline techniques

The pipeline techniques try to find all possible entity pairs available in the text in the first phase and then attempt to classify the relations of the pairs in the second phase. However, there are a lot of drawbacks accompanying this pipeline method that we describe in the following points:

- ➢ Error in the NER subtask can propagate to the RE subtask. For example, if the system misidentifies a named entity, the approach misassigns it to a class of relation.

- ➢ There will be a highly imbalanced classes of relations because there are many pairs available in texts that do not belong to any relation class. Therefore, the false class, which is the class of not having a relation, will get the most members.

Joint Extraction

- pipeline techniques

The pipeline techniques try to find all possible entity pairs available in the text in the first phase and then attempt to classify the relations of the pairs in the second phase. However, there are a lot of drawbacks accompanying this pipeline method that we describe in the following points:

- Error in the NER subtask can propagate to the RE subtask. For example, if the system misidentifies a named entity, the approach misassigns it to a class of relation.

- There will be a highly imbalanced classes of relations because there are many pairs available in texts that do not belong to any relation class. Therefore, the false class, which is the class of not having a relation, will get the most members.

- Uncovering a massive number of entity pairs increases the complexity of the problem.

Joint Extraction

- pipeline techniques

The pipeline techniques try to find all possible entity pairs available in the text in the first phase and then attempt to classify the relations of the pairs in the second phase. However, there are a lot of drawbacks accompanying this pipeline method that we describe in the following points:

- Error in the NER subtask can propagate to the RE subtask. For example, if the system misidentifies a named entity, the approach misassigns it to a class of relation.

- There will be a highly imbalanced classes of relations because there are many pairs available in texts that do not belong to any relation class. Therefore, the false class, which is the class of not having a relation, will get the most members.

- Uncovering a massive number of entity pairs increases the complexity of the problem.

- The classifier will be confused when the same subject and object belong to more than one relation classes.

Joint Extraction

- Joint Extraction

  To avoid this propagation of errors, there is a line of research which models or extracts entities and relations jointly

  ➢ 《Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme》(ACL2017)

Joint Extraction

- Novel Tagging Scheme

Input Sentence:	The	United	States	President	Trump	will	visit	the	Apple	Inc	founded	by	Steven	Paul	Jobs

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Novel Tagging Scheme

Input Sentence: The **United States** President **Trump** will visit the **Apple Inc** founded by **Steven Paul Jobs**

Entities:
- United States
- Trump
- Apple Inc
- Steven Paul Jobs

Relations:
- {United States, Country-President, Trump}
- {Apple Inc, Company-Founder, Steven Paul Jobs}

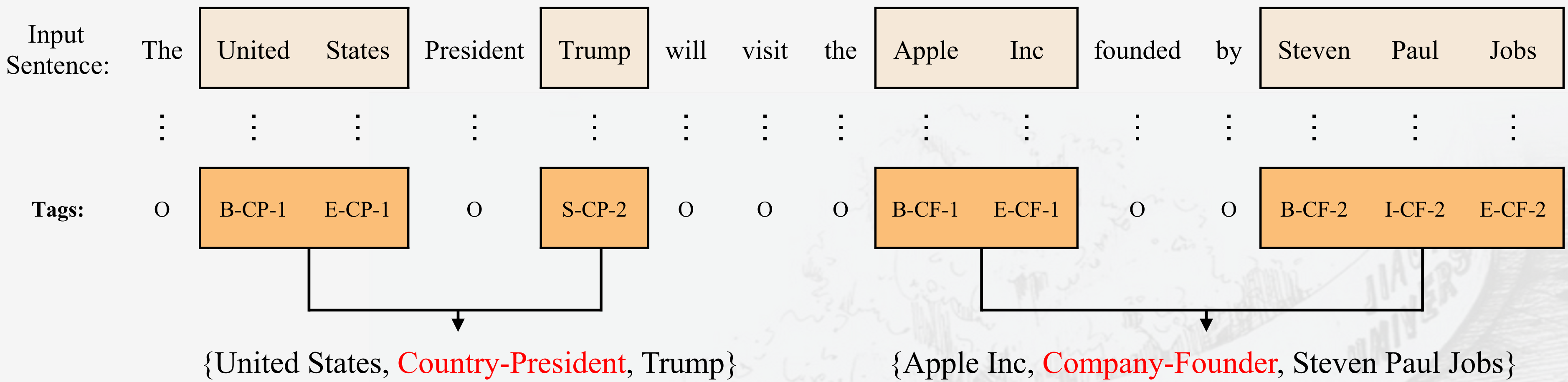
"Country-President" and "Company-Founder" are the predefined relation types.

The input sentence contains two triplets: {United States, Country-President, Trump} and {Apple Inc, Company-Founder, Steven Paul Jobs}, where "Country-President" and "Company-Founder" are the predefined relation types. The words "United", "States", "Trump", "Apple", "Inc", "Steven", "Paul" and "Jobs" are all related to the final extracted results.

