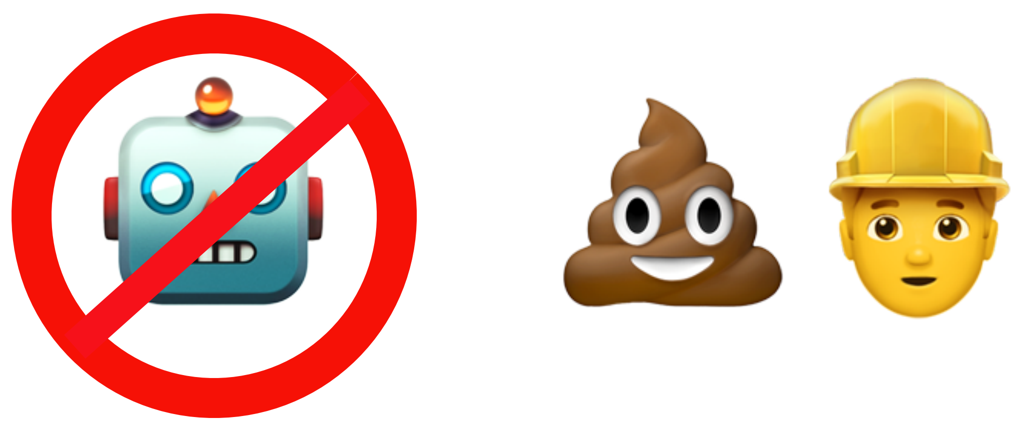
Learning Program Representations:

Symbols to Vectors to Semantics

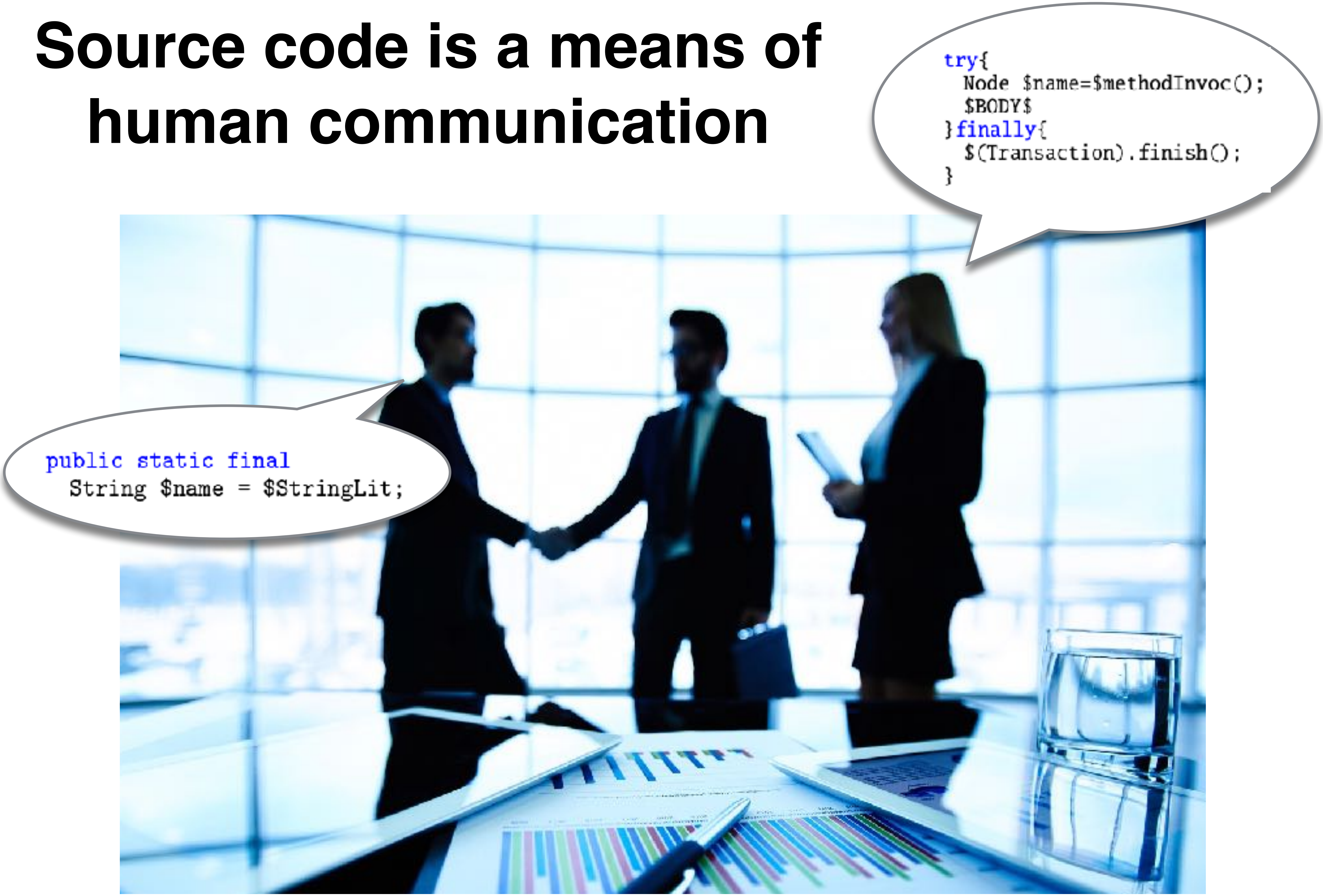
Charles Sutton
University of Edinburgh
& The Alan Turing Institute

10 December 2016

http://edin.ac/2ggR9uK



Source code is a means of human communication



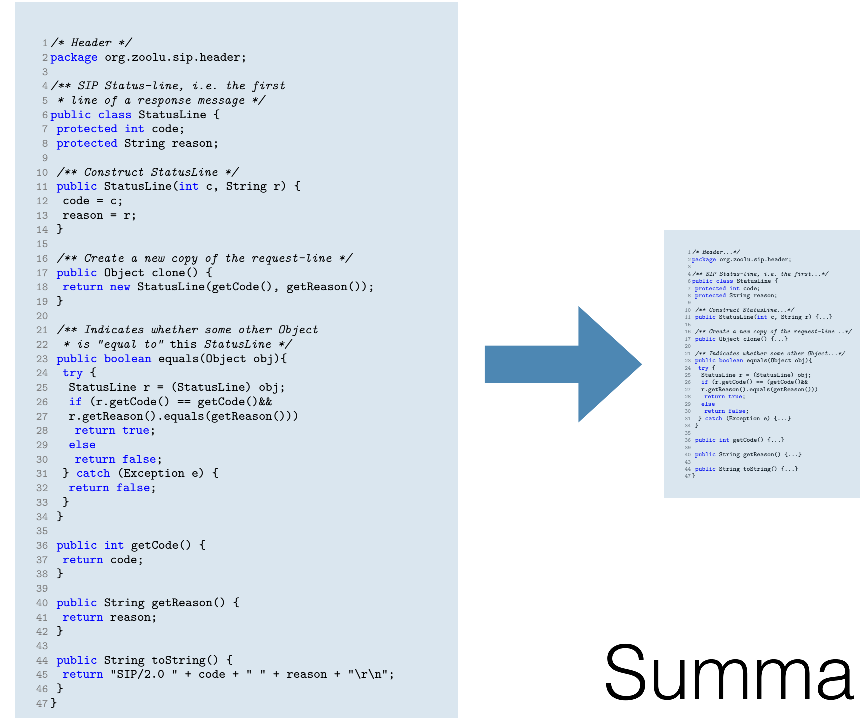
```
ConfigurationBuilder.<init>
ConfigurationBuilder.setOAuthConsumerKey
ConfigurationBuilder.setOAuthConsumerSecret
ConfigurationBuilder.build
TwitterFactory.<init>
TwitterFactory.getInstance
```

```
try{
  Node $name=$methodInvoc();
  $BODY$
}finally{
  $(Transaction).finish();
}
```

Learning how libraries are used

Nonparametric Bayes grammars *[FSE 2014]*

Probabilistic pattern mining *[KDD 2016; FSE 2016]*



Summarisation

Topic models *[ICSE 2016]*

Learning coding conventions

[FSE 2014, 2015]

Further ahead…

- Defining requirements
- Architecting
- Navigation
- Maintenance
- Optimising performance
- Testing, verification
- Refactoring
- Porting
- Debugging

Neural networks that capture program semantics

Why not solved?

