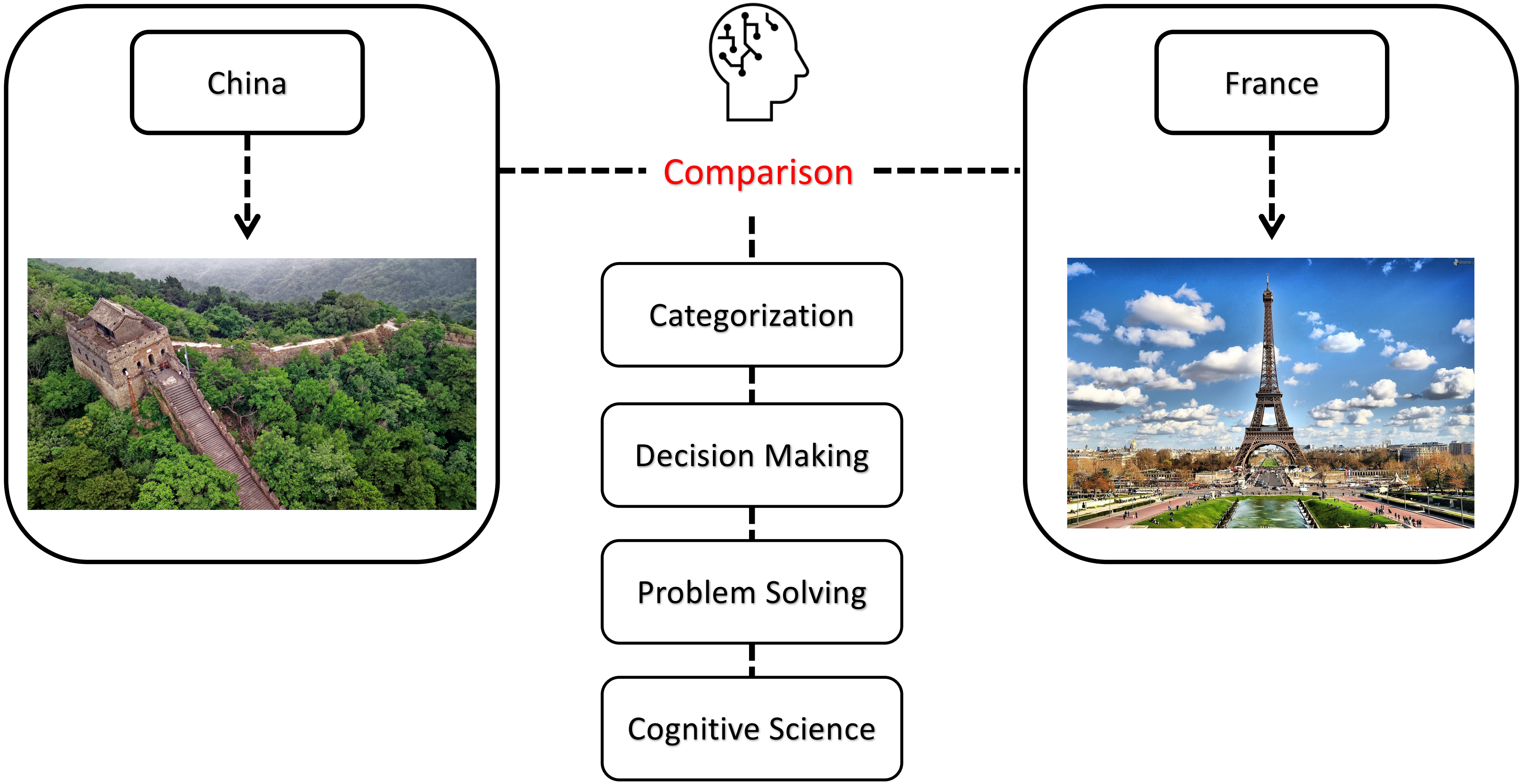
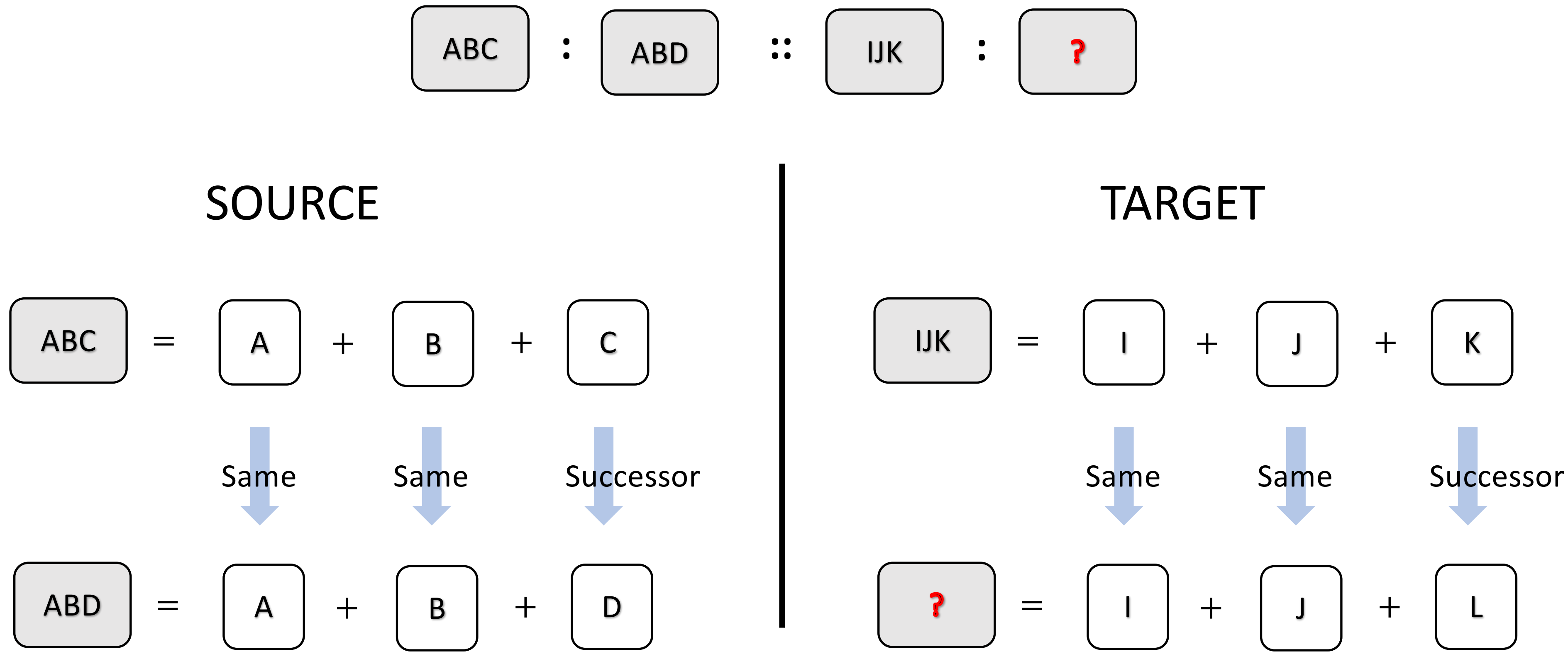
Logic Based Neural Networks for Analogy Solving