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Novel Tagging Scheme

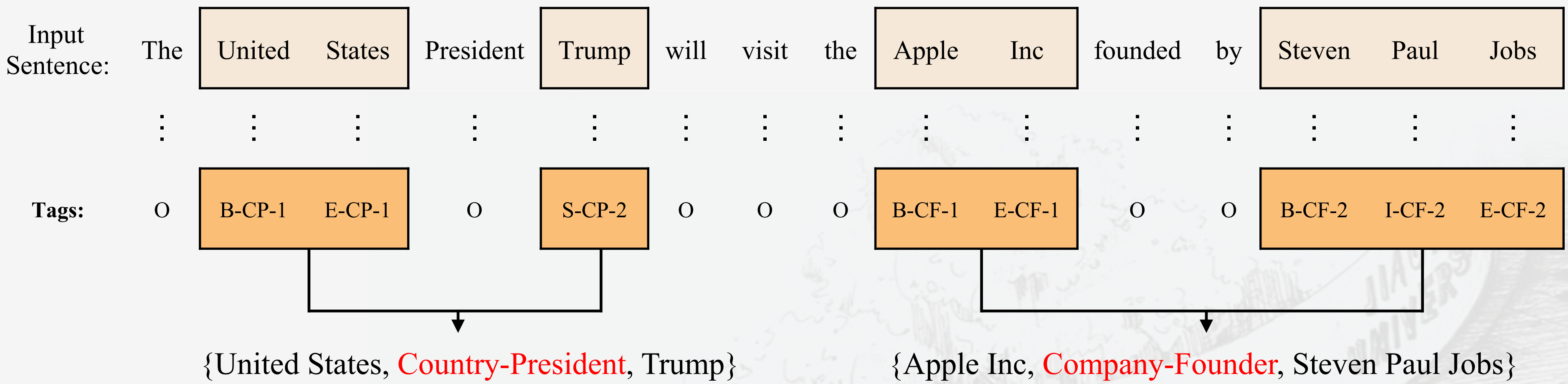


Each word is assigned a label that contributes to extract the results. Tag "O" represents the "Other" tag, which means that the corresponding word is independent of the extracted results. In addition to "O", the other tags consist of three parts: the word position in the entity, the relation type, and the relation role. We use the "BIES" (Begin, Inside, End,Single) signs to represent the position information of a word in the entity. The relation type information is obtained from a predefined set of relations and the relation role information is represented by the numbers "1" and "2". An extracted result is represented by a triplet: (Entity1, RelationT ype, Entity2). "1" means that the word belongs to the first entity in the triplet, while "2" belongs to second entity that behind the relation type.

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Novel Tagging Scheme

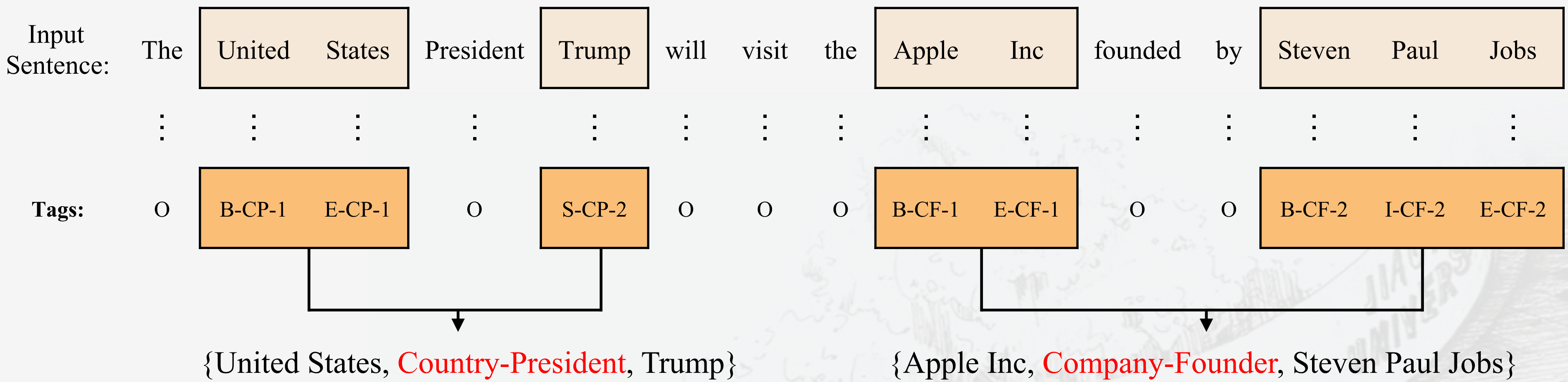


For example, the word of "United" is the first word of entity "United States" and is related to the relation "Country-President", so its tag is "B-CP-1". The other entity "Trump", which is corresponding to "United States", is labeled as "S-CP-2". Besides, the other words irrelevant to the final result are labeled as "O".

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Novel Tagging Scheme

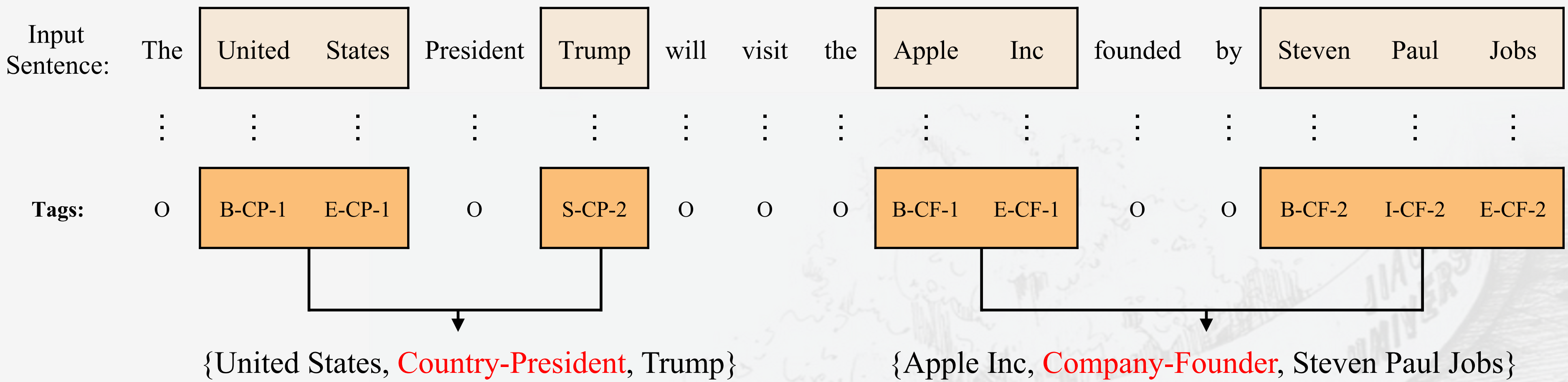


➢ Besides, if a sentence contains two or more triplets with the same relation type, we combine every two entities into a triplet based on the **nearest principle**. For example, if the relation type "Country-President" in Figure 2 is "Company-Founder", then there will be four entities in the given sentence with the same relation type. "United States" is closest to entity "Trump" and the "Apple Inc" is closest to "Jobs", so the results will be {United States, Company-Founder, Trump} and {Apple Inc, Company-Founder, Steven Paul Jobs}.