1. Domain connections
2. Typical programs

Learning to Name

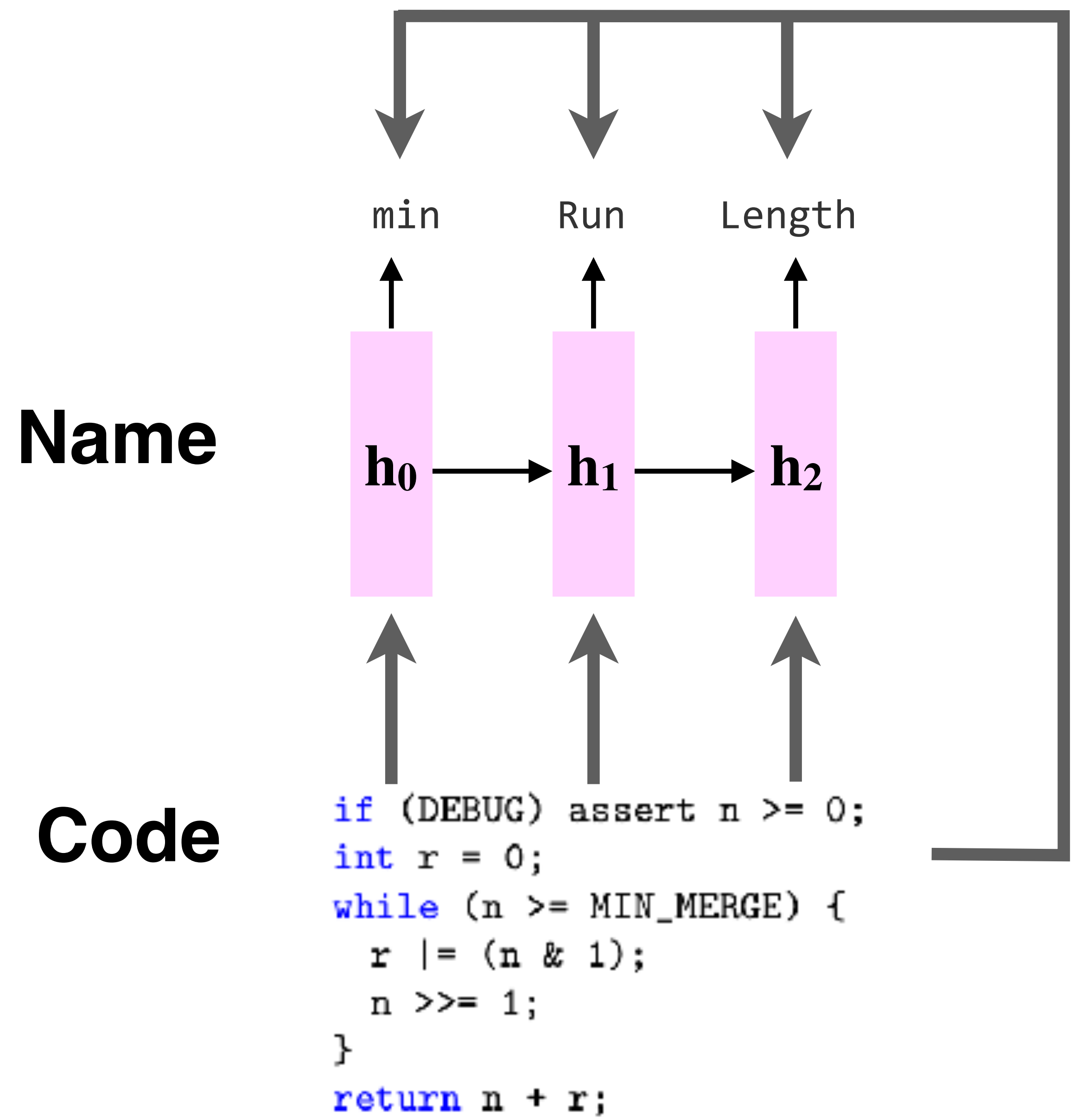
```
if (DEBUG) assert n >= 0;
int r = 0;
while (n >= MIN_MERGE) {
  r |= (n & 1);
  n >>= 1;
}
return n + r;
```

→ minRunLength

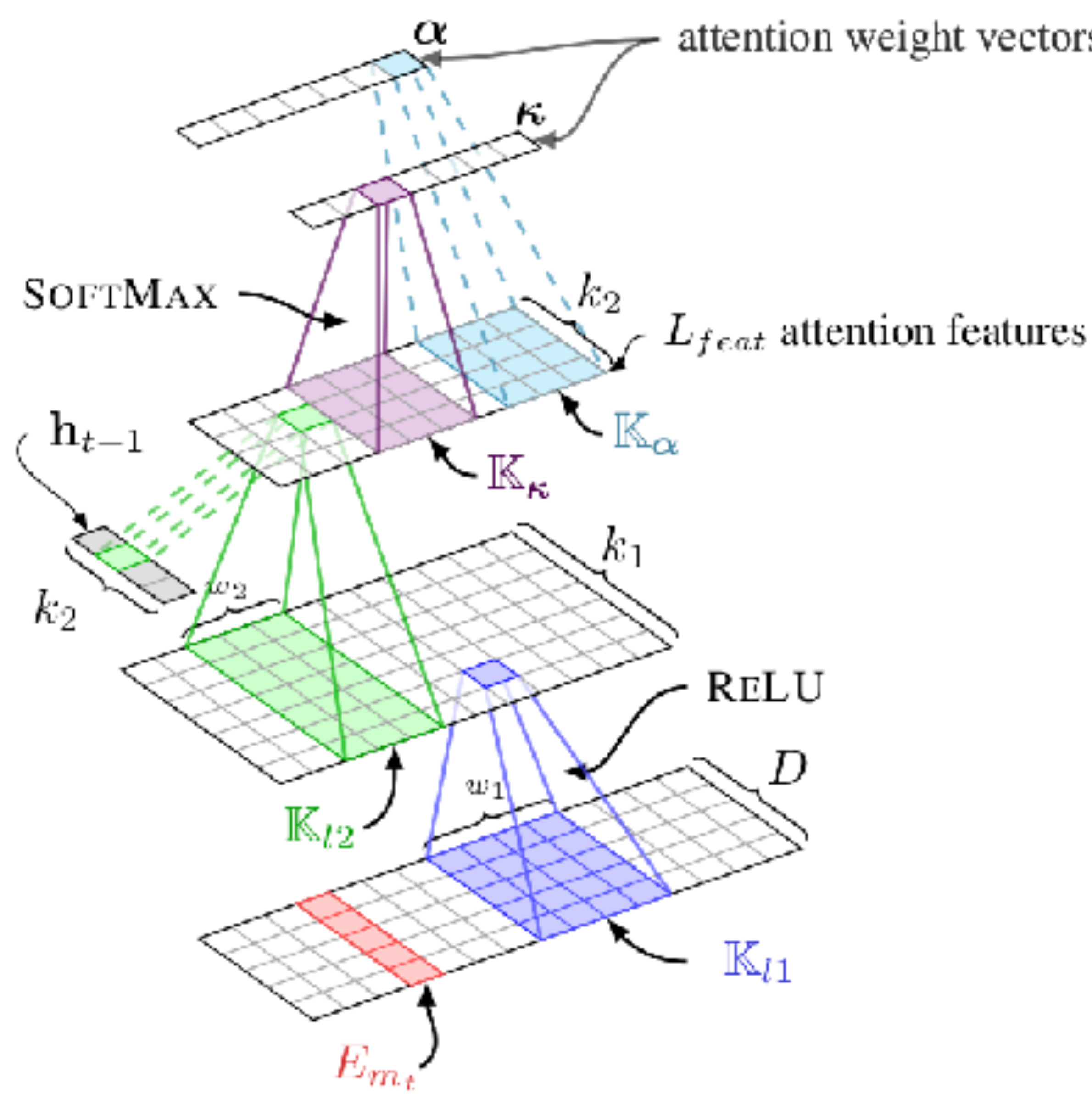
[Allamanis, Peng, and Sutton ICML 2016]

