

NLP Applications II

Relation Extraction and Distant Supervision

Slides source (NLU, MacCartney and Potts, StanfordU)

Slides source (NLP, Jurafsky and Manning, StanfordU)

Slides source (ANLP, David Bamman, UC Berkeley)

Outline

- Relation Extraction
- Distant Supervision
- Matching The Blanks

Relation extraction

Unstructured

The Big Sleep is a 1946 film noir directed by Howard Hawks,^{[2][3]} the first film version of Raymond Chandler's 1939 novel of the same name. The film stars Humphrey Bogart as private detective Philip Marlowe and Lauren Bacall as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."^[4] William Faulkner, Leigh Brackett and Jules Furthman co-wrote the screenplay.

Structured

<i>subject</i>	<i>predicate</i>	<i>object</i>
The Big Sleep	directed_by	Howard Hawks
The Big Sleep	stars	Humphrey Bogart
The Big Sleep	stars	Lauren Bacall
The Big Sleep	screenplay_by	William Faulkner
The Big Sleep	screenplay_by	Leigh Brackett
The Big Sleep	screenplay_by	Jules Furthman

Relation extraction

Automated Content Extraction, “Relation Extraction Task”

- 17 sub-relation of 6 relations from 2008

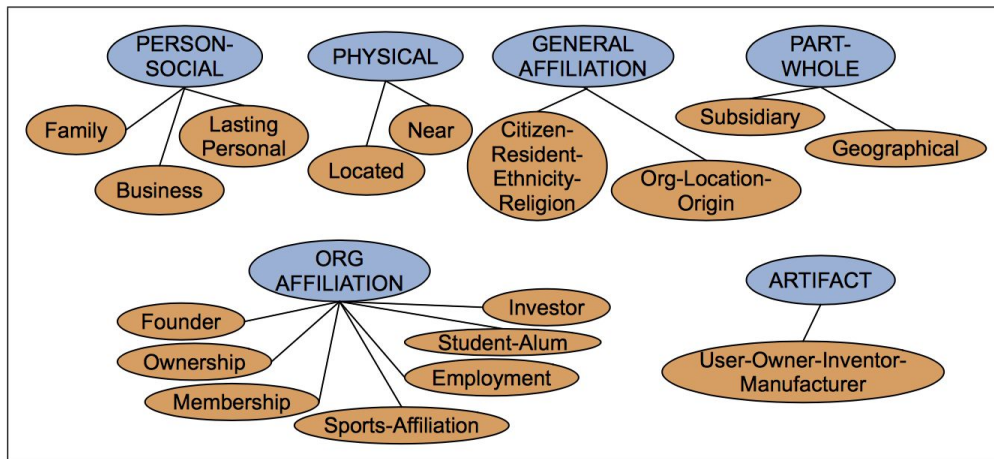


Figure 17.9 The 17 relations used in the ACE relation extraction task.

Automated Content Extraction (ACE)

- Physical-Located PER-GPE
He was in Tennessee
- Part-Whole-Subsidiary ORG-ORG
XYZ, the parent company of ABC
- Person-Social-Family PER-PER
John's wife Yoko
- Org-AFF-Founder PER-ORG
Steve Jobs, co-founder of Apple...

Relation extraction

UMLS: Unified Medical Language System

- 134 entity types, 54 relations

Entity	Relation	Entity
Injury	disrupts	Physiological Function
Bodily Location	location-of	Biologic Function
Anatomical Structure	part-of	Organism
Pharmacologic Substance	causes	Pathological Function
Pharmacologic Substance	treats	Pathologic Function

Unified Medical Language System (UMLS), SLP3

Extracting UMLS relation from a sentence

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



Echocardiography, Doppler **DIAGNOSES** Acquired stenosis

Wikipedia Infoboxes

The Big Sleep is a 1946 [film noir](#) directed by [Howard Hawks](#),^{[2][3]} the first film version of [Raymond Chandler's](#) 1939 [novel of the same name](#). The film stars [Humphrey Bogart](#) as private detective [Philip Marlowe](#) and [Lauren Bacall](#) as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."^[4] [William Faulkner](#), [Leigh Brackett](#) and [Jules Furthman](#) co-wrote the screenplay.

A [remake](#) starring [Robert Mitchum](#) as [Philip Marlowe](#) was [released in 1978](#). This was the [second film](#) in three years featuring Mitchum as Marlowe. The remake was arguably more faithful to the novel, possibly due to fewer restrictions in 1978 on what could be portrayed on screen, however, it was far less successful than the original 1946 version. In 1997, the U.S. [Library of Congress](#) deemed the film "culturally, historically, or aesthetically significant," and added it to the [National Film Registry](#).

DBPedia: 2 billion RDF
triples, from Wikipedia

Freebase extracted from
Wikipedia Infoboxes

Resource Description Framework (RDF) triples
subject - predicate - object

The Big Sleep - **DistributedBy** - Warner Bros.



Theatrical release lobby card

Directed by	Howard Hawks
Produced by	Howard Hawks
Screenplay by	William Faulkner Leigh Brackett Jules Furthman
Based on	The Big Sleep by Raymond Chandler
Starring	Humphrey Bogart Lauren Bacall
Music by	Max Steiner
Cinematography	Sidney Hickox
Edited by	Christian Nyby
Distributed by	Warner Bros.
Release date	August 23, 1946 (United States)
Running time	114 minutes (released cut) 116 minutes (re-released original cut)

Ontological relations

Examples from the WordNet thesaurus

- IS-A (hypernym): subsumption between classes
 - Giraffe IS-A ruminant IS-A ungulate IS-A mammal IS-A vertebrate IS-A animal...
- Instance-of: relation between individual and class
 - San Francisco instance-of city

How to build relation extractors

1. **Hand-written patterns**
2. **Supervised machine learning**
3. **Semi-supervised and unsupervised**
 - a. Bootstrapping (using seeds)
 - b. Distant supervision
 - c. Unsupervised learning from the web (Open IE)

Regular expressions

- Regular expressions are precise ways of extracting high-precisions relations
 - “NP₁ is a film directed by NP₂” → directed_by(NP₁, NP₂)
 - “NP₁ was the director of NP₂” → directed_by(NP₂, NP₁)

Hearst patterns

<i>pattern</i>	<i>sentence</i>
NP {, NP}* {,} (and or) other NP _H	temples, treasures, and other important civic buildings
NP _H such as {NP,}* {(or and)} NP	red algae such as Gelidium
such NP _H as {NP,}* {(or and)} NP	such authors as Herrick, Goldsmith, and Shakespeare
NP _H {,} including {NP,}* {(or and)} NP	common-law countries , including Canada and England
NP _H {,} especially {NP,}* {(or and)} NP	European countries , especially France, England, and Spain

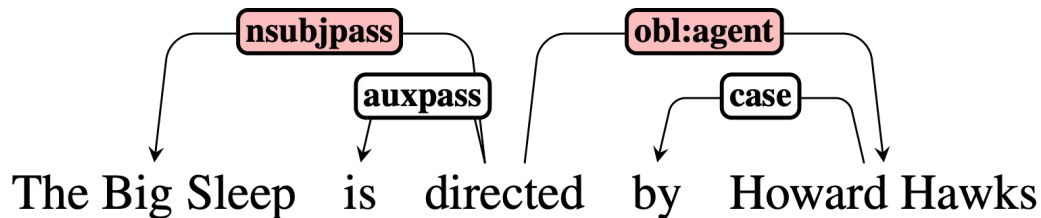
Supervised relation extraction

[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}, the first film version of Raymond Chandler's 1939 novel of the same name.

feature(m1, m2)	Feature instance
headwords of m1, m2	Sleep, Hawks
bag of words in m1, m2	The, Big, Sleep, Howard, Hawks
bag of bigrams	The Big, Big Sleep, Howard Hawks
bag of words between m1, m2	is, a, 1974, film, noir, directed, by
named entity types of m1, m2	ORG, PER
syntactic path between m1, m2	See next slide

Supervised relation extraction

[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}, the first film version of Raymond Chandler's 1939 novel of the same name.