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Novel Tagging Scheme



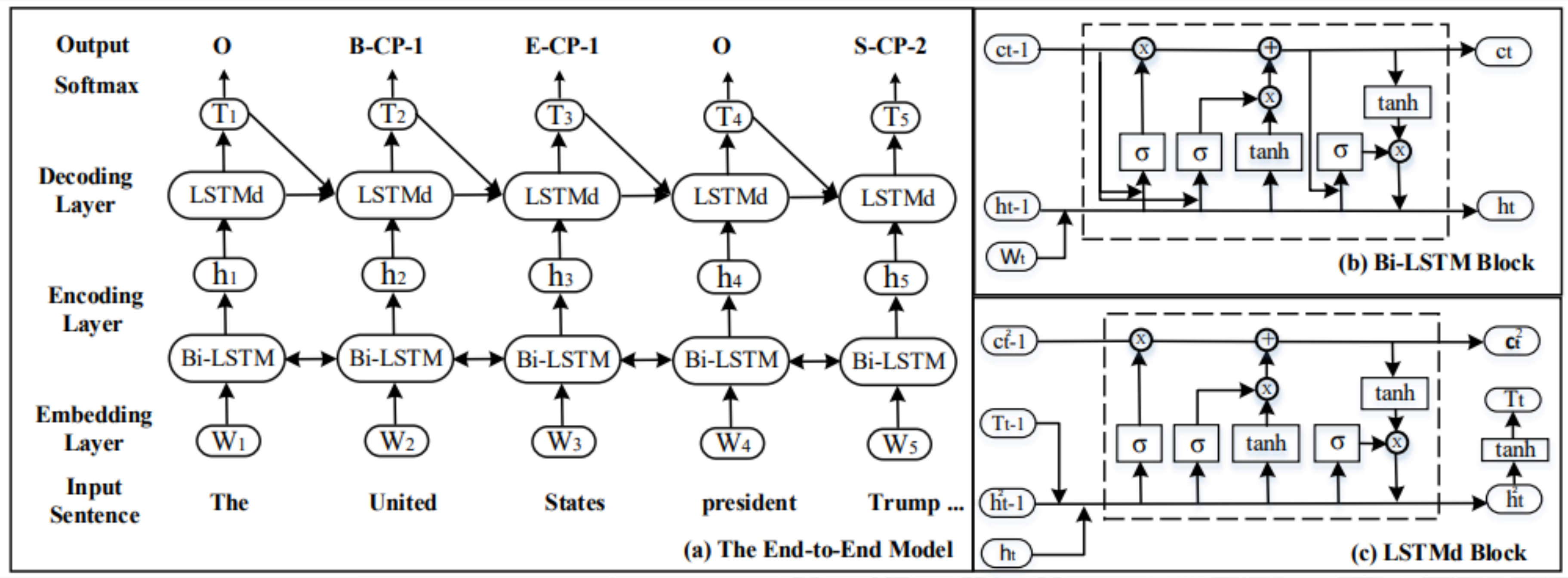
➢ Besides, if a sentence contains two or more triplets with the same relation type, we combine every two entities into a triplet based on the **nearest principle**. For example, if the relation type "Country-President" in Figure 2 is "Company-Founder", then there will be four entities in the given sentence with the same relation type. "United States" is closest to entity "Trump" and the "Apple Inc" is closest to "Jobs", so the results will be {United States, Company-Founder, Trump} and {Apple Inc, Company-Founder, Steven Paul Jobs}.

➢ In this paper, we only consider the situation where **an entity belongs to a triplet**, and we leave identification of **overlapping relations** for future work.

Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

- Model Structure



Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.

Joint Extraction

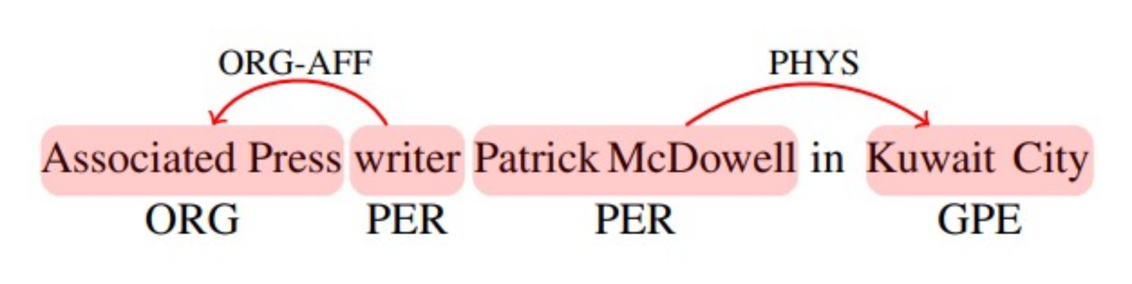
- Joint Extraction

To avoid this propagation of errors, there is a line of research which models or extracts entities and relations jointly

- 《Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme》(ACL2017)
- 《End-to-End Neural Relation Extraction with Global Optimization. 》(EMNLP2017)

Joint Extraction

- Table-filling

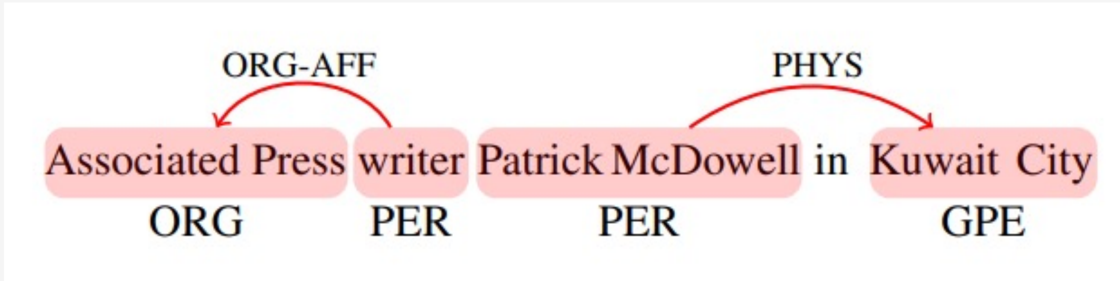


The example is chosen from the ACE05 dataset, where ORG, PER and GPE denote organization, person and geo-political entities, respectively; ORG-AFF and PHYS denote organization affiliation and physical relations, respectively.

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Table-filling



The example is chosen from the ACE05 dataset, where ORG, PER and GPE denote organization, person and geo-political entities, respectively; ORG-AFF and PHYS denote organization affiliation and physical relations, respectively.

