Learning Sentence Representation from Weak Supervision

Yacine Jernite, Sam Bowman

Learning sentence representation
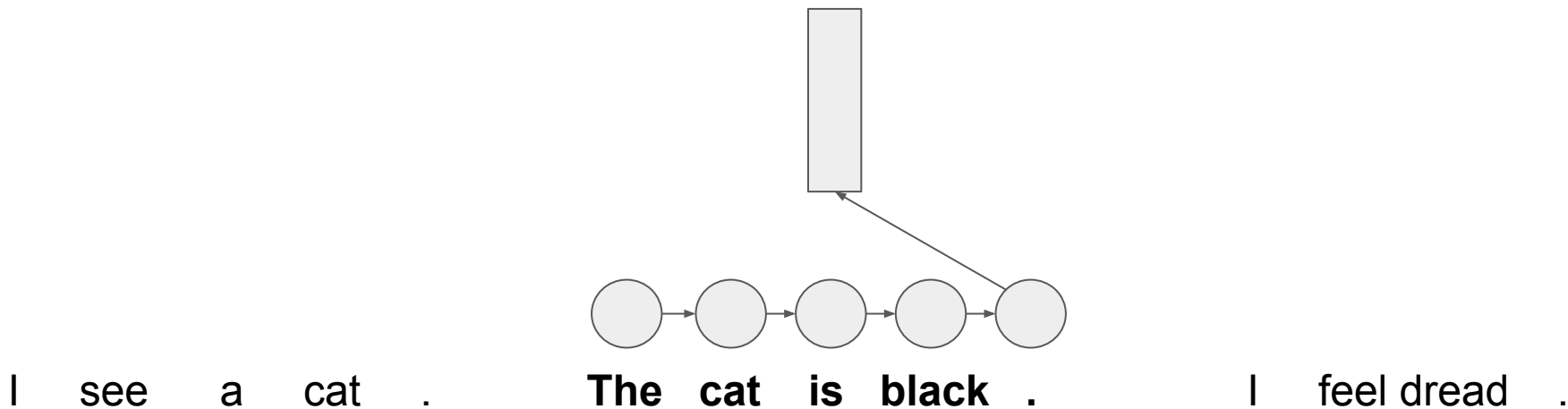
- Option 1: task-specific supervised training
    - Sentiment analysis, meh…
    - SNLI
    - Expensive training data

- Option 2: predictive / generative training

Discourse modeling

- Discourse modeling
    - Coherence and Coreference (Hobbs, 1979)
    - Relevance, pragmatics…

- Requires everything that reading comprehension does, but noisier signal.

- Tons of available data: wikipedia, BookCorpus
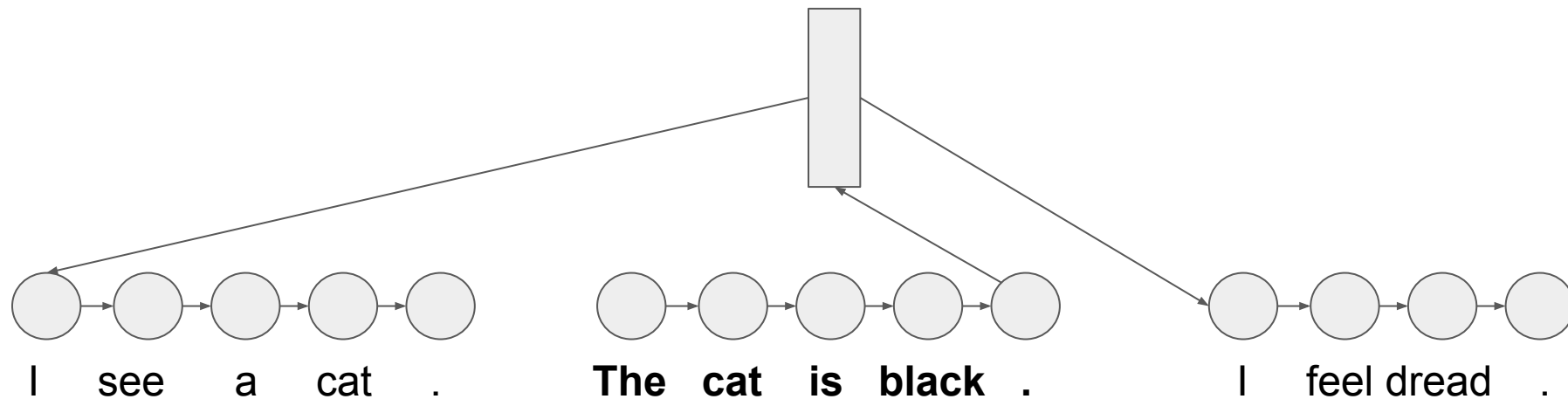
- Skip-thought type language modeling

Discourse modeling

Skip-thought model (Kiros et al., 2014)

I see a cat . **The cat is black .** I feel dread .