[The Big Sleep]_{m1} \leftarrow *nsubjpass* directed \rightarrow *obl:agent* [Howard Hawks]_{m2},

m1 \leftarrow *nsubjpass* \leftarrow directed \rightarrow *obl:agent* \rightarrow m2

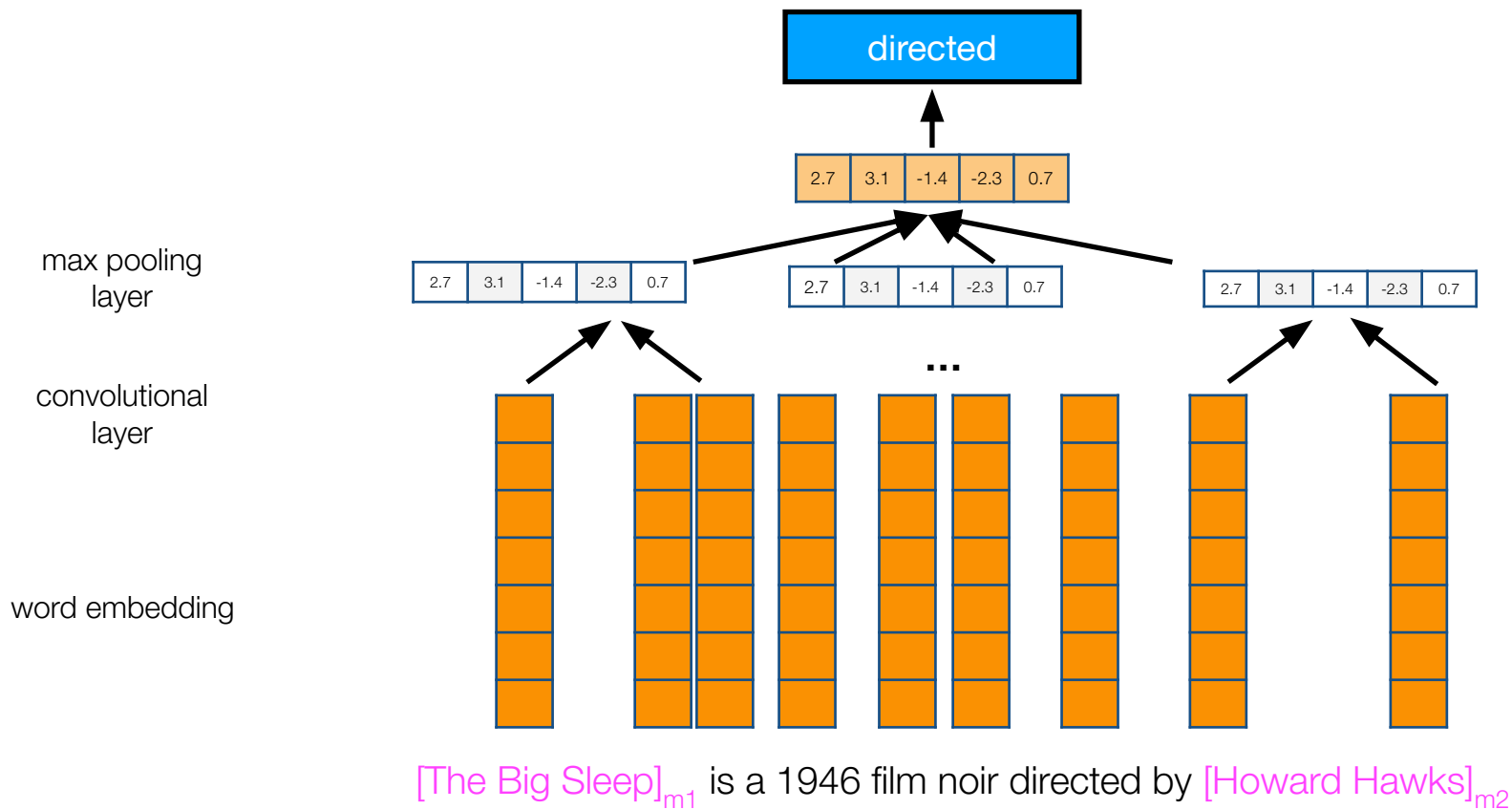
Supervised relation extraction

1. <i>George Bush traveled to France</i>	<i>George Bush</i> \leftarrow <i>traveled</i> \rightarrow <i>France</i> NSUBJ OBL
2. <i>Ahab traveled to Nantucket</i>	<i>Ahab</i> \leftarrow <i>traveled</i> \rightarrow <i>Nantucket</i> NSUBJ OBL
3. <i>George Bush will travel to France</i>	<i>George Bush</i> \leftarrow <i>travel</i> \rightarrow <i>France</i> NSUBJ OBL
4. <i>George Bush wants to travel to France</i>	<i>George Bush</i> \leftarrow <i>wants</i> \rightarrow <i>travel</i> \rightarrow <i>France</i> NSUBJ XCOMP OBL
5. <i>Ahab traveled to a city in France</i>	<i>Ahab</i> \leftarrow <i>traveled</i> \rightarrow <i>city</i> \rightarrow <i>France</i> NSUBJ OBL NMOD
6. <i>We await Ahab's visit to France</i>	<i>Ahab</i> \leftarrow <i>visit</i> \rightarrow <i>France</i> NMOD:POSS NMOD

Supervised relation extraction

```
function FINDRELATIONS(words) returns relations  
  
  relations  $\leftarrow$  nil  
  entities  $\leftarrow$  FINDENTITIES(words)  
  forall entity pairs  $\langle e1, e2 \rangle$  in entities do  
    if RELATED?(e1, e2)  
      relations  $\leftarrow$  relations + CLASSIFYRELATION(e1, e2)
```

Figure 17.13 Finding and classifying the relations among entities in a text.



We don't know which entities
we're classifying!

directed(Howard Hawks, The Big Sleep)
genre(The Big Sleep, Film Noir)
year_of_release(The Big Sleep, 1946)

Neural RE

- To solve this, we'll add positional embeddings to our representation of each word — the distance from each word w in the sentence to $m1$ and $m2$