	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 $\perp$	16 $\perp$	22 $\perp$	27 $\perp$	31 $\perp$	34 $\perp$	36 $\perp$
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 $\perp$	23 $\perp$	28 $\perp$	32 $\perp$	35 $\perp$
writer			3 U-PER	11 $\perp$	18 $\perp$	24 $\perp$	29 $\perp$	33 $\perp$
Patrick				4 B-PER	12 $\perp$	19 $\perp$	25 $\perp$	30 $\perp$
McDowell					5 L-PER	13 $\perp$	20 $\perp$	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 $\perp$	21 $\perp$
Kuwait							7 B-GPE	15 $\perp$
City								8 L-GPE

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Table-filling

	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 ⊥	16 ⊥	22 ⊥	27 ⊥	31 ⊥	34 ⊥	36 ⊥
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 ⊥	23 ⊥	28 ⊥	32 ⊥	35 ⊥
writer			3 U-PER	11 ⊥	18 ⊥	24 ⊥	29 ⊥	33 ⊥
Patrick				4 B-PER	12 ⊥	19 ⊥	25 ⊥	30 ⊥
McDowell					5 L-PER	13 ⊥	20 ⊥	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 ⊥	21 ⊥
Kuwait							7 B-GPE	15 ⊥
City								8 L-GPE

- Formally, given a sentence $w_1 w_2 \ldots w_n$, we maintain a table $T^{n \times n}$, where $T(i, j)$ denotes the relation between w_i and w_j . When $i = j$, $T(i, j)$ denotes an entity boundary label. We map entity words into labels under the BILOU (Begin, Inside, Last, Outside, Unit) scheme, assuming that there are no overlapping entities in one sentence. Only the upper triangular table is necessary for indicating the relations.

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Table-filling

	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 ⊥	16 ⊥	22 ⊥	27 ⊥	31 ⊥	34 ⊥	36 ⊥
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 ⊥	23 ⊥	28 ⊥	32 ⊥	35 ⊥
writer			3 U-PER	11 ⊥	18 ⊥	24 ⊥	29 ⊥	33 ⊥
Patrick				4 B-PER	12 ⊥	19 ⊥	25 ⊥	30 ⊥
McDowell					5 L-PER	13 ⊥	20 ⊥	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 ⊥	21 ⊥
Kuwait							7 B-GPE	15 ⊥
City								8 L-GPE

- Formally, given a sentence $w_1w_2 \ldots w_n$, we maintain a table $T^{n \times n}$, where $T(i, j)$ denotes the relation between w_i and w_j . When $i = j$, $T(i, j)$ denotes an entity boundary label. We map entity words into labels under the BILOU (Begin, Inside, Last, Outside, Unit) scheme, assuming that there are no overlapping entities in one sentence. Only the upper triangular table is necessary for indicating the relations.

- We adopt the close-first left-to-right order to map the two-dimensional table into a sequence, in order tofill the table incrementally. First $\{T(i, i)\}$ are filled by growing i, and then the sequence $\{T(i, i + 1)\}$ is filled, and then $\{T(i, i + 2)\}, \cdots, \{T(i, i + n)\}$ are filled incrementally, until the table is fully annotated.

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Table-filling

	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 ⊥	16 ⊥	22 ⊥	27 ⊥	31 ⊥	34 ⊥	36 ⊥
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 ⊥	23 ⊥	28 ⊥	32 ⊥	35 ⊥
writer			3 U-PER	11 ⊥	18 ⊥	24 ⊥	29 ⊥	33 ⊥
Patrick				4 B-PER	12 ⊥	19 ⊥	25 ⊥	30 ⊥
McDowell					5 L-PER	13 ⊥	20 ⊥	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 ⊥	21 ⊥
Kuwait							7 B-GPE	15 ⊥
City								8 L-GPE

- During the table-filling process, we take two label sets for entity detection ($i = j$) and relation classification ($i < j$), respectively. The labels for entity detection include {B-*, I-*, L-*, O, U-* }, where * denotes the entity type, and the labels for relation classification are $\{\overrightarrow{*}, \overleftarrow{*}, \perp\}$, where * denotes the relation category and ⊥ denotes a NULL relation.
- At each step, given a partially-filled table T, we determine the most suitable label l for the next step using a scoring function:

$$score(T, l) = W_l h_T$$

- where W_l is a model parameter and h_T is the vector representation of T. Based on the function, we aim to find the best label sequence $l_1 \cdots l_m$, where $m=n(n+1)/2$, and the resulting sequence of partially-filled tables is $T_0 T_1 \cdots T_m$, where $T_i = F_{ILL}(Ti-1, li)$, and T_0 is an empty table. Different from previous work, we investigate a structural model that is optimized for the label sequence $l_1 \cdots l_m$ globally, rather than for each l_i locally.

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Representation

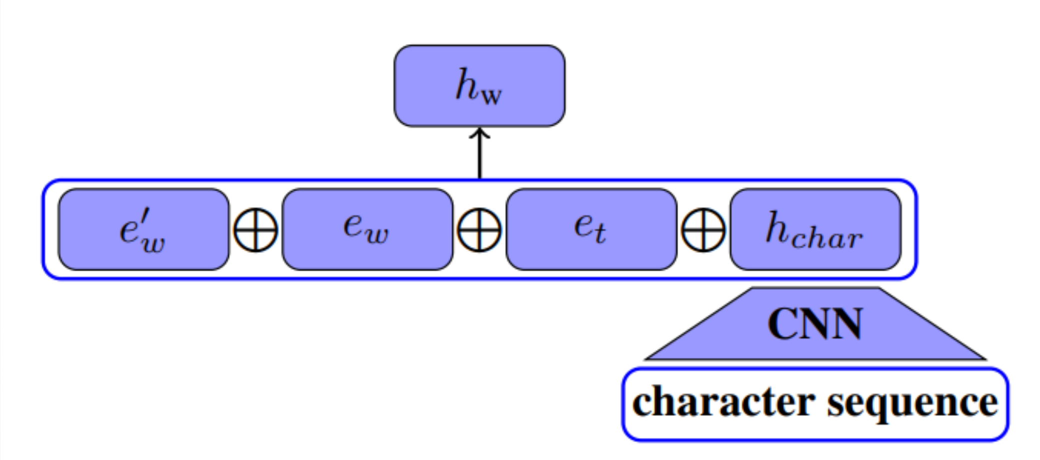
  - **Word Representation:** we represent each word w_i by a vector h_w^i using its word form, POS tag and characters. Two different forms of embeddings are used based on the word form, one being obtained by using a randomly initialized look-up table E_w, tuned during training and represented by e_w, and the other being a pre-trained external word embedding from E'_w, which is fixed and represented by e'_w. For a POS tag t, its embedding e_t is obtained from a look-up table Et similar to E_w.