http://edin.ac/2ggR9uK

Predicting Names of Methods



RNN for generating summary



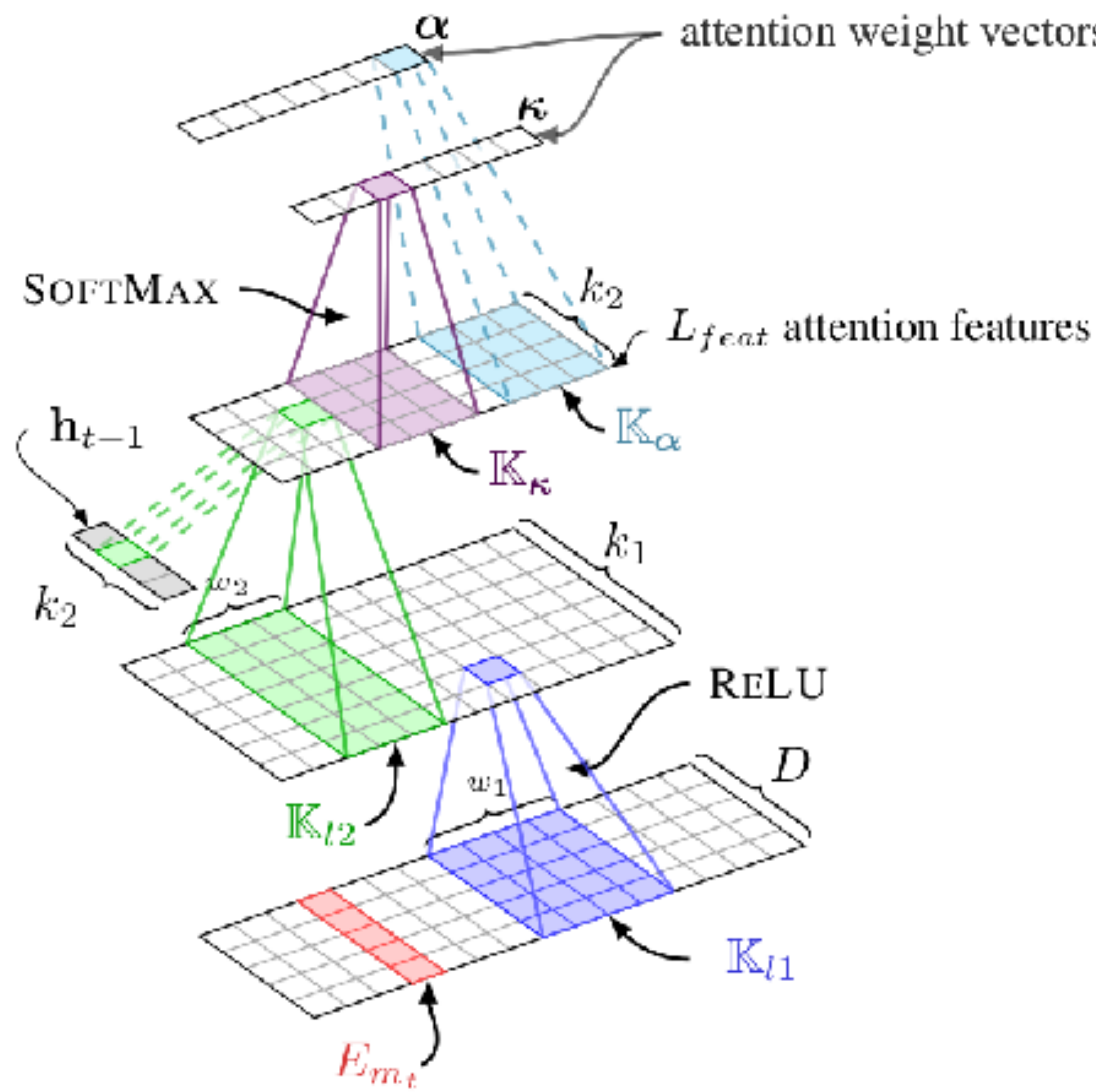
convolutional attention mechanism

[Allamanis, Peng, and Sutton ICML 2016]

http://edin.ac/2ggR9uK

Three Attention Mechanisms

- α : Distribution over input locations
  - Weights for averaging input embeddings
- κ : Distribution over input locations
- Weights for copying tokens from input to output (even OOV)
  - Related to pointer networks
    [Vinyals et al, 2015]
- λ: Scalar [0, 1]
  - weight to decide two mechanisms



[Allamanis, Peng, and Sutton ICML 2016]

http://edin.ac/2ggR9uK

	F1 (%)		Exact Match (%)	
	Rank 1	Rank 5	Rank 1	Rank 5
tf-idf	40.0	52.1	**24.3**	29.3
Standard Attention	33.6	45.2	17.4	24.9
conv_attention	43.6	57.7	20.6	29.8
copy_attention	**44.7**	**59.6**	23.5	**33.7**

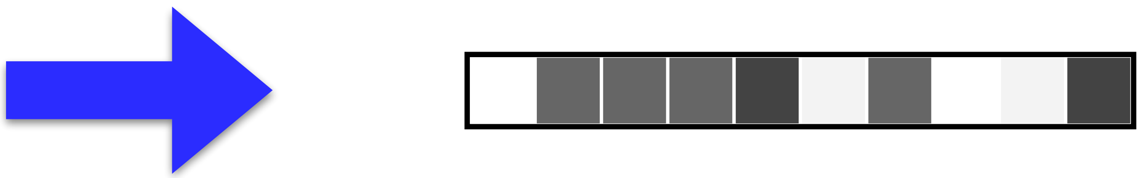
Standard attention: [Bahdanau, Cho, and Bengio, 2015]

http://edin.ac/2ggR9uK

Continuous Semantics for Symbolic Expressions

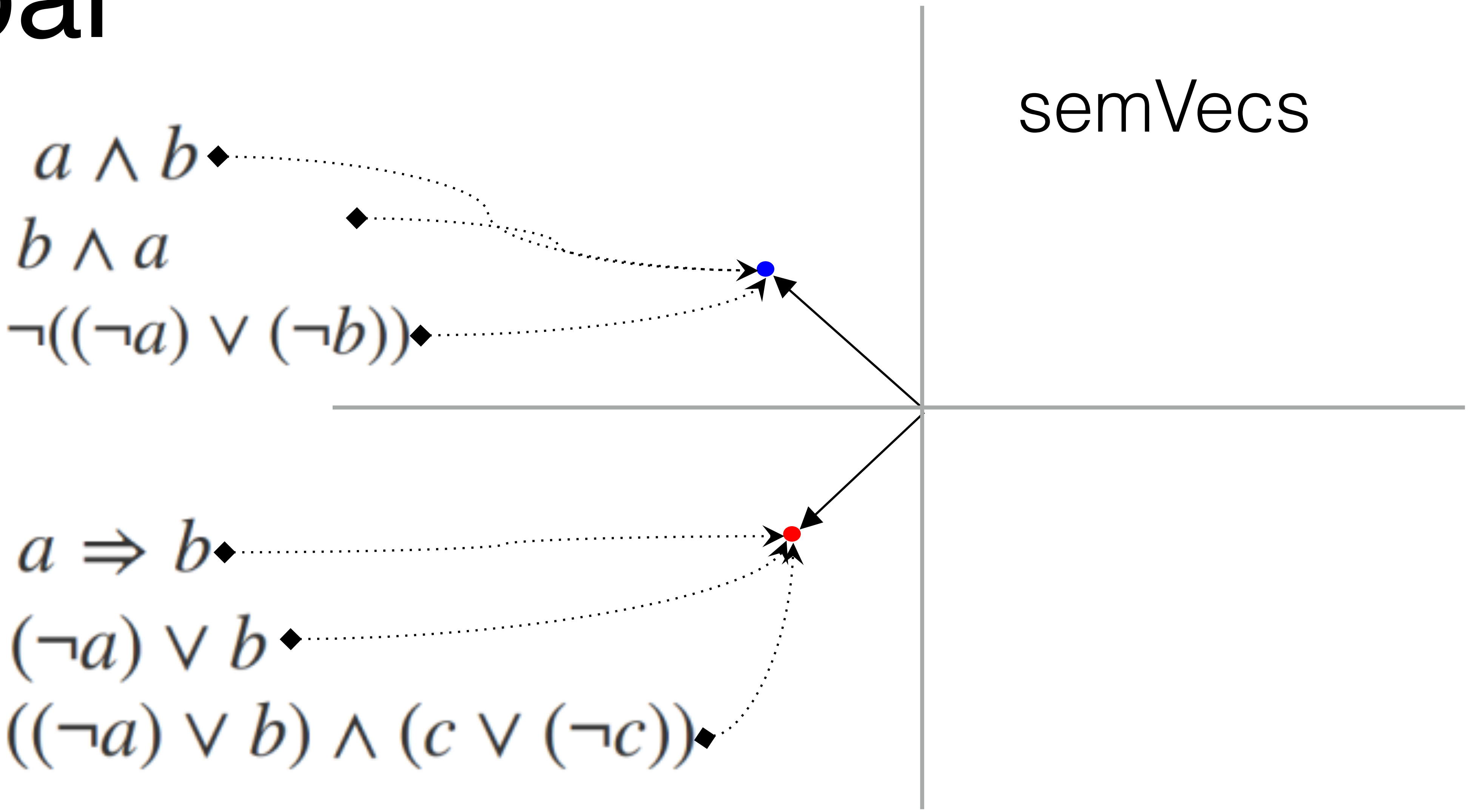
[Allamanis, Chanthirasegaran Kohli, and Sutton, 2017]

```
(a-b)*(b+c)+(b-b)
a*b+a*c-b*(b+c)
a*c+b*(a-b-c)
```