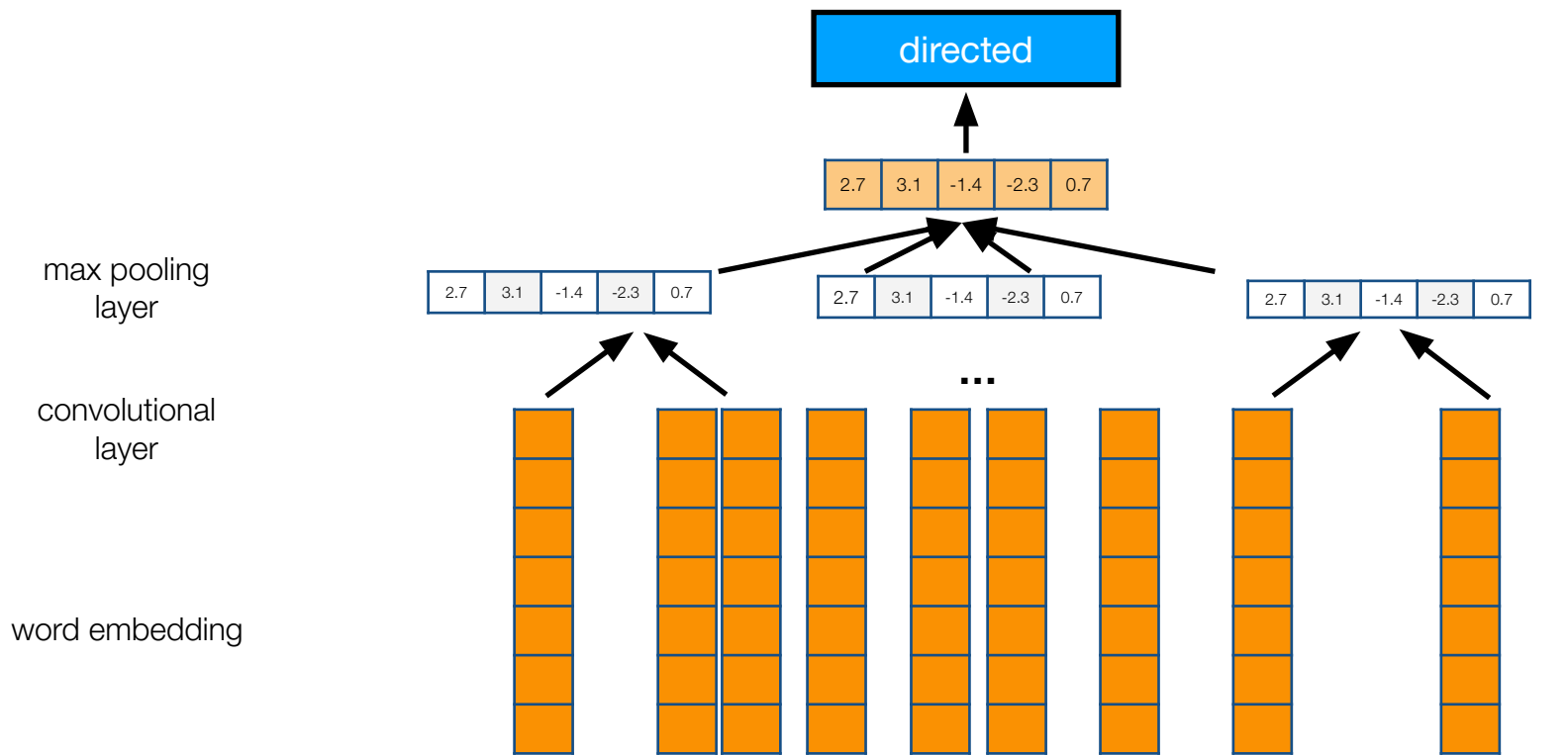
dist from m1	0	1	3	4	5	6	7	8	9
dist from m2	-8	-7	-6	-5	-4	-3	-2	-1	0
	[The Big Sleep]	is	a	1946	film	noir	directed	by	[Howard Hawks]

- 0 here uniquely identifies the head and tail of the relation; other position indicate how close the word is (maybe closer words matter more)