Discourse modeling

Skip-thought model (Kiros et al., 2014)

I see a cat .   **The cat is black .**   I feel dread .

Discourse modeling

- Generative training: expensive, focused on language modeling

- Alternative: discourse-inspired discriminative tasks
    - Discourse organization
    - Discourse relations...

Order prediction task, 40M

- Example 1: Book ()

    - I'll bet I've mentioned it in previous offerings . ////\\\\ Well , I've mentioned it again .

- Example 2: Wiki: Psychological contract ()

    - Responses may occur in the form of reduced loyalty , commitment , and organizational citizenship behaviors . ////\\\\ Employees or employers who perceive a breach are likely to respond negatively .

- Example 3: Book ()

    - That was all tin , we used to sit in there every opportunity we had when it rained , just listening . ////\\\\ Remember the shed when we were young ?

Order prediction task, 40M

- Example 1: Book (1)

  - I'll bet I've mentioned it in previous offerings . ////\\\\ Well , I've mentioned it again .

- Example 2: Wiki: Psychological contract ()

  - Responses may occur in the form of reduced loyalty , commitment , and organizational citizenship behaviors . ////\\\\ Employees or employers who perceive a breach are likely to respond negatively .

- Example 3: Book ()

  - That was all tin , we used to sit in there every opportunity we had when it rained , just listening . ////\\\\ Remember the shed when we were young ?

Order prediction task, 40M

- Example 1: Book (1)

  - I'll bet I've mentioned it in previous offerings . ////\\\\ Well , I've mentioned it again .

- Example 2: Wiki: Psychological contract (0)

  - Responses may occur in the form of reduced loyalty , commitment , and organizational citizenship behaviors . ////\\\\ Employees or employers who perceive a breach are likely to respond negatively .

- Example 3: Book ()

  - That was all tin , we used to sit in there every opportunity we had when it rained , just listening . ////\\\\ Remember the shed when we were young ?

Order prediction task, 40M

- Example 1: Book (1)

  - I'll bet I've mentioned it in previous offerings . ////\\\\ Well , I've mentioned it again .

- Example 2: Wiki: Psychological contract (0)

  - Responses may occur in the form of reduced loyalty , commitment , and organizational citizenship behaviors . ////\\\\ Employees or employers who perceive a breach are likely to respond negatively .

- Example 3: Book (0)

  - That was all tin , we used to sit in there every opportunity we had when it rained , just listening . ////\\\\ Remember the shed when we were young ?

Next sentence prediction task: 600K

- Context: Homer 's Iliad is the most famous narrative of Achilles ' deeds in the Trojan War . Achilles ' wrath is the central theme of the poem . The Homeric epic only covers a few weeks of the decade-long war , and does not narrate Achilles ' death .

Agamemnon refuses and Apollo sends a plague amongst the Greeks .
The prophet Calchas correctly determines the source of the troubles but will not speak unless Achilles vows to protect him .
Achilles does so and Calchas declares Chryseis must be returned to her father .
It begins with Achilles ' withdrawal from battle after he is dishonored by Agamemnon , the commander of the Achaean forces .
Agamemnon had taken a woman named Chryseis as his slave .

Next sentence prediction task: 600K

- Context: Homer 's Iliad is the most famous narrative of Achilles ' deeds in the Trojan War . Achilles ' wrath is the central theme of the poem . The Homeric epic only covers a few weeks of the decade-long war , and does not narrate Achilles ' death .

Agamemnon refuses and Apollo sends a plague amongst the Greeks .
The prophet Calchas correctly determines the source of the troubles but will not speak unless Achilles vows to protect him .
Achilles does so and Calchas declares Chryseis must be returned to her father .
It begins with Achilles ' withdrawal from battle after he is dishonored by Agamemnon , the commander of the Achaean forces .
Agamemnon had taken a woman named Chryseis as his slave .