http://edin.ac/2ggR9uK

Goal



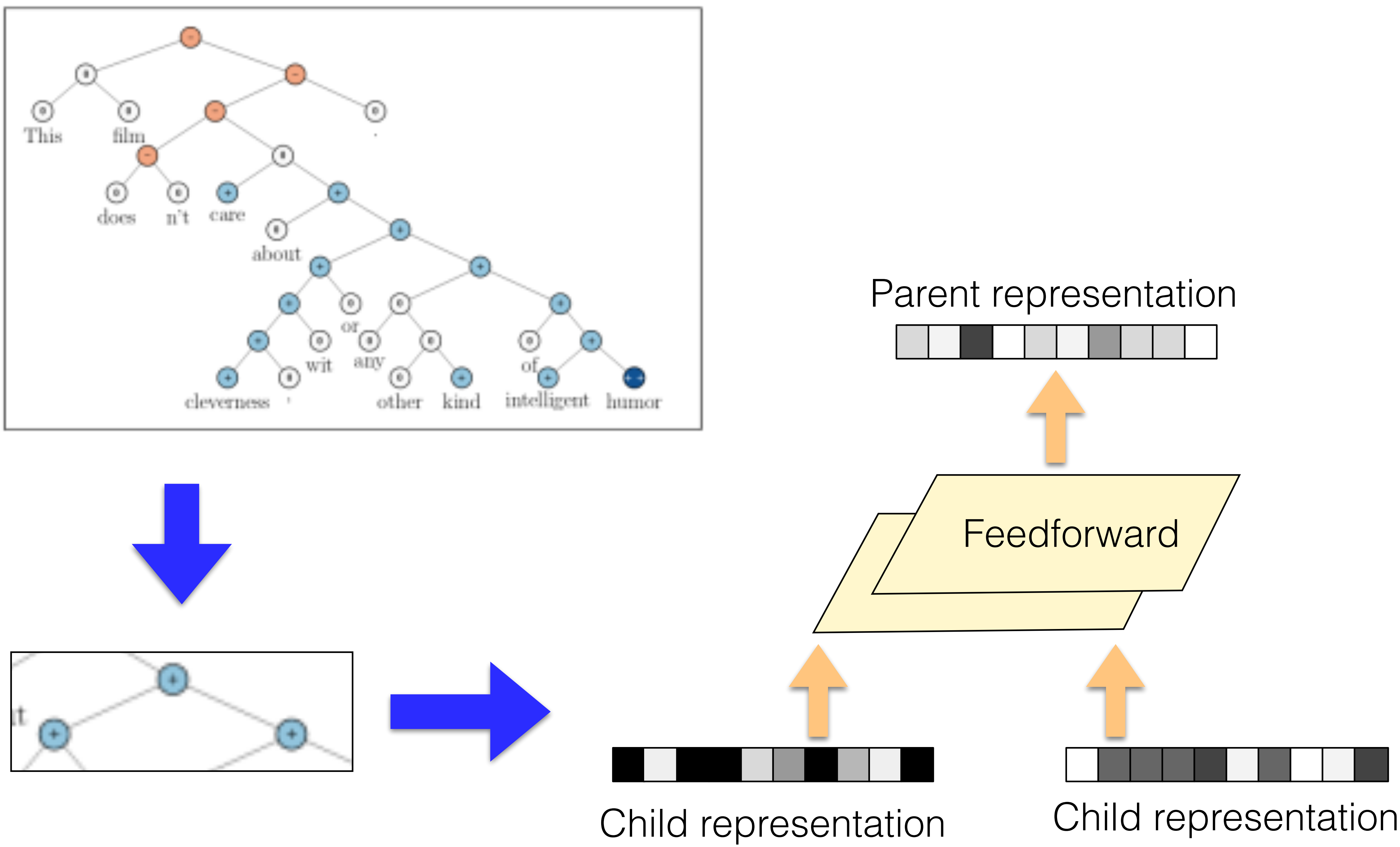
How much symbolic semantics (**semantic equivalence**)
can we compress into **continuous** vector?

In this work: semantics = equivalence
Assume we have an oracle

Symbolic reasoning: ~~search~~ **pattern recognition**

http://edin.ac/2ggR9uK

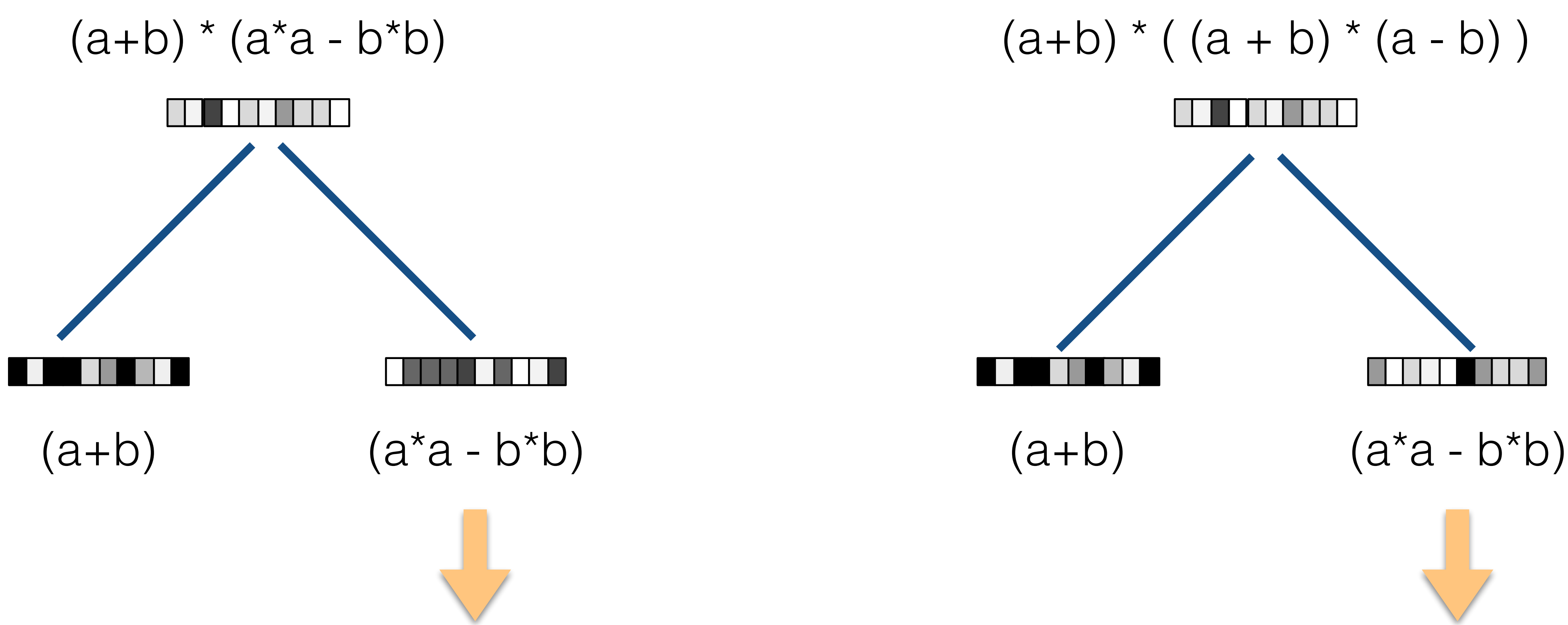
Recursive NN (TreeNN)