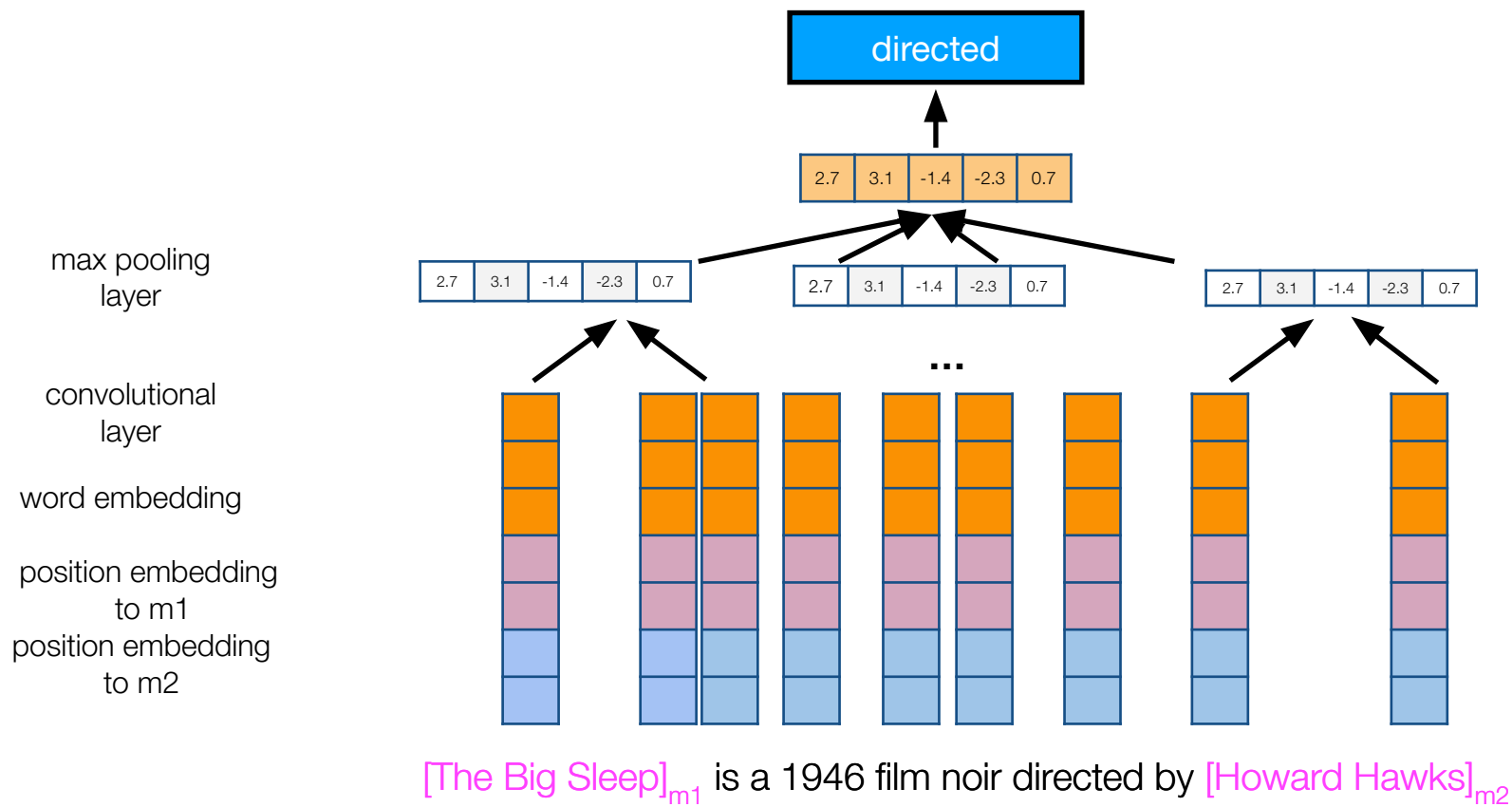
Neural RE

Each position then has an embedding

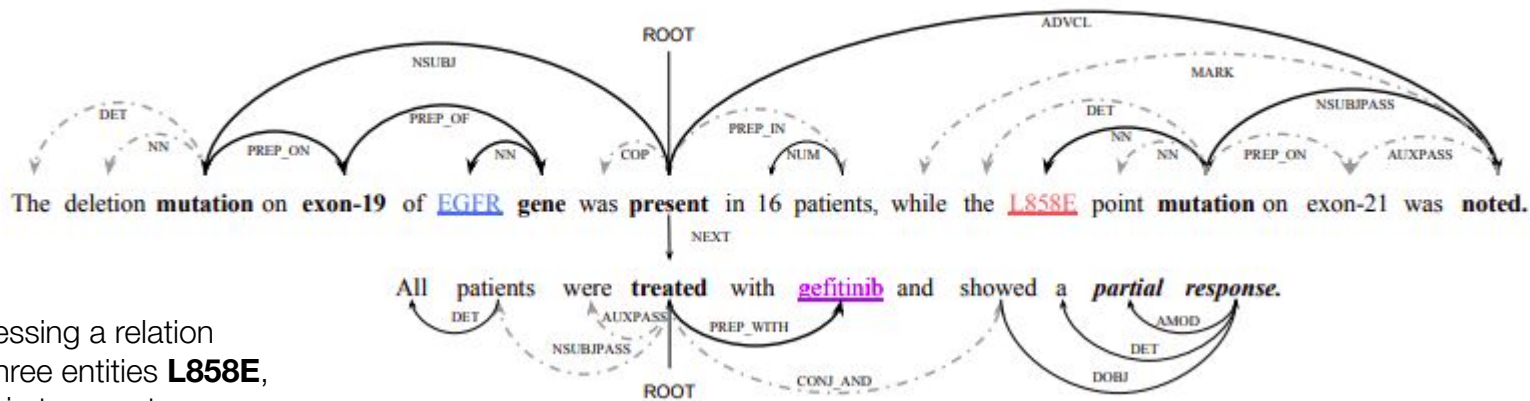
-4	2	-0.5	1.1	0.3	0.4	-0.5
-3	-1.4	0.4	-0.2	-0.9	0.5	0.9
-2	-1.1	-0.2	-0.5	0.2	-0.8	0
-1	0.7	-0.3	1.5	-0.3	-0.4	0.1
0	-0.8	1.2	1	-0.7	-1	-0.4
1	0	0.3	-0.3	-0.9	0.2	1.4
2	0.8	0.8	-0.4	-1.4	1.2	-0.9
3	1.6	0.4	-1.1	0.7	0.1	1.6
4	1.2	-0.2	1.3	-0.4	0.3	-1.0



[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}



Graph Conv Networks



Example about expressing a relation **sensitivity** among three entities **L858E**, **EGFR** and **gefitinib** in two sentences.

AGGCN (Guo et al. 2019a)

Most dependency based model decide manually what syntactic relation are useful.

Is it possible to automatically learn how to selectively attend to the relevant sub-structures?

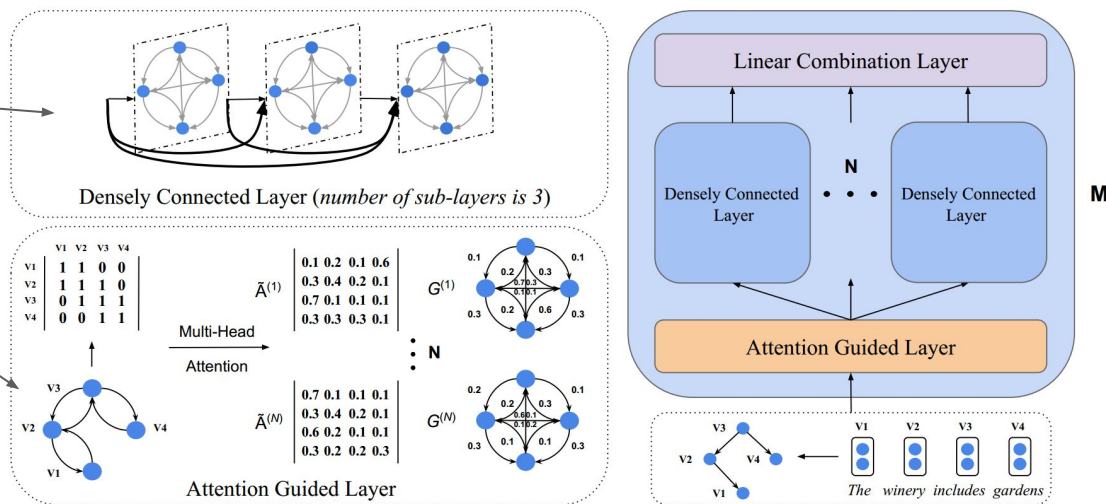
Graph Conv Networks

M identical blocks and each block has three types of layers a

The original dependency tree is transformed into N different fully connected edge-weighted graphs.

The convolution computation for node i at the lth layer:

$$\mathbf{h}_i^{(l)} = \rho \left(\sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)} + \mathbf{b}^{(l)} \right)$$

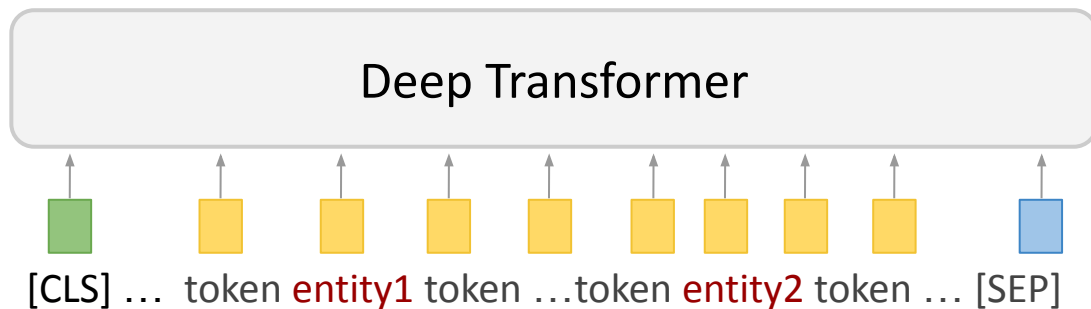


Take full dependency graph as input.

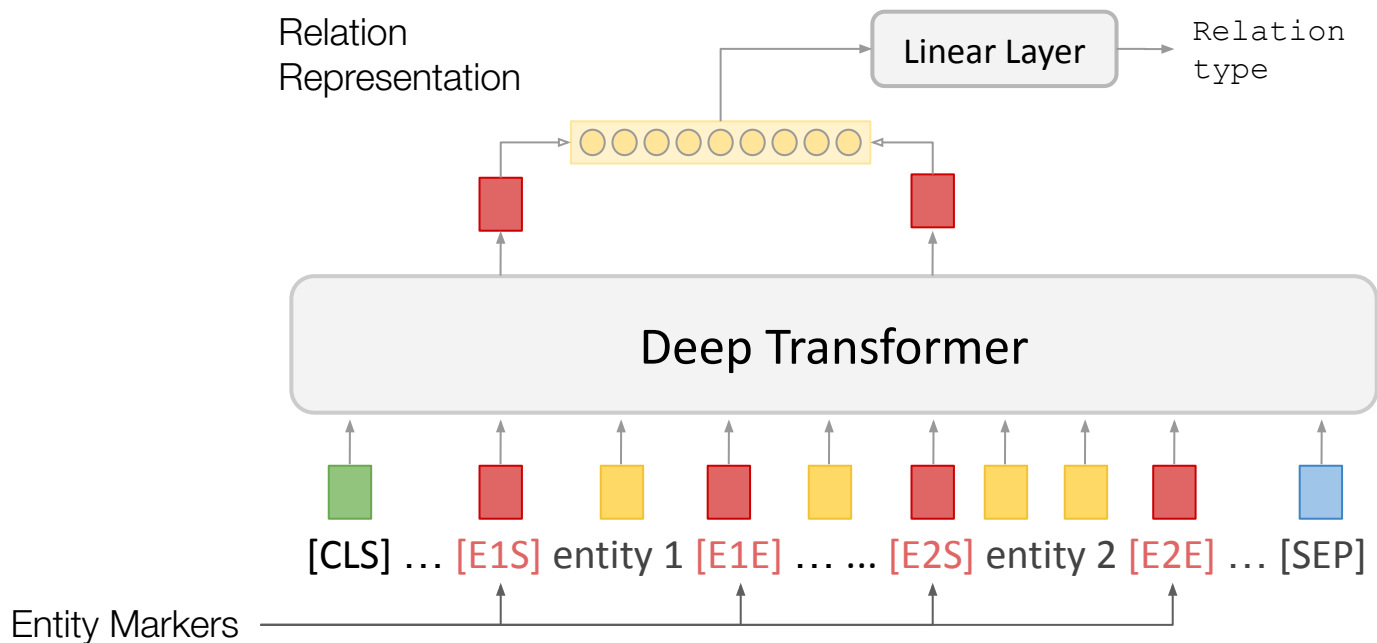
Entity Markers

Start with BERT model (deep Transformers, pretrained on masked LM)
Input is relation statement: {tokens, **e1**, **e2**}, where **e1**, **e2** encode positions.