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Representation

  - **Word Representation:** we represent each word w_i by a vector h_w^i using its word form, POS tag and characters. Two different forms of embeddings are used based on the word form, one being obtained by using a randomly initialized look-up table E_w, tuned during training and represented by e_w, and the other being a pre-trained external word embedding from E'_w, which is fixed and represented by e'_w. For a POS tag t, its embedding e_t is obtained from a look-up table Et similar to E_w.We further enhance the word representation by using its character sequence, taking a convolution neural network (CNN) to derive a character-based word representation h_{char}, which has been demonstrated effective for several NLP tasks.We obtain the final h_w^i based on a non-linear feed-forward layer on $e'_w \oplus e_w \oplus e_t \oplus h_{char}$, where $\oplus$ denotes concatenation



Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Representation

  - **Word Representation:** we represent each word w_i by a vector h_w^i using its word form, POS tag and characters. Two different forms of embeddings are used based on the word form, one being obtained by using a randomly initialized look-up table E_w, tuned during training and represented by e_w, and the other being a pre-trained external word embedding from E'_w, which is fixed and represented by e'_w. For a POS tag t, its embedding e_t is obtained from a look-up table Et similar to E_w.We further enhance the word representation by using its character sequence, taking a convolution neural network (CNN) to derive a character-based word representation h_{char}, which has been demonstrated effective for several NLP tasks.We obtain the final h_w^i based on a non-linear feed-forward layer on $e'_w \oplus e_w \oplus e_t \oplus h_{char}$, where $\oplus$denotes concatenation

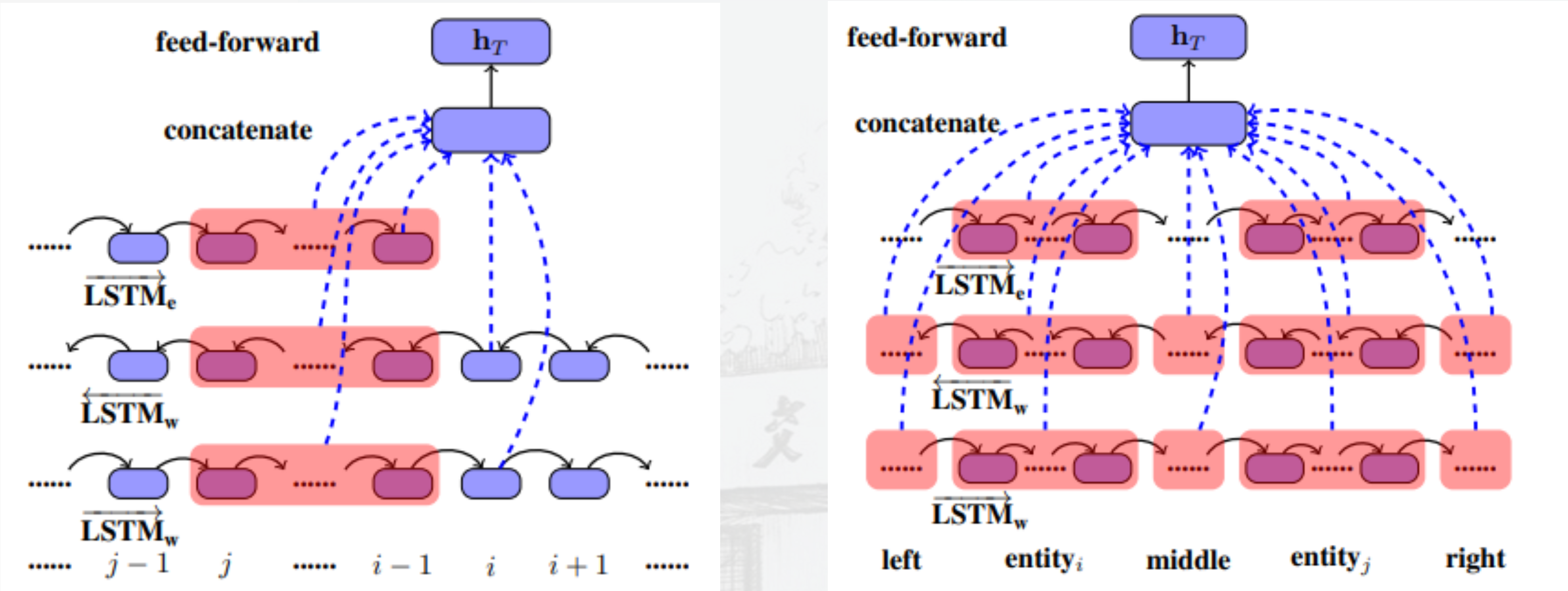
  - **Label Representation:**In addition to the word sequence, the history label sequence $l_1 l_2 \cdots l_{i-1}$, and especially the labels representing detected entities, are also useful disambiguation. For example, the previous entity boundary label can be helpful to deciding the boundary label of the current word. During relation classification, the types of the entities involved can indicate the relation category between them. We exploit the diagonal label sequence of partial table *T,* which denotes entity boundaries, to enhance the representation learning. A word's entity boundary label embedding e_l is obtained by

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Representation

  - We use separate feature representations for entity detection and relation classification, both of which are extracted from the above three LSTM structures. In particular, we first extract a set of base neural features, and then concatenate them and feed them into a non-linear neural layer for entity detection and relation classification.