Lingsong Meng

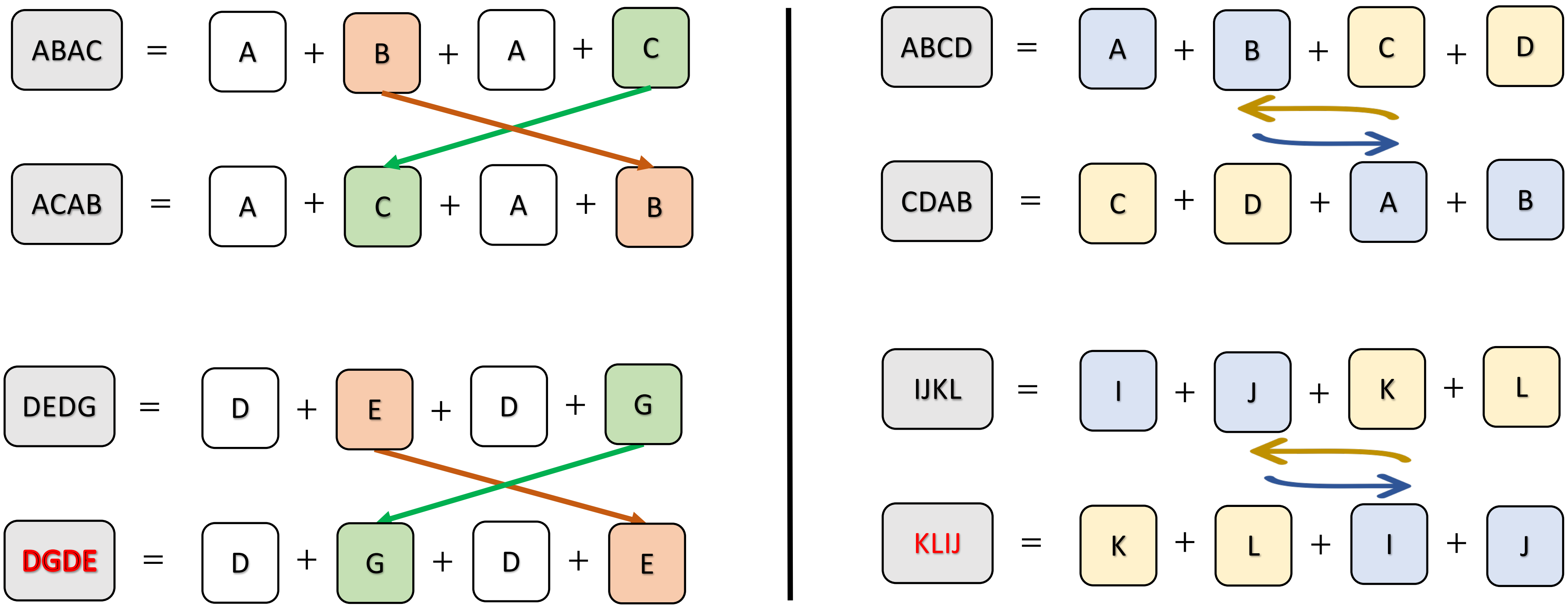
What is analogy?