1. How to expose **entity1** and **entity2** to the model?
2. How to extract a relation representation from the model?

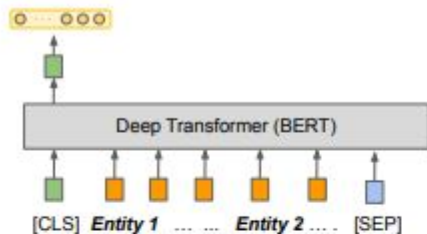


Entity Markers

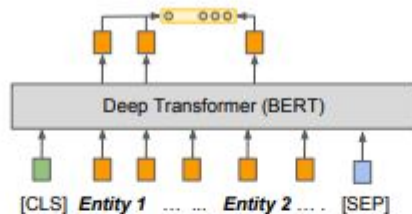


Entity Markers

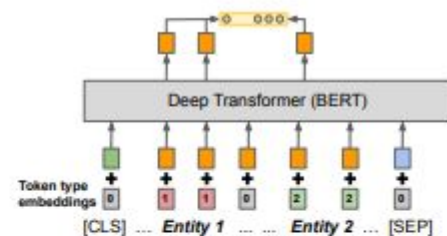
Soares et al (2019) explore variants for extracting relation representations



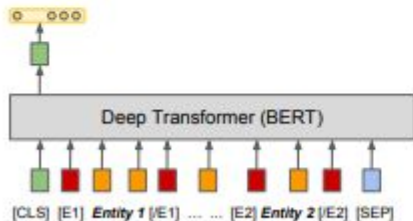
(a) STANDARD - [CLS]



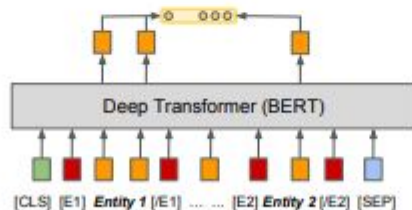
(b) STANDARD - MENTION POOLING



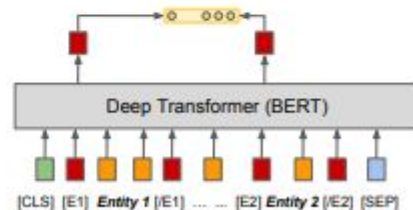
(c) POSITIONAL EMB. - MENTION POOL.



(d) ENTITY MARKERS - [CLS]



(e) ENTITY MARKERS - MENTION POOL.



(f) ENTITY MARKERS - ENTITY START

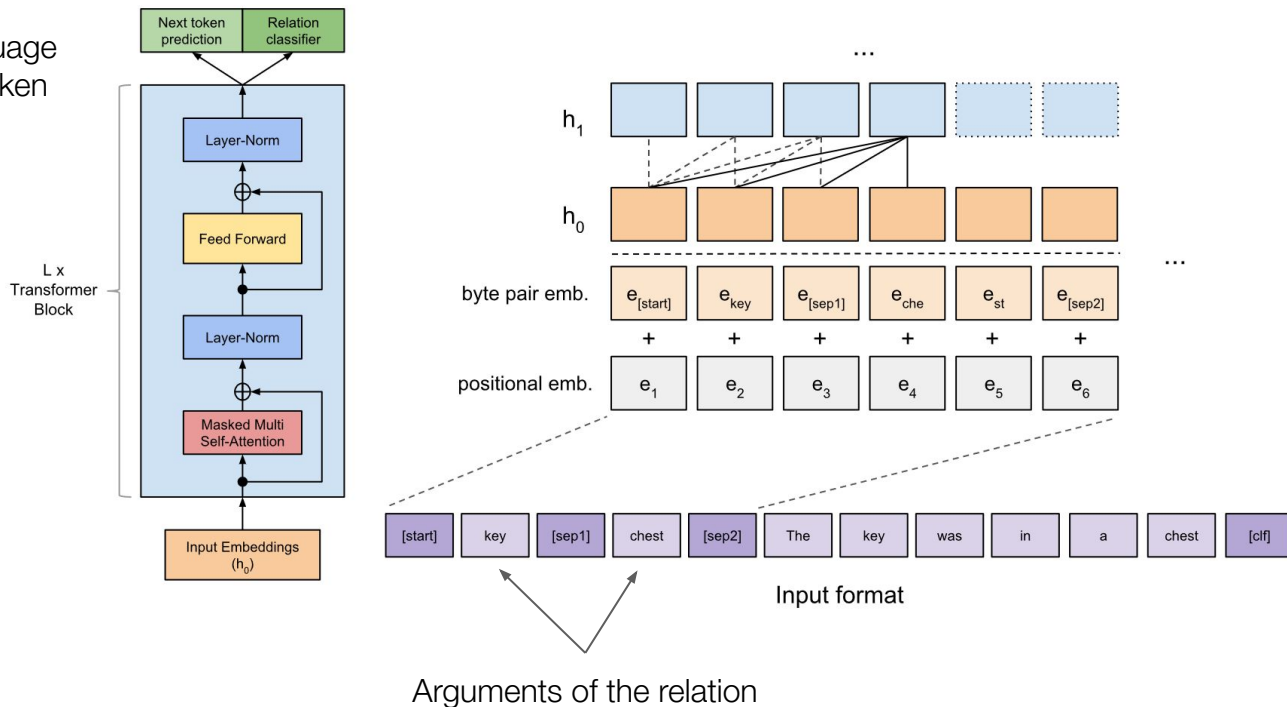
Entity Markers

	SemEval 2010 Task 8		KBP37		TACRED		FewRel 5-way-1-shot
# training annotated examples	8,000 (6,500 for dev)		15,916		68,120		44,800
# relation types	19		37		42		100
	Dev F1	Test F1	Dev F1	Test F1	Dev F1	Test F1	Dev Acc.
Wang et al. (2016)*	–	88.0	–	–	–	–	–
Zhang and Wang (2015)*	–	79.6	–	58.8	–	–	–
Bilan and Roth (2018)*	–	84.8	–	–	–	68.2	–
Han et al. (2018)	–	–	–	–	–	–	71.6

Input type	Output type							
STANDARD	[CLS]	71.6	–	41.3	–	23.4	–	85.2
STANDARD	MENTION POOL.	78.8	–	48.3	–	66.7	–	87.5
POSITIONAL EMB.	MENTION POOL.	79.1	–	32.5	–	63.9	–	87.5
ENTITY MARKERS	[CLS]	81.2	–	68.7	–	65.7	–	85.2
ENTITY MARKERS	MENTION POOL.	80.4	–	68.2	–	69.5	–	87.6
ENTITY MARKERS	ENTITY START	82.1	89.2	70	68.3	70.1	70.1	88.9

Transformer based RE (TRE)

- Based on GTP pretrained language model (autoregressive - next token prediction)
- Multitask:
 - Next token prediction
 - Relation type
- Simple input representation



Distant supervision

- It's uncommon to have labeled data in the form of <sentence, relation> pairs

<i>sentence</i>	<i>relations</i>
<p>[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}, the first film version of Raymond Chandler's 1939 novel of the same name.</p>	<p>directed_by(The Big Sleep, Howard Hawks)</p>

Distant supervision

- More common to have knowledge base data about entities and their relations that's separate from text.
- We know the text likely expresses the relations somewhere, but not **exactly where**.