Conjunction prediction task: 400K

- Example 1: Wikipedia: Anarchism

Anarchists participated alongside the Bolsheviks in both February and October revolutions , and were initially enthusiastic about the Bolshevik revolution .	?
Following a political falling out with the Bolsheviks by the anarchists and other left-wing opposition , the conflict culminated in the 1921 Kronstadt rebellion .	

- Example 2: Wikipedia: Alabama

They held on to old representation to maintain political and economic power in agricultural areas .	?
The state legislature gerrymandered the few Birmingham legislative seats to ensure election by persons living outside Birmingham .	

Conjunction prediction task: 400K

- Example 1: Wikipedia: Anarchism

Anarchists participated alongside the Bolsheviks in both February and October revolutions , and were initially enthusiastic about the Bolshevik revolution .	However
Following a political falling out with the Bolsheviks by the anarchists and other left-wing opposition , the conflict culminated in the 1921 Kronstadt rebellion .	

- Example 2: Wikipedia: Alabama

They held on to old representation to maintain political and economic power in agricultural areas .	In addition
The state legislature gerrymandered the few Birmingham legislative seats to ensure election by persons living outside Birmingham .	

Conjunction prediction task

however	62803		therefore	7717		furthermore	2797
instead	43232		nevertheless	7020		thus	2572
in fact	40751		meanwhile	6753		rather	2474
then	39831		otoh	6592		nonetheless	2452
finally	30361		otherwise	6155		moreover	2366
also	13226		that is	4974		likewise	1031
now	10493		besides	4159		further	955
still	10300		next	3580		consequently	937
again	9385		in addition	3496		hence	765
indeed	8836		for example	3152		specifically	698

Title prediction task (working)

Paragraph 1: The state has invested in aerospace , education , health care , banking , and various heavy industries , including automobile manufacturing , mineral extraction , steel production and fabrication . ///\\\ By 2006 , crop and animal production in Alabama was valued at $ 1.5 billion . ///\\\ In contrast to the primarily agricultural economy of the previous century , this was only about 1 % of the state 's gross domestic product . ///\\\ The number of private farms has declined at a steady rate since the 1960s , as land has been sold to developers , timber companies , and large farming conglomerates .

Paragraph 2: The state 's two U.S. senators are Jefferson B . ///\\\ Sessions III and Richard C. Shelby , both Republicans . ///\\\ Shelby was originally elected to the Senate as a Democrat in 1986 and re-elected in 1992 , but switched parties immediately following the November 1994 general election .