(a) entity detection

(b) relation classification

Zhang M , Yue Z , Fu G . End-to-End Neural Relation Extraction with Global Optimization[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

Joint Extraction

- Joint Extraction

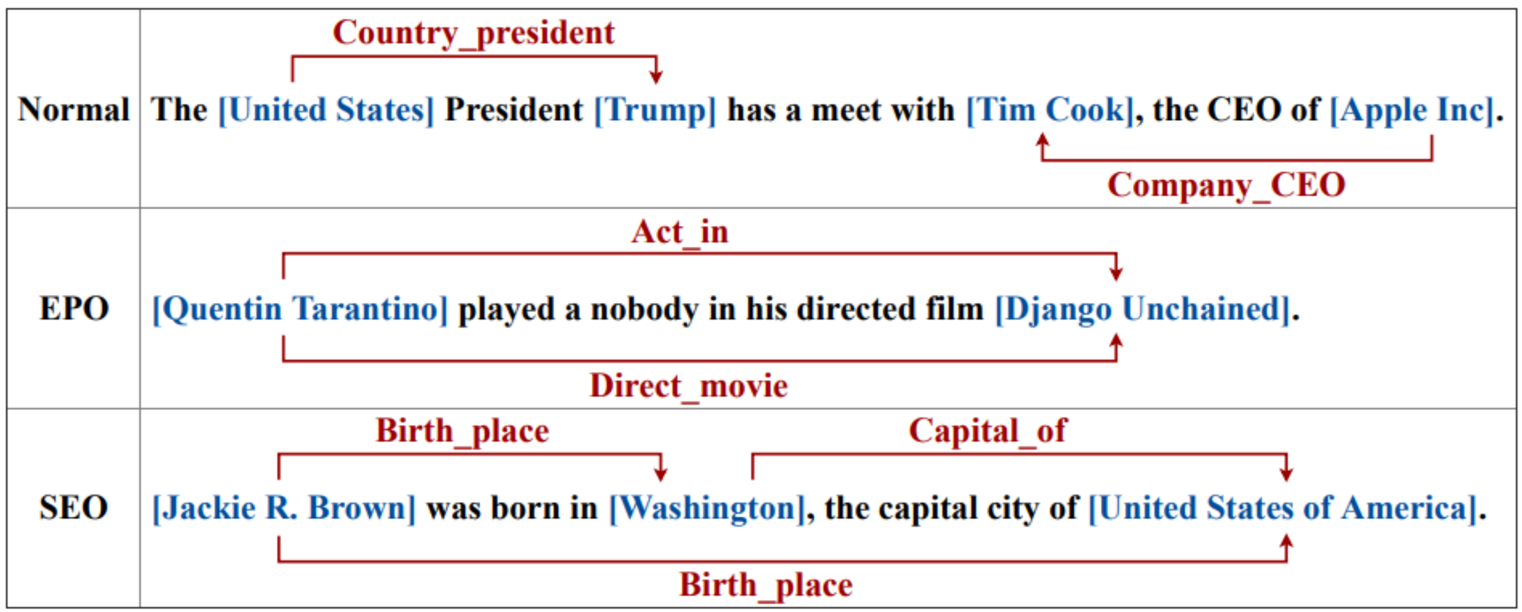
To avoid this propagation of errors, there is a line of research which models or extracts entities and relations jointly

- 《Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme》(ACL2017)
- 《End-to-End Neural Relation Extraction with Global Optimization.》(EMNLP2017)
- 《A Novel Cascade Binary Tagging Framework for Relational Triple Extraction》(ACL2020)

Joint Extraction

- Overlapping Triple

Few existing works excel in solving the overlapping triple problem where multiple relational triples in the same sentence share the same entities.



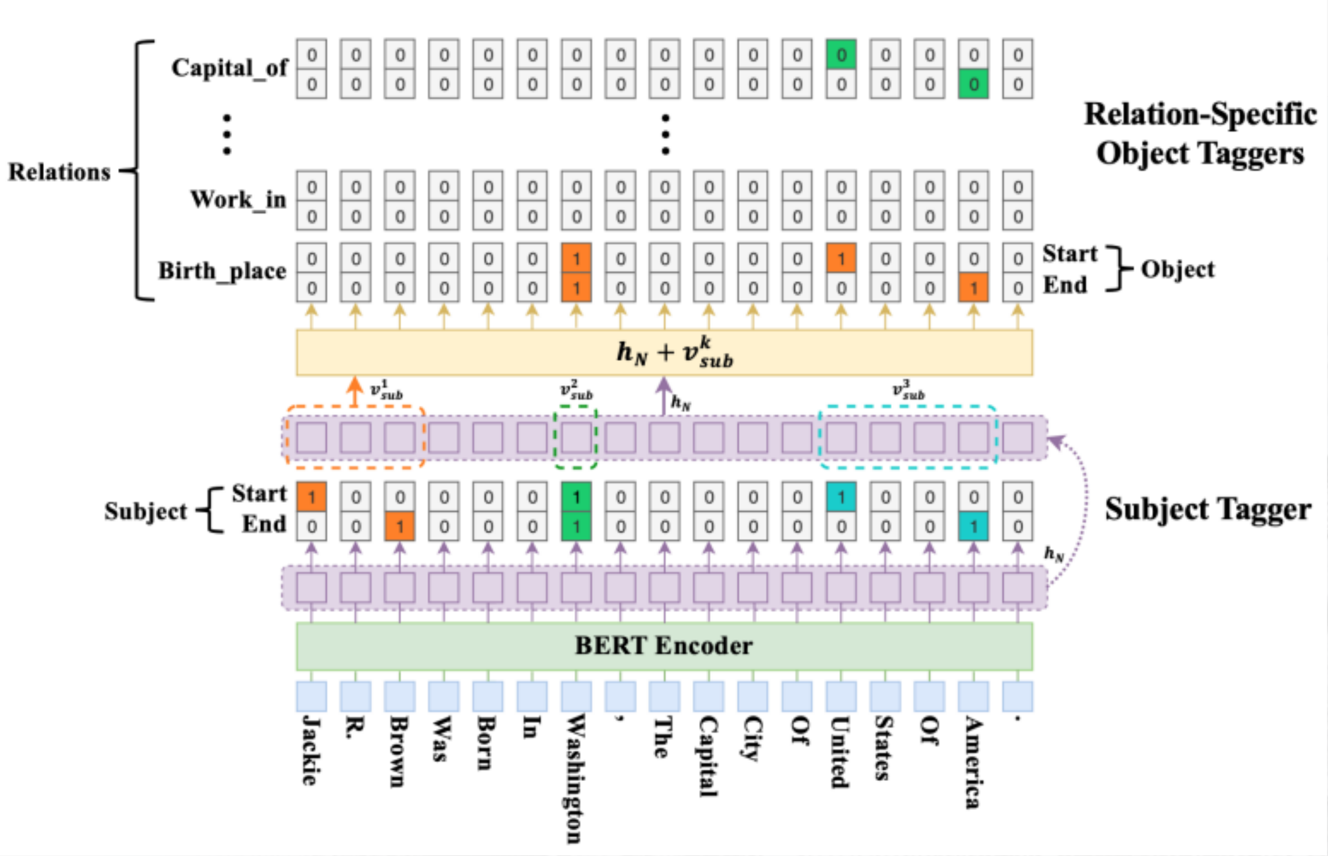
Examples of Normal, EntityPairOverlap (EPO) and SingleEntityOverlap (SEO) overlapping patterns.