Wikipedia Infoboxes

The Big Sleep is a 1946 [film noir](#) directed by [Howard Hawks](#),^{[2][3]} the first film version of [Raymond Chandler](#)'s 1939 [novel of the same name](#). The film stars [Humphrey Bogart](#) as private detective [Philip Marlowe](#) and [Lauren Bacall](#) as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."^[4] [William Faulkner](#), [Leigh Brackett](#) and [Jules Furthman](#) co-wrote the screenplay.

A [remake](#) starring [Robert Mitchum](#) as [Philip Marlowe](#) was [released in 1978](#). This was the [second film](#) in three years featuring Mitchum as Marlowe. The remake was arguably more faithful to the novel, possibly due to fewer restrictions in 1978 on what could be portrayed on screen, however, it was far less successful than the original 1946 version. In 1997, the U.S. [Library of Congress](#) deemed the film "culturally, historically, or aesthetically significant," and added it to the [National Film Registry](#).

The Big Sleep



Theatrical release lobby card

Directed by	Howard Hawks
Produced by	Howard Hawks
Screenplay by	William Faulkner Leigh Brackett Jules Furthman
Based on	<i>The Big Sleep</i> by Raymond Chandler
Starring	Humphrey Bogart Lauren Bacall
Music by	Max Steiner
Cinematography	Sidney Hickox
Edited by	Christian Nyby
Distributed by	Warner Bros.
Release date	August 23, 1946 (United States)
Running time	114 minutes (released cut) 116 minutes (re-released original cut)

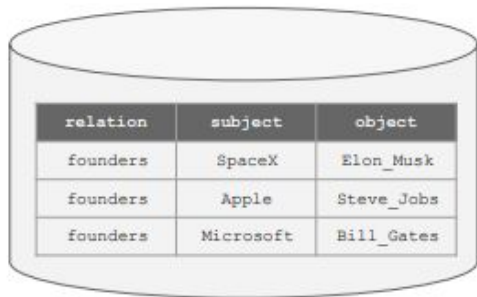
Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

Table 2: The 23 largest Freebase relations we use, with their size and an instance of each relation.

Distant supervision

Idea: Derive labels from a existing knowledge base (KB):

- Assume sentences with related entities are **positive examples**
- Assume sentences with unrelated entities are **negatives example**



relation	subject	object
founders	SpaceX	Elon_Musk
founders	Apple	Steve_Jobs
founders	Microsoft	Bill_Gates

Elon Musk, co-founder of PayPal, went on to establish SpaceX, one of the most ...



Entrepreneur Elon Musk announced the latest addition to the SpaceX arsenal ...



Elon Musk dismissed concerns that Apple was poaching the company's talent.



Now we know what Apple would have done with Elon Musk if that deal had ...



Massive quantities of training data, practically for free.
But, are those assumptions reliable?

Distant supervision

mayor(Maynard Jackson, Atlanta)

Elected mayor of Atlanta in 1973, Maynard Jackson...

Atlanta's airport will be renamed to honor Maynard Jackson, the city's first Black mayor

Born in Dallas, Texas in 1938, Maynard Holbrook Jackson, Jr. moved to Atlanta when he was 8.

mayor(Fiorello LaGuardia, New York)

Fiorello LaGuardia was Mayor of New York for three terms...

Fiorello LaGuardia, then serving on the New York City Board of Aldermen...

Distant supervision

- For feature-based models, we can represent the tuple **<m1, m2>** by aggregating together the representations from all the sentences they appear in.
- Features devised in Mintz et al 2009 are high precision - low recall
- Aggregation increases the coverage.

Feature type	Left window	NE1	Middle	NE2	Right window
Lexical	[]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[]
Lexical	[Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[.]
Lexical	[#PAD#, Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[, Missouri]
Syntactic	[]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[]
Syntactic	[Edwin Hubble ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[]
Syntactic	[Astronomer ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[]
Syntactic	[]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{lex-mod} ,]
Syntactic	[Edwin Hubble ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{lex-mod} ,]
Syntactic	[Astronomer ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{lex-mod} ,]
Syntactic	[]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{inside} Missouri]
Syntactic	[Edwin Hubble ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{inside} Missouri]
Syntactic	[Astronomer ↓ _{lex-mod}]	PER	[↑ _s was ↓ _{pred} born ↓ _{mod} in ↓ _{pcomp-n}]	LOC	[↓ _{inside} Missouri]

Table 3: Features for ‘Astronomer Edwin Hubble was born in Marshfield, Missouri’.

Distant supervision

[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}, the first film version of Raymond Chandler's 1939 novel of the same name.

[Howard Hawks]_{m2} directed the [The Big Sleep]_{m1}

feature(m1, m2)	value (e.g., normalized over all sentences)
“directed” between m1, m2	0.37
“by” between m1, m2	0.42
m1 ← <i>nsubjpass</i> ← directed → <i>obl:agent</i> → m2	0.13
m2 ← <i>nsubj</i> ← directed → <i>obj</i> → m2	0.08

Distant supervision

- Discovering Hearst patterns from distant supervision using WordNet (Snow et al. 2005)

<i>pattern</i>	<i>sentence</i>
NP _H like NP	Many hormones like leptin...
NP _H called NP	a markup language called XHTML
NP is a NP _H	Ruby is a programming language...
NP, a NP _H	IBM, a company with a long...

Distant supervision: Limitations

Powerful idea, but it has two limitations:

1. Not all sentences with related entities are truly positive
(But the benefit of more data outweighs the harm of noisier data)
2. Need an existing KB to start from - can't start from scratch.

Entrepreneur **Elon Musk** announced the latest addition to the **SpaceX** arsenal ...