[Socher et al, 2011, 2013]

http://edin.ac/2ggR9uK

Problem: Separating out syntax

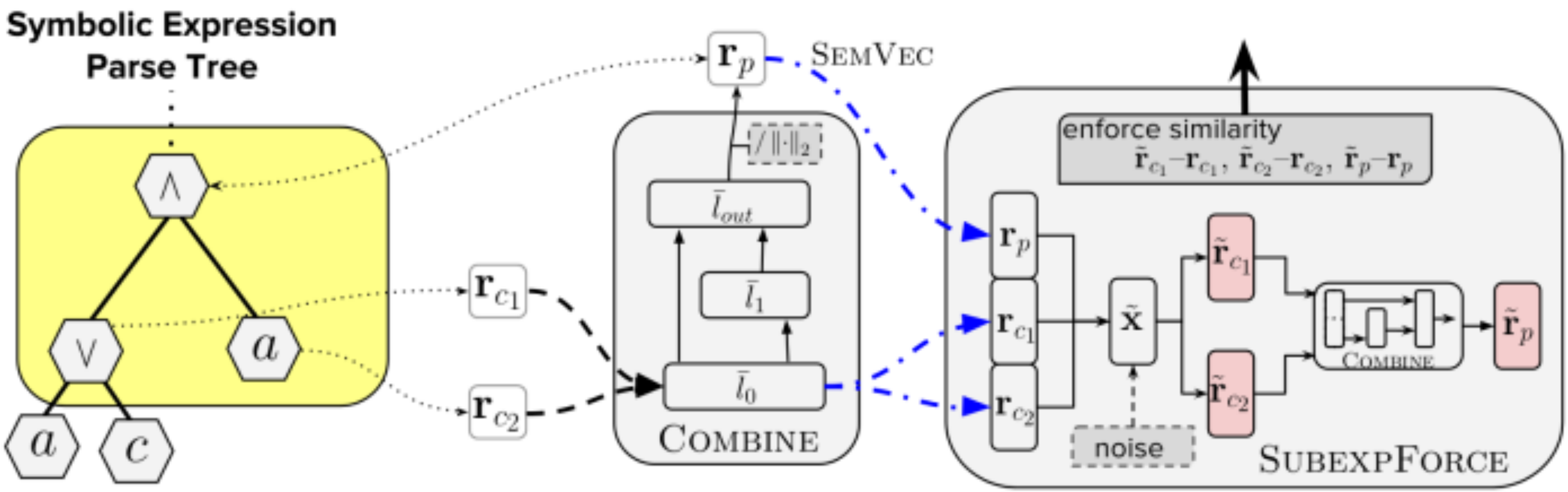


semantically equivalent, different vectors!

Result: nearest neighbours mostly reflect syntax

http://edin.ac/2ggR9uK

EqNet



http://edin.ac/2ggR9uK

Motivation via Unification

Semantic information is bidirectional

Not only do **children** provide info re **parents**

But **parents** provide info re **children**

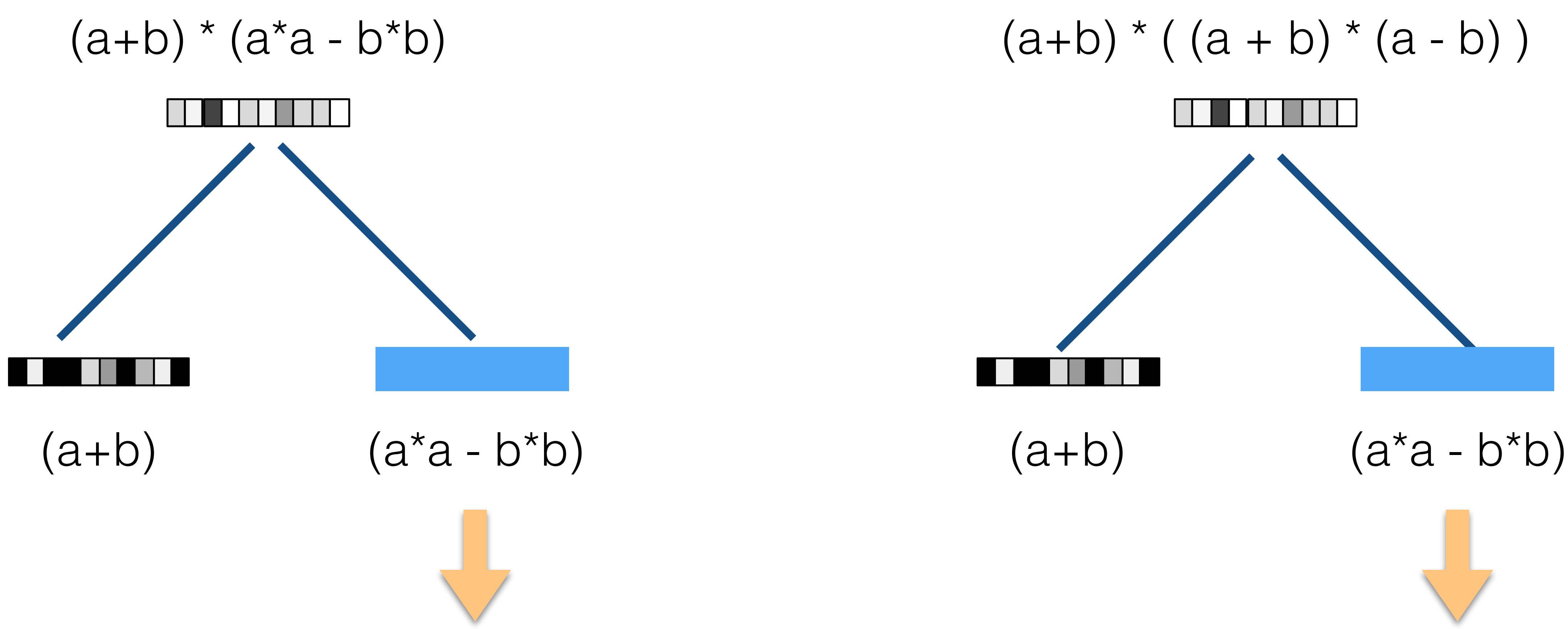
```
uncle(?A,?B) :- parent(?A,?Z), brother(?Z,?B)
```

Unification propagates this info automatically

How to map to continuous space?

http://edin.ac/2ggR9uK

Subexpression Forcing

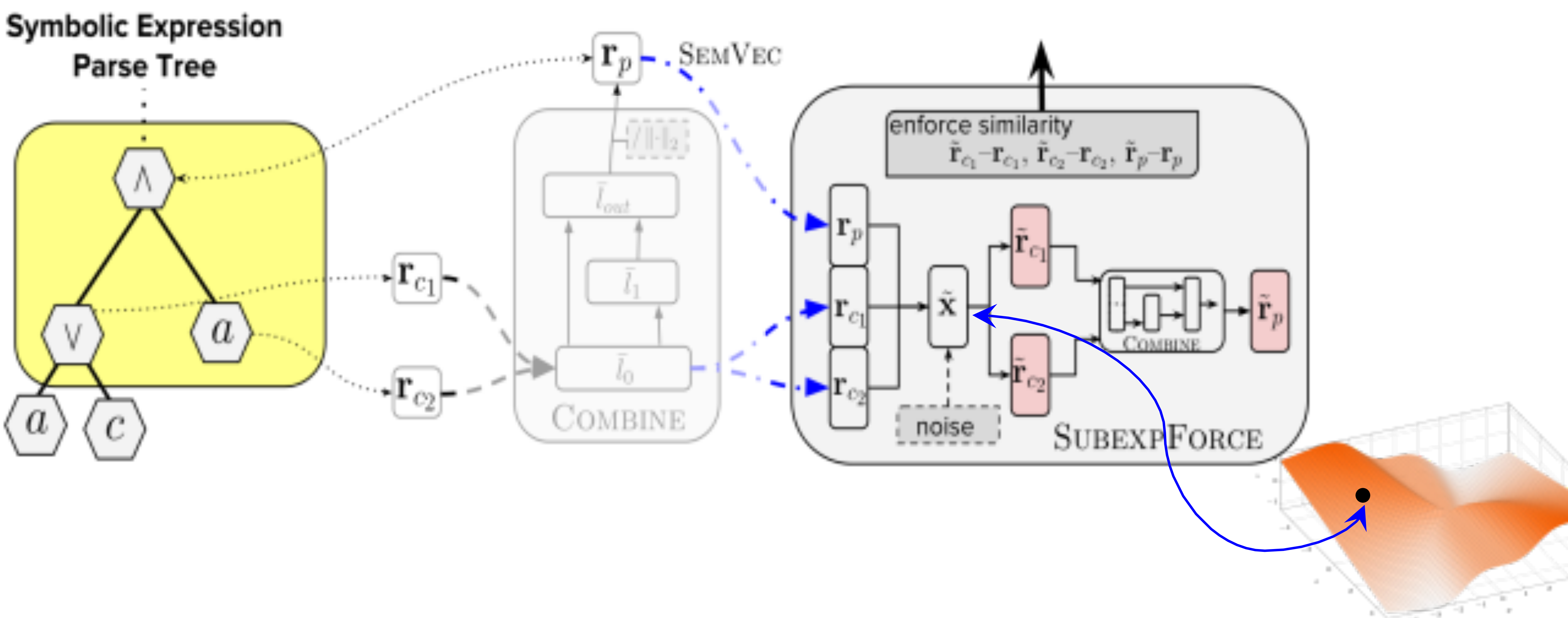


ensure this prediction problem is “easy”