Wei Z , Su J , Wang Y , et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Joint Extraction

- The CasRel Framework

Identify all entities as subjects, and for each subject traverse all relations to determine whether there is an object

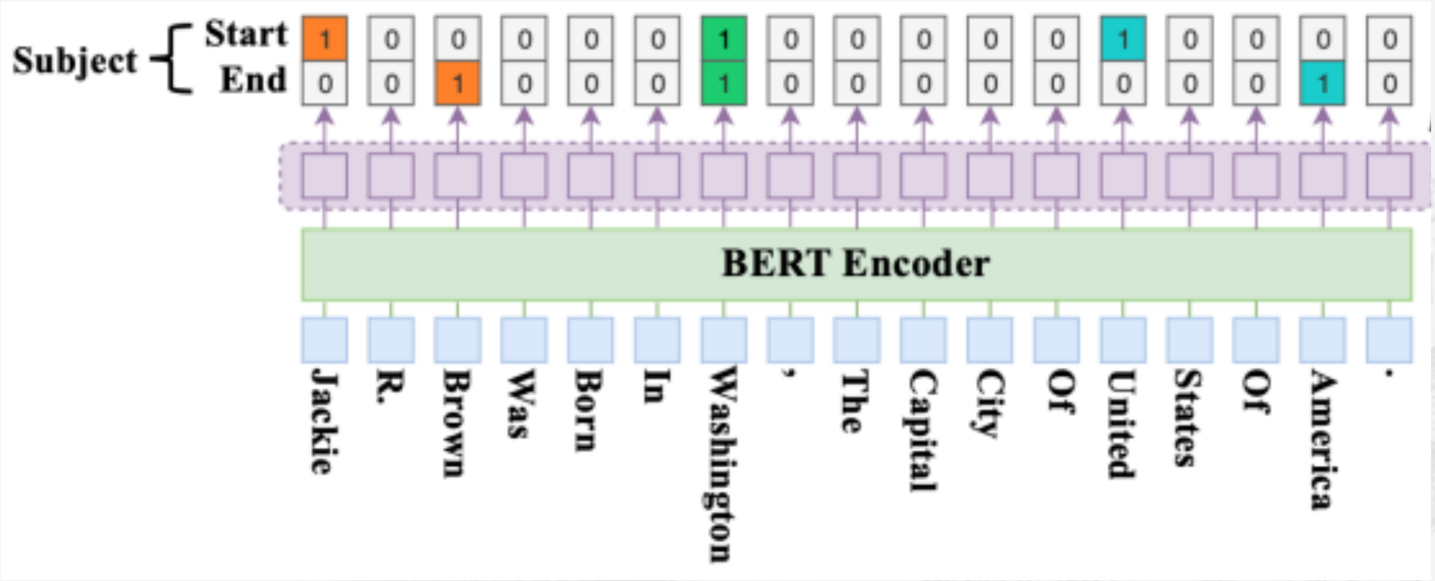


Wei Z , Su J , Wang Y , et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Joint Extraction

- The CasRel Framework

  - **BERT Encoder:** employ a pre-trained BERT model to encode the context information.

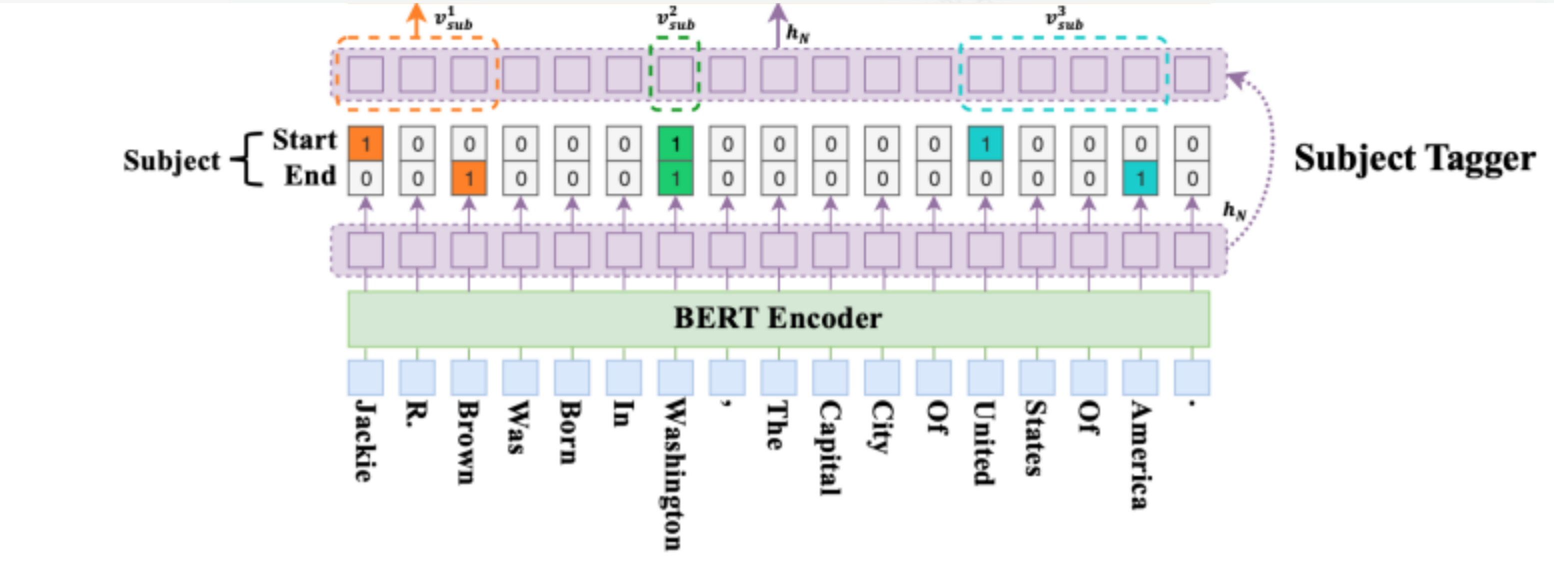


Wei Z , Su J , Wang Y , et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Joint Extraction