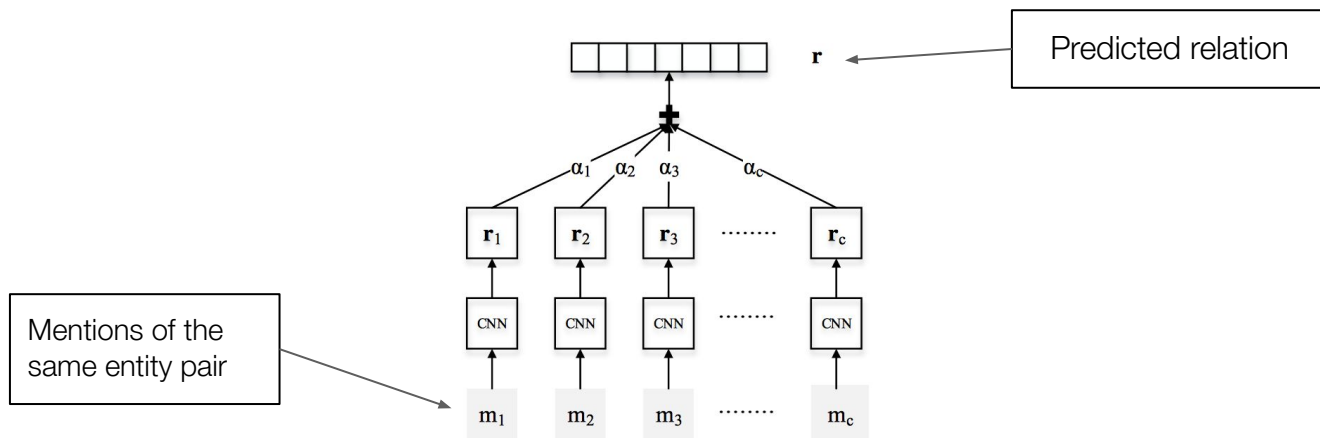


Multiple Instance Learning

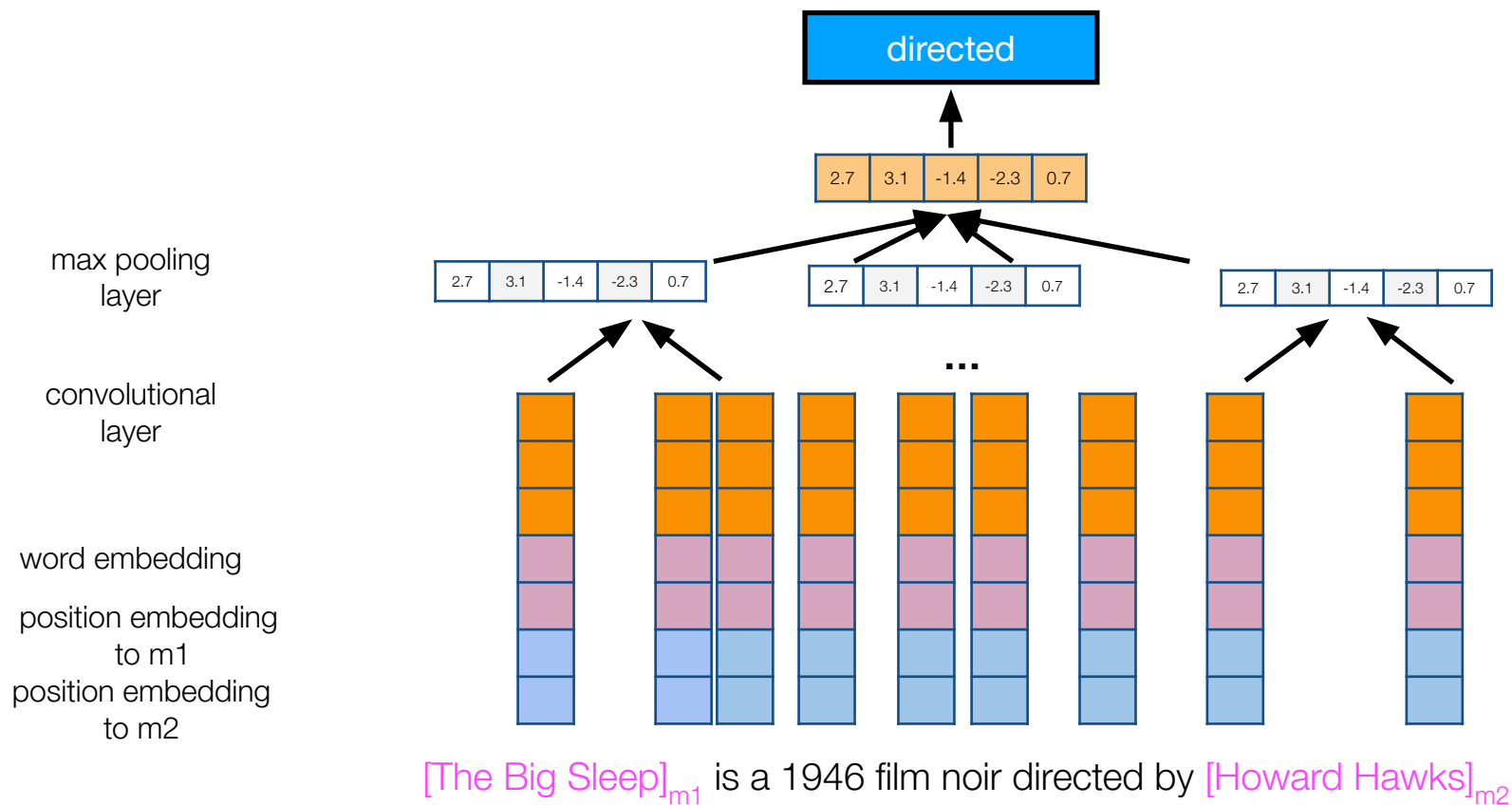
- Labels are assigned to a set of sentences, each containing the pair of entities **m1** and **m2**;
- not all of those sentences express the relation between **m1** and **m2**.

Attention

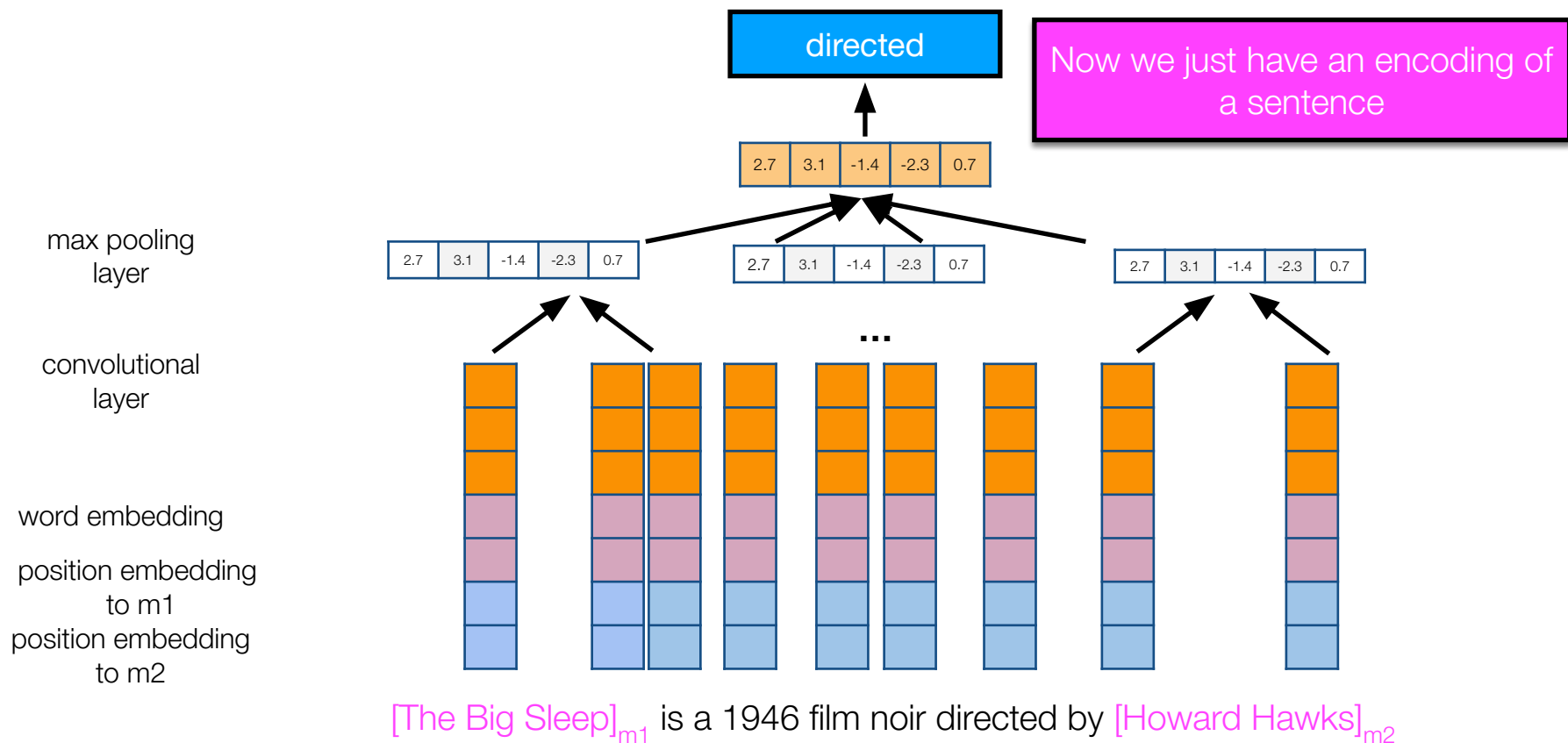
- Let's incorporate structure (and parameters) into a network that captures which **sentences** in the input we should be **attending** to (and which we can ignore).