Titles: Health . ////\\\\ Economy . ////\\\\ Largest employers . ////\\\\ Agriculture . ////\\\\ Healthcare . ////\\\\ State elections . ////\\\\ Local elections . ////\\\\ Federal elections . ////\\\\ Climate . ////\\\\ Flora and fauna .
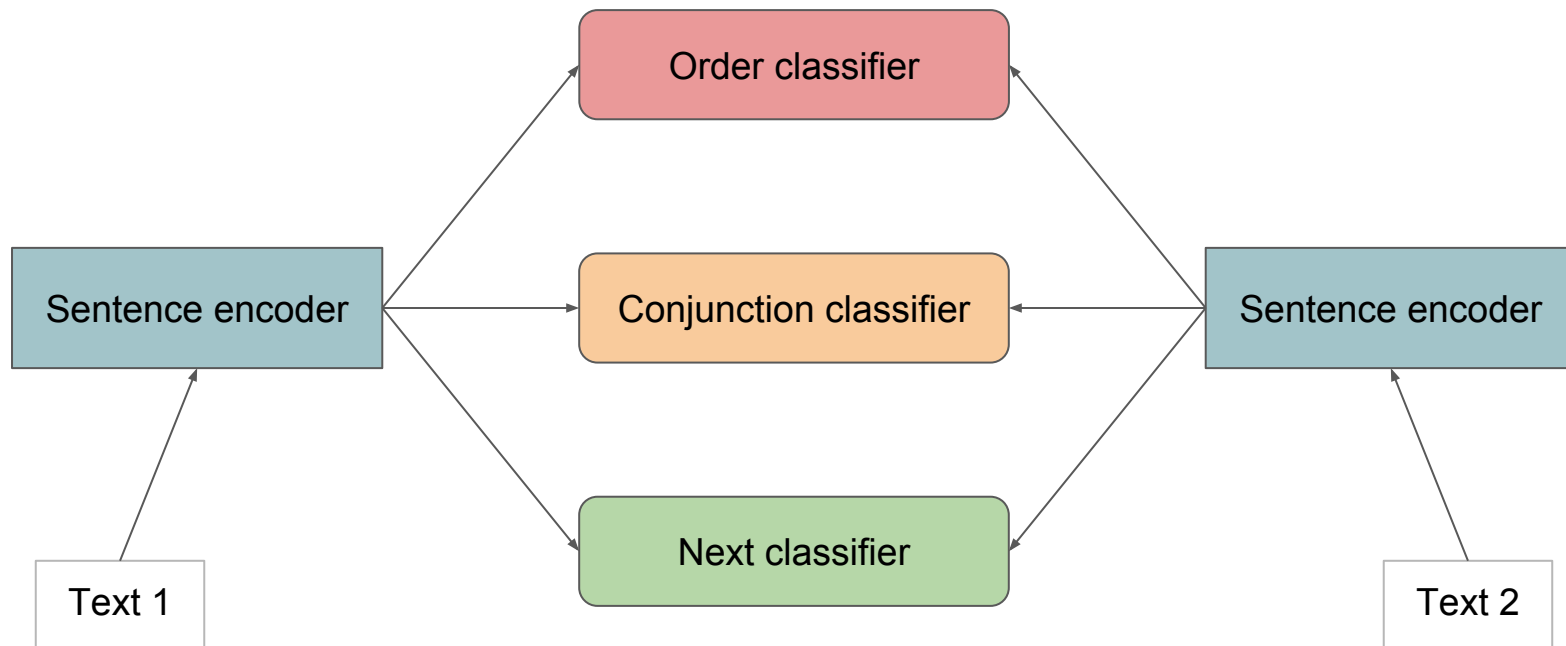
Title prediction task (working)

Paragraph 1: The state has invested in aerospace , education , health care , banking , and various heavy industries , including automobile manufacturing , mineral extraction , steel production and fabrication . ///\\\ By 2006 , crop and animal production in Alabama was valued at $ 1.5 billion . ///\\\ In contrast to the primarily agricultural economy of the previous century , this was only about 1 % of the state 's gross domestic product . ///\\\ The number of private farms has declined at a steady rate since the 1960s , as land has been sold to developers , timber companies , and large farming conglomerates .
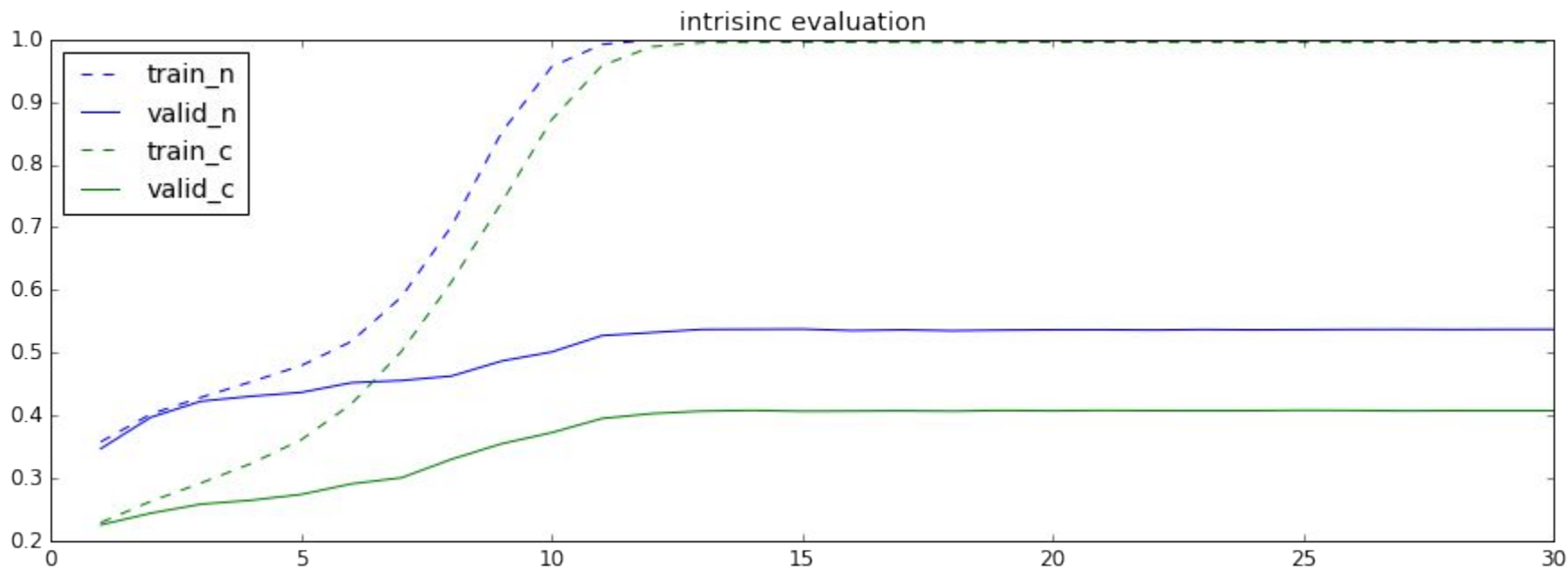
Paragraph 2: The state 's two U.S. senators are Jefferson B . ///\\\ Sessions III and Richard C. Shelby , both Republicans . ///\\\ Shelby was originally elected to the Senate as a Democrat in 1986 and re-elected in 1992 , but switched parties immediately following the November 1994 general election .