semantic classes will be clustered together

http://edin.ac/2ggR9uK

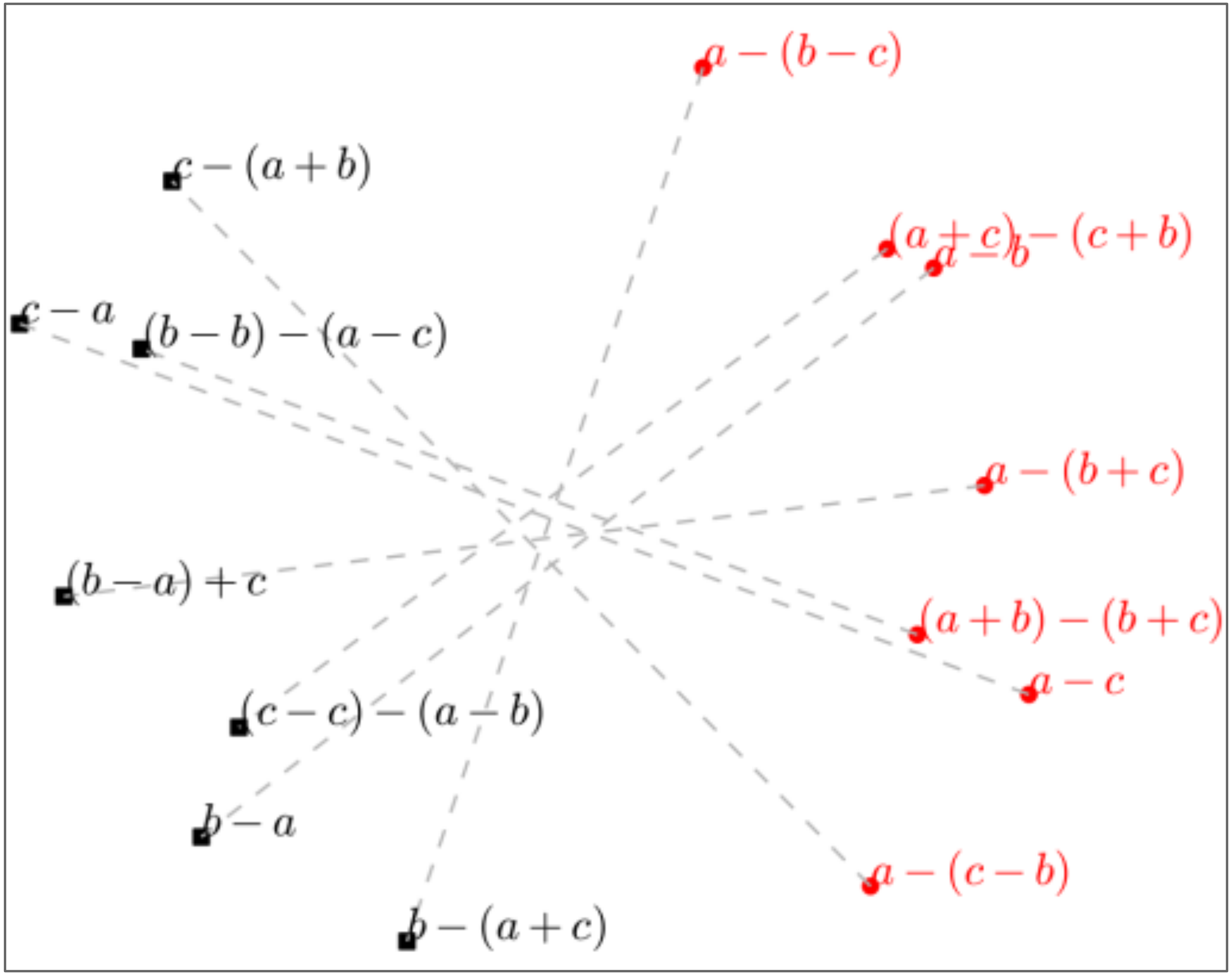
Subexpression Forcing



Denoising autoencoder plus bottleneck
on (parent, child1, child2) representations
(Additional regulariser)

http://edin.ac/2ggR9uK

Visualizing polynomials



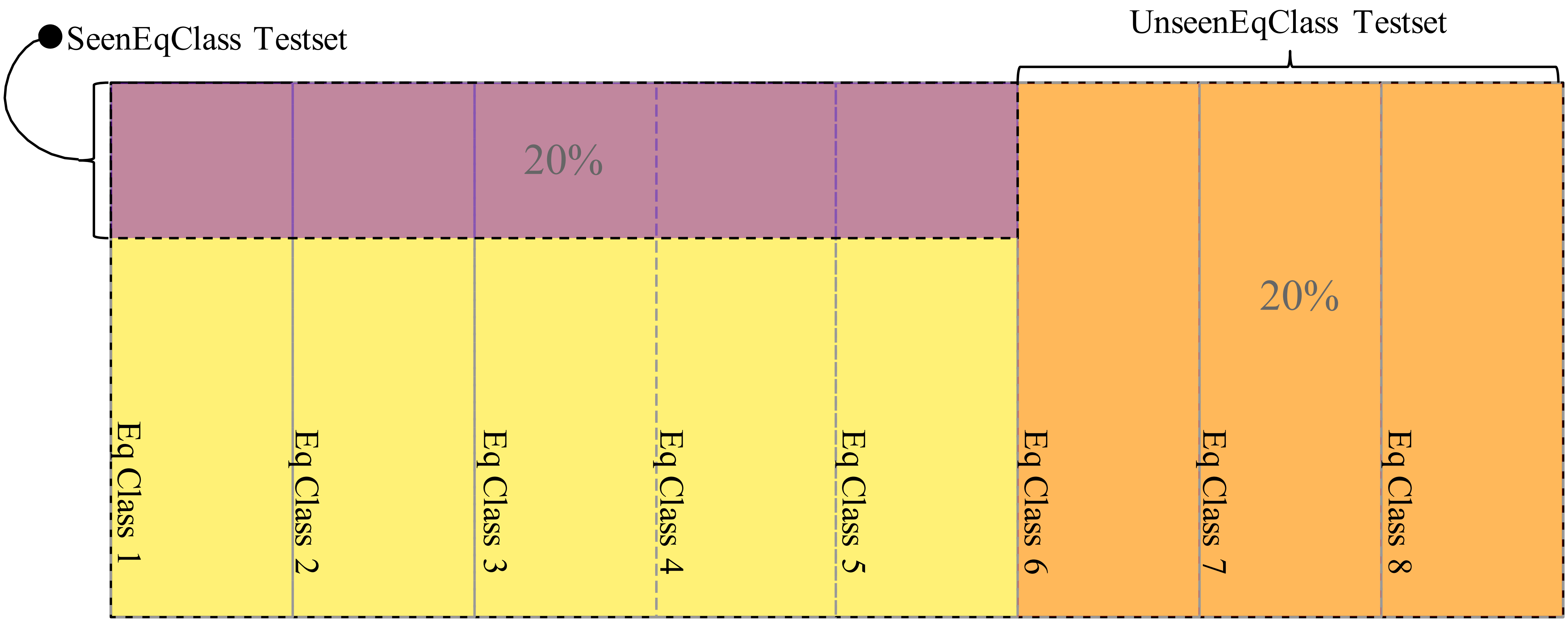
http://edin.ac/2ggR9uK

Evaluation

Dataset	# Vars	# Equiv Classes	# Exprs	H
SimpBool8	3	120	39,048	5.6
SimpBool10^S	3	191	26,304	7.2
Bool5	3	95	1,239	5.6
Bool8	3	232	257,784	6.2
Bool10^S	10	256	51,299	8.0
SimpBoolL5	10	1,342	10,050	9.9
BoolL5	10	7,312	36,050	11.8
SimpPoly5	3	47	237	5.0
SimpPoly8	3	104	3,477	5.8
SimpPoly10	3	195	57,909	6.3
oneV-Poly10	1	83	1,291	5.4
oneV-Poly13	1	677	107,725	7.1
Poly5	3	150	516	6.7
Poly8	3	1,102	11,451	9.0