Lin et al (2016), "Neural Relation Extraction with Selective Attention over Instances" (ACL)



Lin et al (2016), "Neural Relation Extraction with Selective Attention over Instances" (ACL)



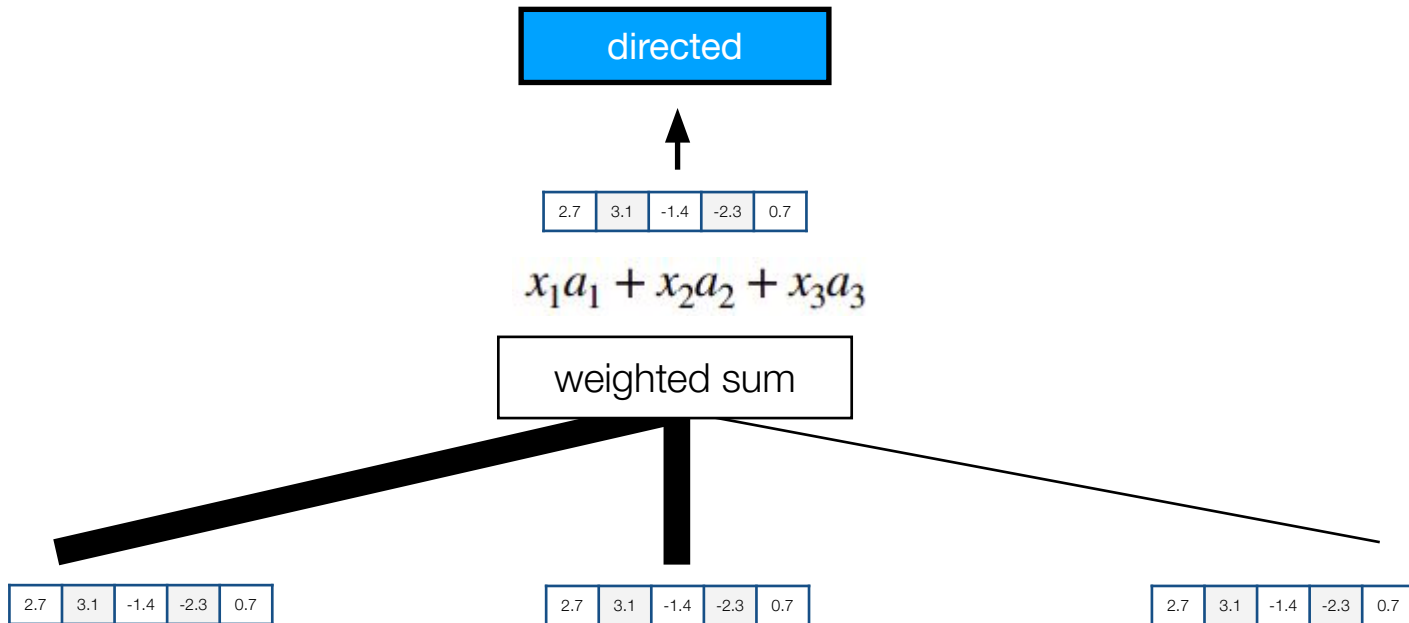
Lin et al (2016), “Neural Relation Extraction with Selective Attention over Instances” (ACL)

sentence
encoding

[The Big Sleep]_{m1} is a
1946 film noir directed by
[Howard Hawks]_{m2}

[Howard Hawks]_{m2}
directed [The Big Sleep]_{m1}

After [The Big Sleep]_{m1}
[Howard Hawks]_{m2}
married Dee Hartford



Matching the Blanks

Hypothesis: if two entities occur in two different contexts, we assume that the likelihood of they exhibit same relation is high.

Se observó actividad de CK en [BLANK] con dengue con presencia de [BLANK] como vómito, hematemesis y dolor abdominal.

(pacientes, sintomas) = (pacientes, sintomas)

(pacientes, sintomas) ≠ (pacientes, tiempo)

Al parecer, existen mecanismos comunes a ambas patologías que pueden influir en la exacerbación de los [BLANK] del asma en [BLANK] con obesidad.

El [BLANK] promedio para el inicio de ENT fue de 30 (23,5) horas, y el 88,7% de los [BLANK] alcanzaron el objetivo nutricional en 48 horas.

(pacientes, sintomas) ≠ (pacientes, tiempo)

Matching the Blanks

Automatically generate training data:

- Statement pairs r_A and r_B form a positive example since they share resolution of two entities.
- Statement pairs r_A and r_C as well as r_B and r_C form strong negative pairs since they share one entity in common but contain other non-matching entities

\mathbf{r}_A	In 1976, \mathbf{e}_1 (then of Bell Labs) published \mathbf{e}_2 , the first of his books on programming inspired by the Unix operating system.
\mathbf{r}_B	The “ \mathbf{e}_2 ” series spread the essence of “C/Unix thinking” with makeovers for Fortran and Pascal. \mathbf{e}_1 ’s Ratfor was eventually put in the public domain.
\mathbf{r}_C	\mathbf{e}_1 worked at Bell Labs alongside \mathbf{e}_3 creators Ken Thompson and Dennis Ritchie.
Mentions	\mathbf{e}_1 = Brian Kernighan, \mathbf{e}_2 = Software Tools, \mathbf{e}_3 = Unix

Learn a relation statement encoder

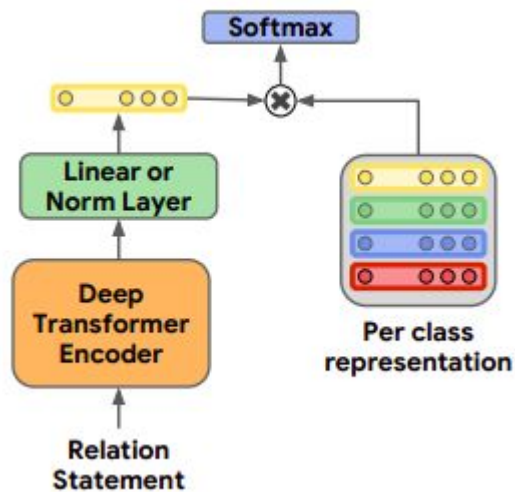
$$p(l = 1 | \mathbf{r}, \mathbf{r}') = \frac{1}{1 + \exp f_{\theta}(\mathbf{r})^{\top} f_{\theta}(\mathbf{r}')}$$

Relation encoder

(Baldini Soares et al. 2019)

Matching the Blanks

1. Once relation encoder is (pre)trained using MTB .
2. Fine-tune the encoder with Entity Maker based classifier.



(Baldini Soares et al. 2019)

Activity

- `labs/4.RE_with_transformers_and_EMs.ipynb`