Titles: Health . ////\\\\ **Economy** . ////\\\\ Largest employers . ////\\\\ Agriculture . ////\\\\ Healthcare . ////\\\\ State elections . ////\\\\ Local elections . ////\\\\ **Federal elections** . ////\\\\ Climate . ////\\\\ Flora and fauna .
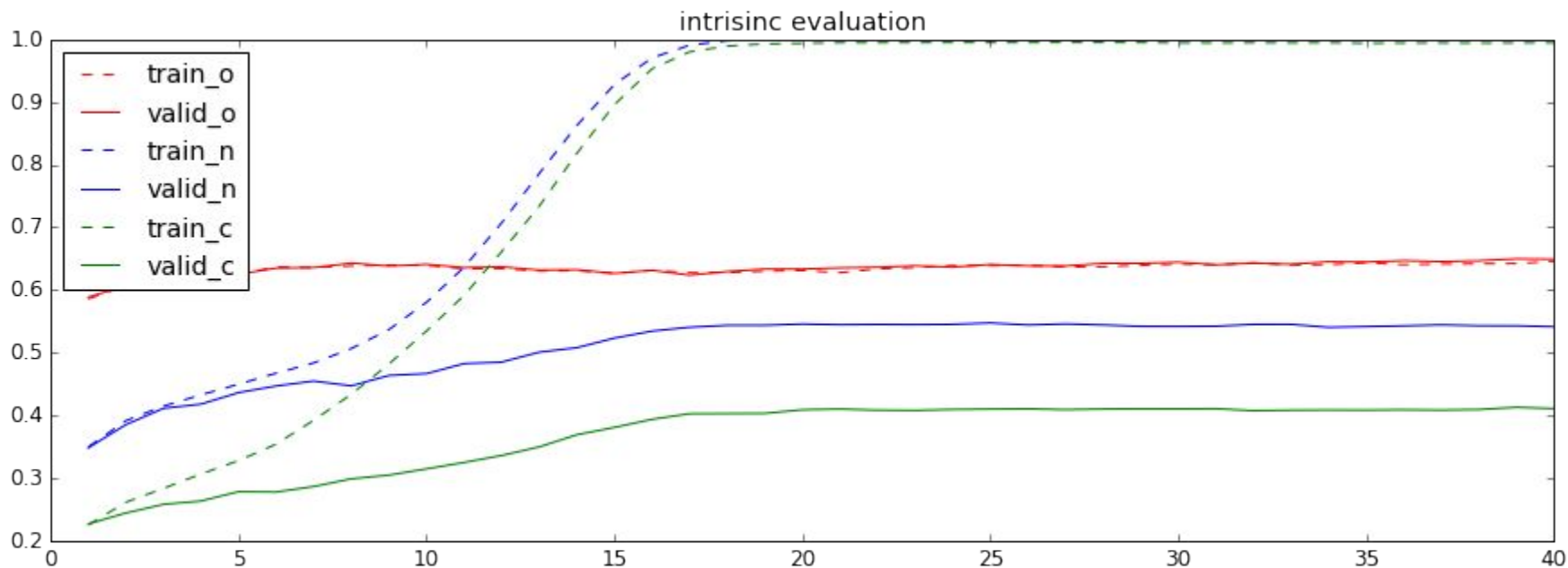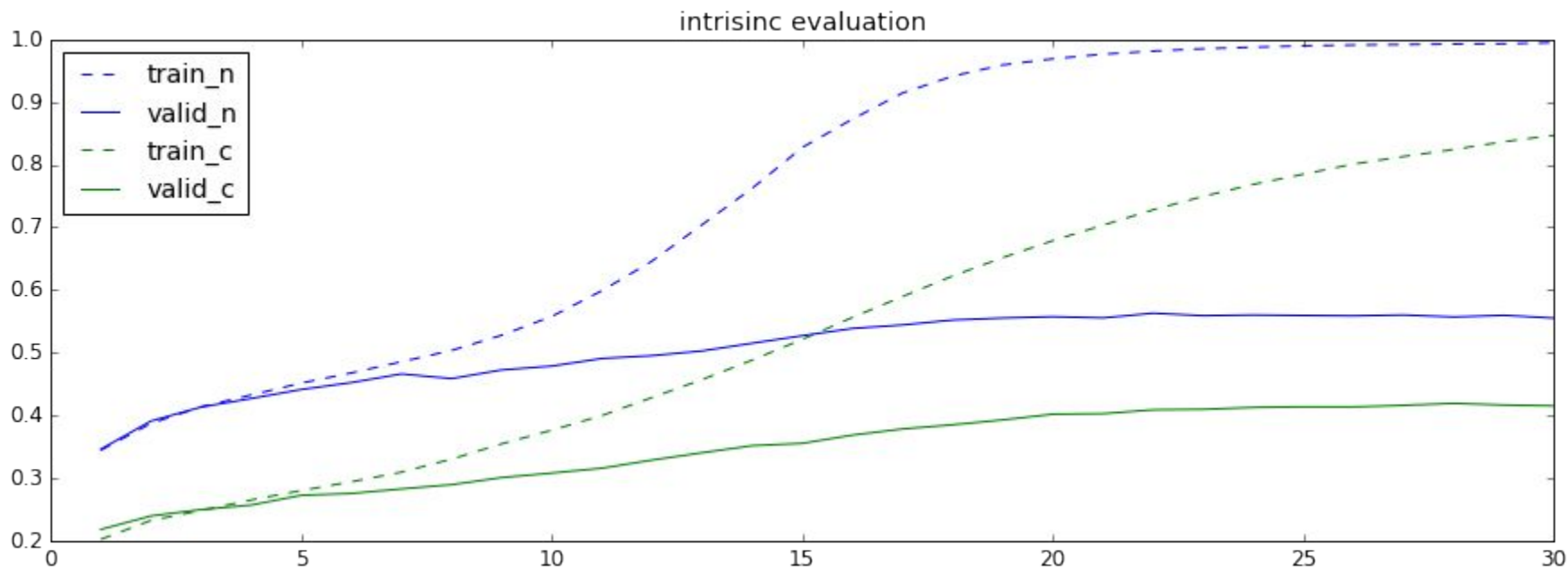
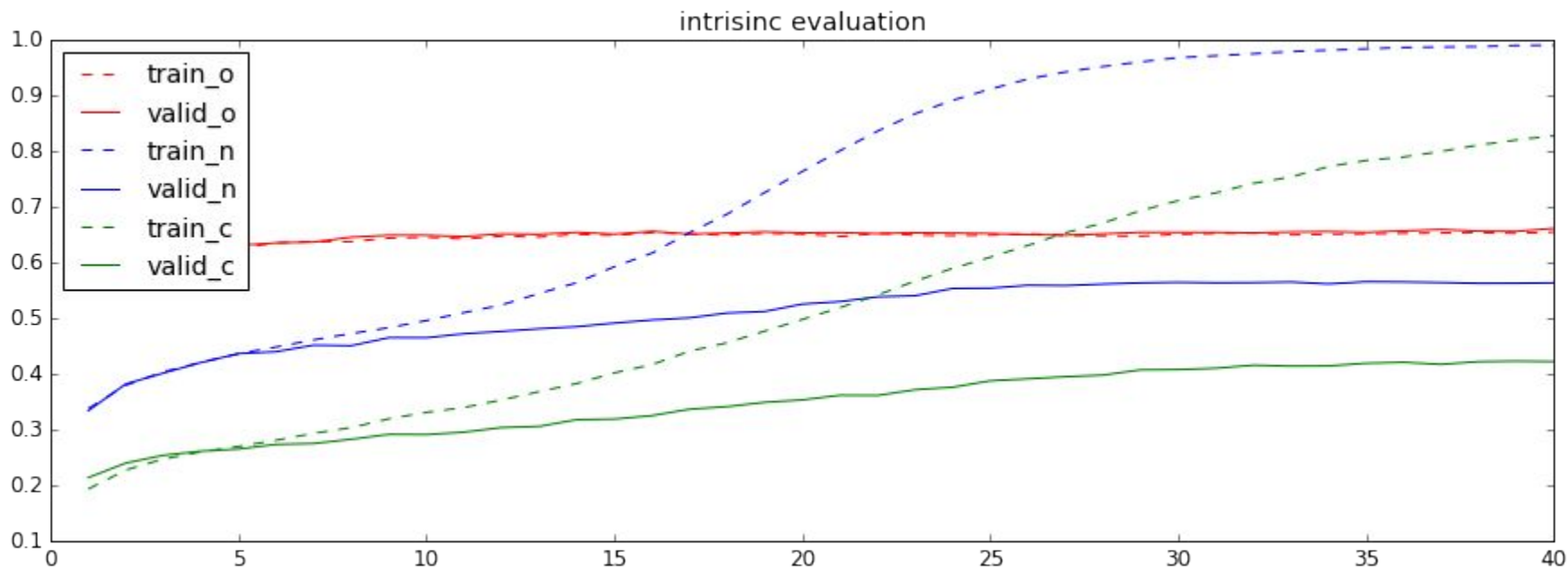# Model architecture