- The CasRel Framework

  ➢ **Subject Tagger:** The low level tagging module is designed to recognize all possible subjects in the input sentence by directly decoding the encoded vector h_N produced by the N-layer BERT encoder

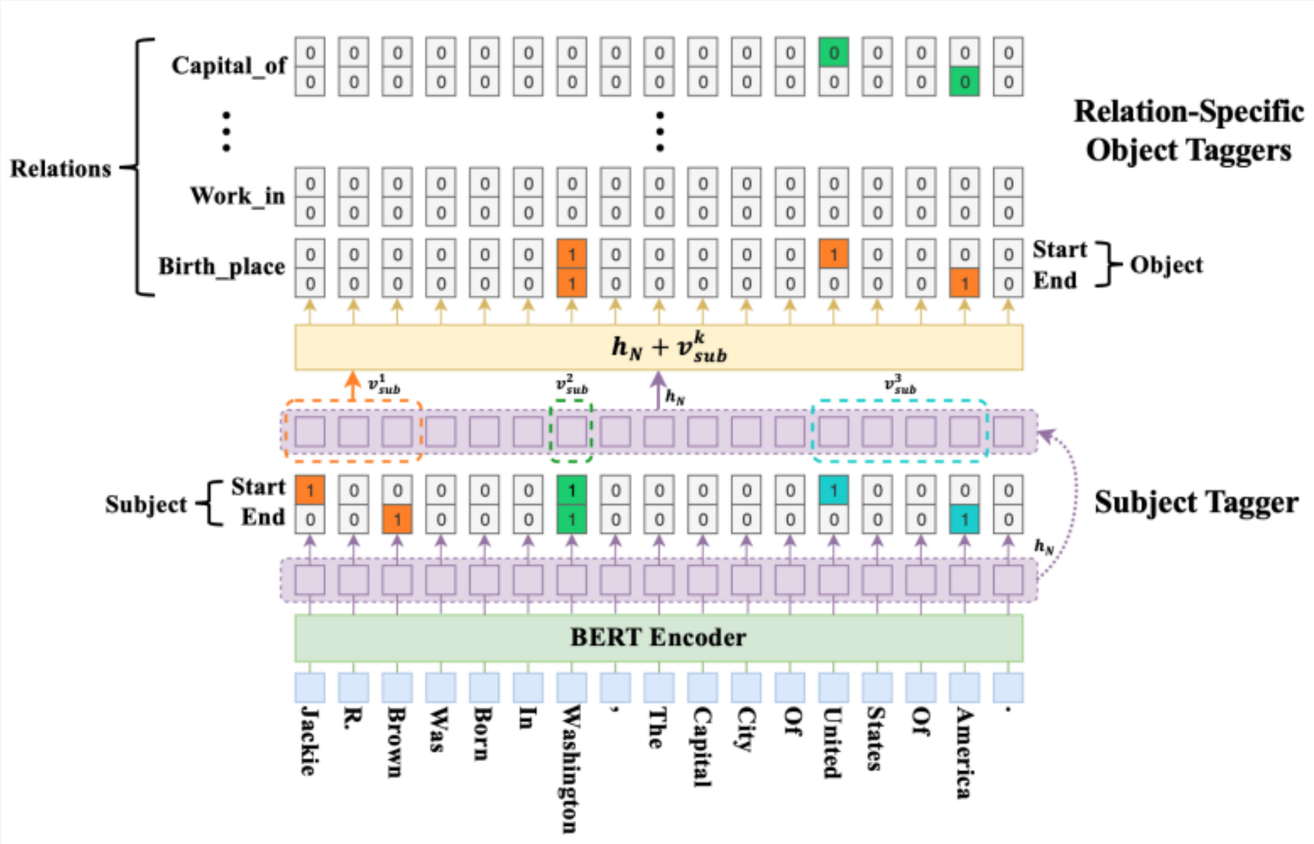


Wei Z , Su J , Wang Y , et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Joint Extraction

- The CasRel Framework

  - **Relation-Specific Object Taggers:** The high level tagging module simultaneously identifies the objects as well the involved relations with respect to the subjects obtained at lower level.



Wei Z , Su J , Wang Y , et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

Joint Extraction

- Joint Extraction

To avoid this propagation of errors, there is a line of research which models or extracts entities and relations jointly

- 《Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme》(ACL2017)
- 《End-to-End Neural Relation Extraction with Global Optimization. 》(EMNLP2017)
- 《A Novel Cascade Binary Tagging Framework for Relational Triple Extraction》(ACL2020)

Other:

[1] Fang C H , Chen Y L , Yeh M Y , et al. Multi-head Attention with Hint Mechanisms for Joint Extraction of Entity and Relation[C]// 2021.

[2] Zheng H , Wen R , Chen X , et al. PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction:, 10.18653/v1/2021.acl-long.486[P]. 2021.

[3] Shang Y M , Huang H , Mao X L . OneRel:Joint Entity and Relation Extraction with One Module in One Step[J]. 2022.

[4]Li X M, Luo X T, Dong C H Dong , et al. TDEER: An Efficient Translating Decoding Schema for Joint Extraction of Entities and Relations.[C]// In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing

[5] Huang W , Cheng X , Wang T , et al. BERT-Based Multi-Head Selection for Joint Entity-Relation Extraction[J]. 2019.

[6]Eberts M, Ulges A. Span-based joint entity and relation extraction with transformer pre-training[J]. arXiv preprint arXiv:1909.07755, 2019.

Assignment

Task:

- Implement a relation extraction method.
- Analyse the semantic relations in a domain-related corpus (e.g. PubMed) or a general corpus (e.g. Wikipedia).

Requirements：

- Complete the task independently.
- Submit a report explaining methods and results. Further analysis and perspectives are welcomed.
- Submit a couple of slides explaining your work. (A random few will be asked to present in class.)
- Try to reuse the embeddings obtained in last assignment.

Q & A