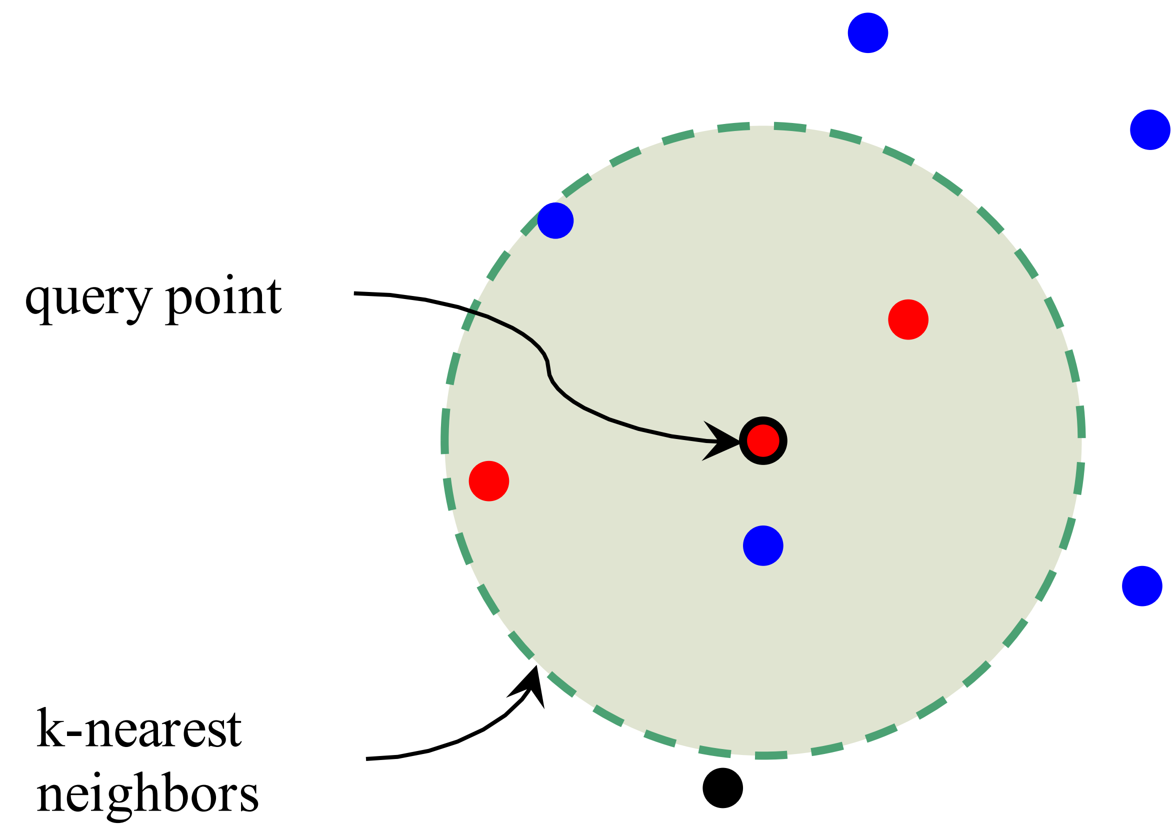
http://edin.ac/2ggR9uK

Training / Test Split



http://edin.ac/2ggR9uK

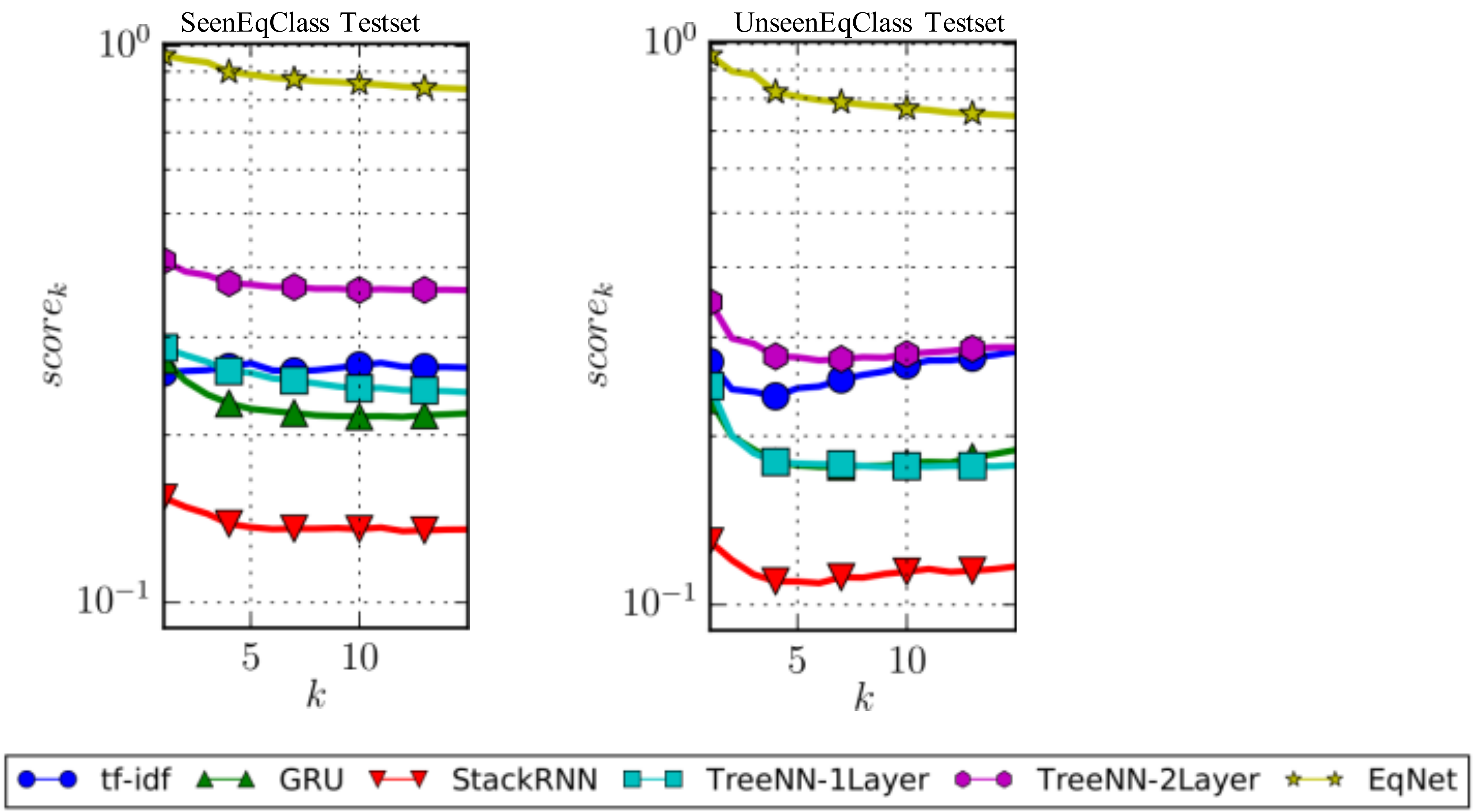
Evaluation Metric



$$score_k(q) = \frac{|\mathbb{N}_k(q) \cap c|}{\min(k, |c|)}$$

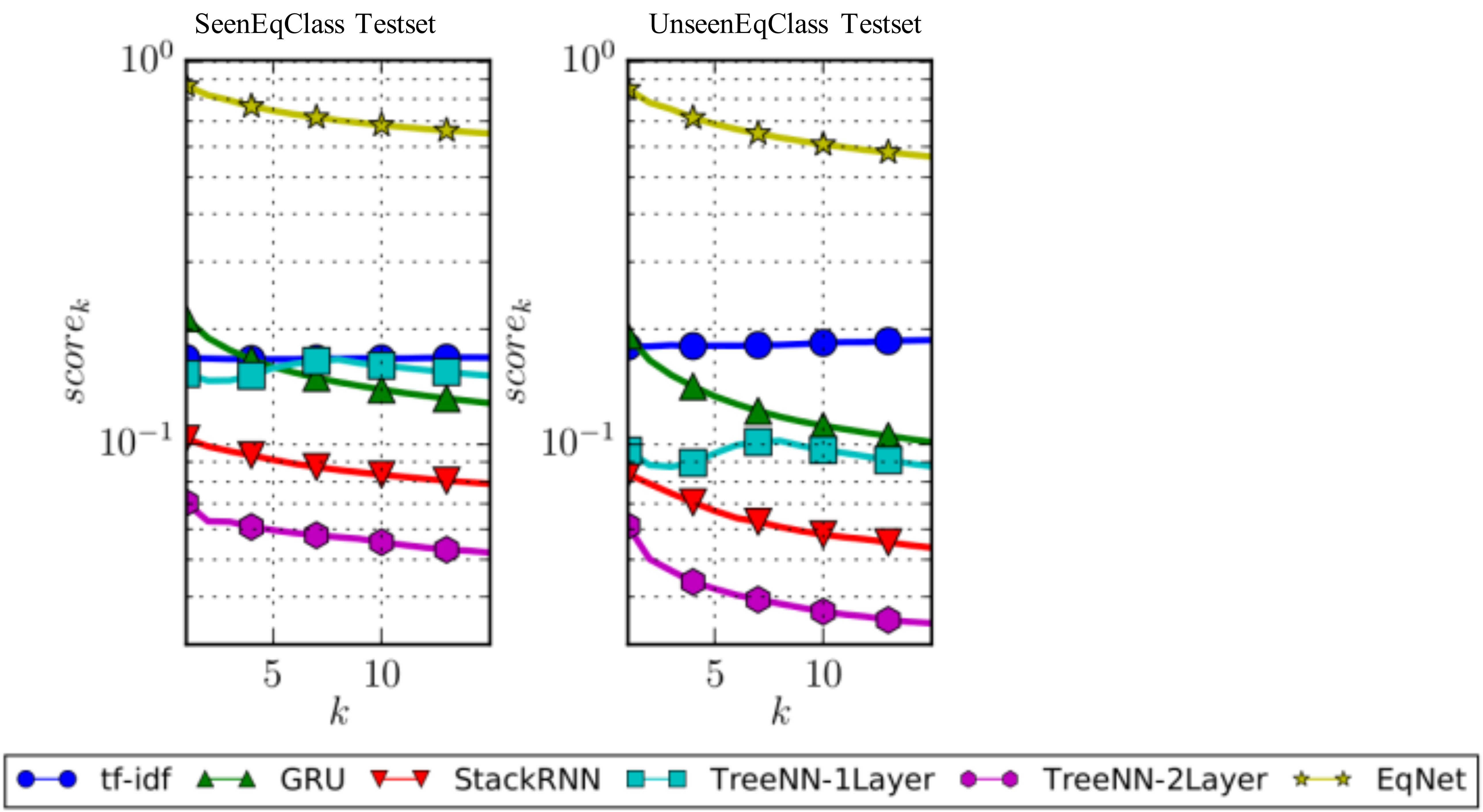
http://edin.ac/2ggR9uK

Results



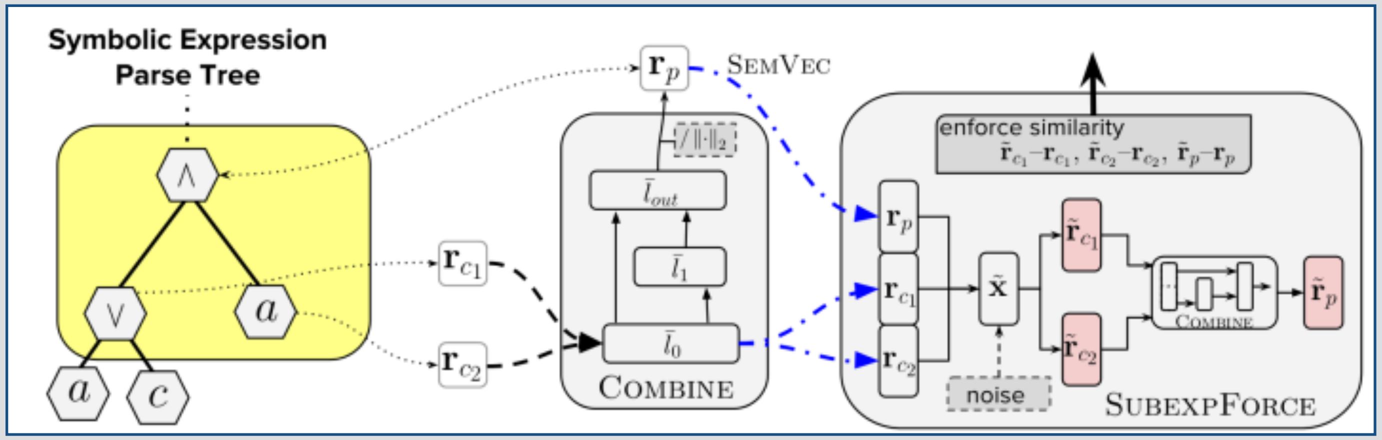
http://edin.ac/2ggR9uK

Evaluating compositionality