Joint training: unregularized

Joint training: unregularized

Joint training: dropouts

intrisinc evaluation

Joint training: dropouts

Joint training: intrinsic evaluation

	next/ conjunction	next/ conjunction/ dropout	order/next/ conjunction	order/next/ conjunction/ dropout
order	-	-	65.0	66.1
next	53.8	56.2	54.7	56.5
conjunction	40.8	41.9	41.3	42.2

Joint training: extrinsic evaluation

	Majority baseline	Skip-thought BiGRU 1200	Weak Supervision GRU 1024	Skip-thought GRU 2400	MTMETRICS
MSR paraphrase accuracy	66.5	71.2	**72.0**	73.0	***77.4***

Joint training: Qualitative evaluation

He held those positions until 1981 , and retired from judging in 1984 .
He held multiple positions before being named vice president for academic affairs in 1982 , as which he served until December 1986 .
Concurrently he served as member of the Council of State in extraordinary service from 1984 to 1989 .
She served as Chairman of UNICEF at the international level from 1987 to 1988 .
He served one two-year term , leaving office in 1983 .

Joint training: Qualitative evaluation

The (UNK) -theory has proven to be important for studying the general properties of quantum field theories .
Indeed , this theory subsumes a large number of mathematically interesting effective quantum field theories and points to new dualities relating these theories .
Leslie 's work has shown important connections between generic generalizations and social cognition .
The construction is widely applicable and is used to define an algebraic K-theory in a non-classical context .
The term quantum machine learning can also be used for approaches that apply classical methods of machine learning to problems of quantum information theory .

Joint training: Qualitative evaluation

To stop the monsters , players use a variety of traps and other objects that they can place in the corridors leading from the entrance point to the core that damage the monsters , as well as traps that are pre-built in the level , such as a trap that when triggered releases a large boulder down stairs .
To protect against attacks against airbases , and from being overrun , some air forces have a force dispersal doctrine that sees aircraft dispersed to secondary and emergency air bases , such as highway strips , and , as was the case with the Royal Air Force 's vertical take off Harriers , dispersals in forest clearings .
These UNK prefer to lurk among aquatic plants such as reeds , dead trees , and artificial structures of various types ; from where they wait for small prey animals to drift by or alight on the surface , before darting from their hiding place to hunt .
On the eve of the event , the gang encounters several challenges , such as being three hours behind schedule , guards on the premises and nearly dropping the heavy cable off the roof .

Conclusion

Can use these tasks to:

- Train sentence representations systems

- Compare different encoding schemes