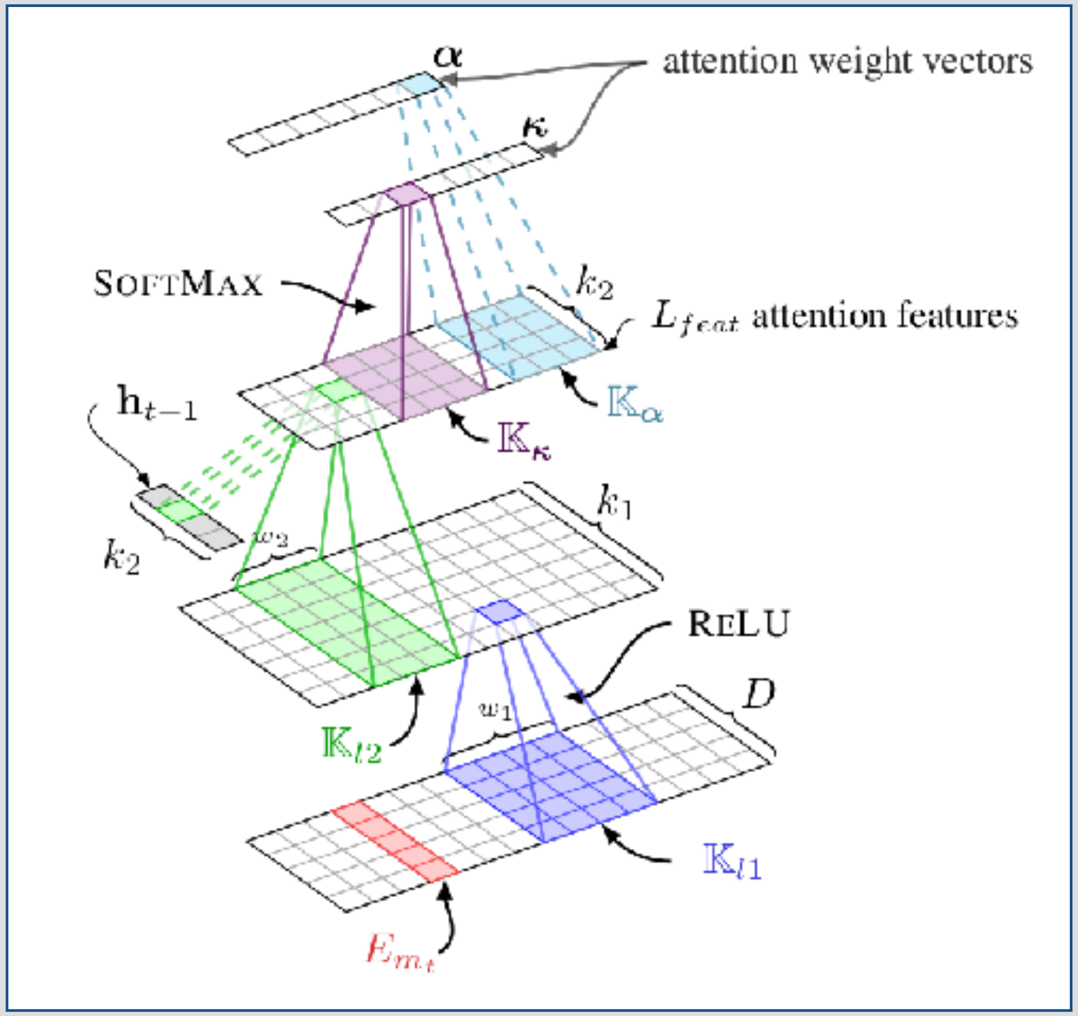


http://edin.ac/2ggR9uK

Learning Program Representations:
Symbols to Vectors to Semantics
Charles Sutton, University of Edinburgh



Equivalence networks
for continuous semantics



Naming methods
convolutional attention

Thanks!

- Miltiadis Allamanis
- Hao Peng
- Pushmeet Kohli
- Pankajan Chantirasagaran

EPSRC
Engineering and Physical Sciences Research Council

Microsoft Research

http://edin.ac/2ggR9uK