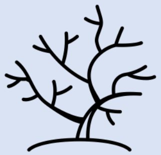
Letter-string Analogy



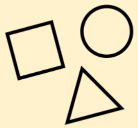
Letter-string Analogy



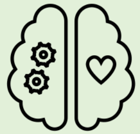
Challenges for Machines



Analogous Relations

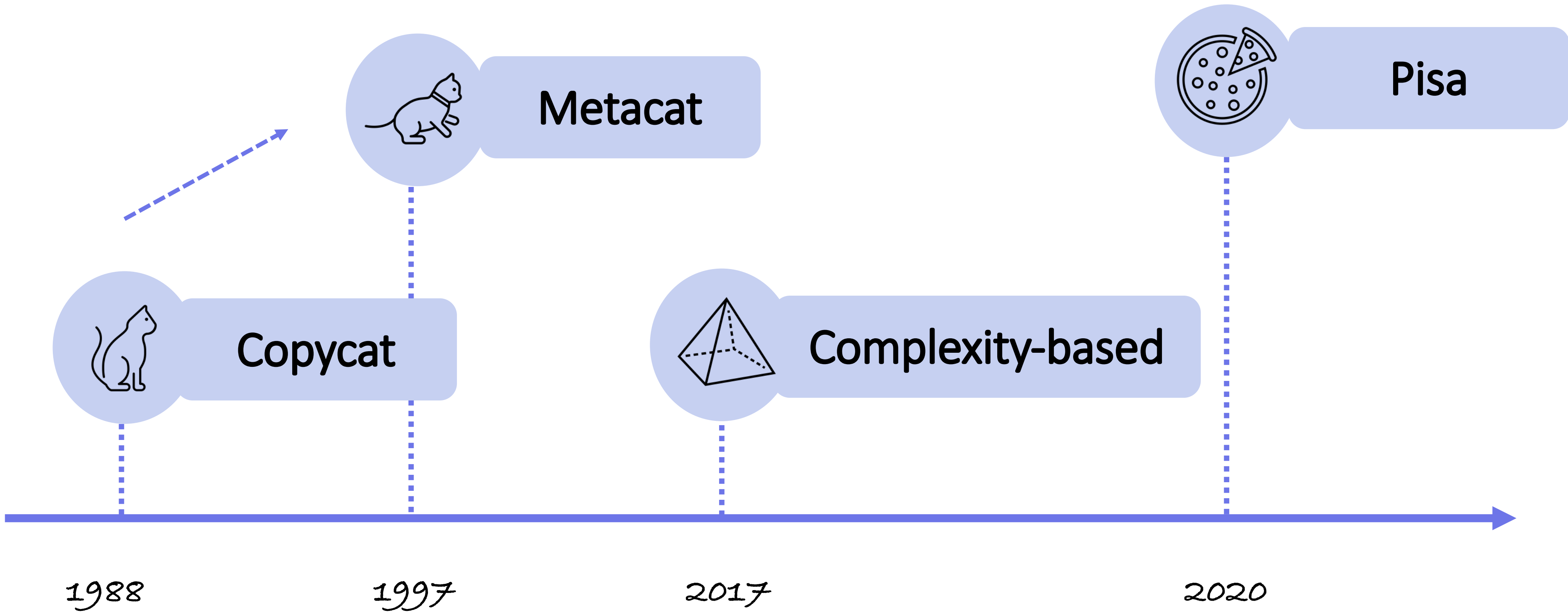


Discrete Manner

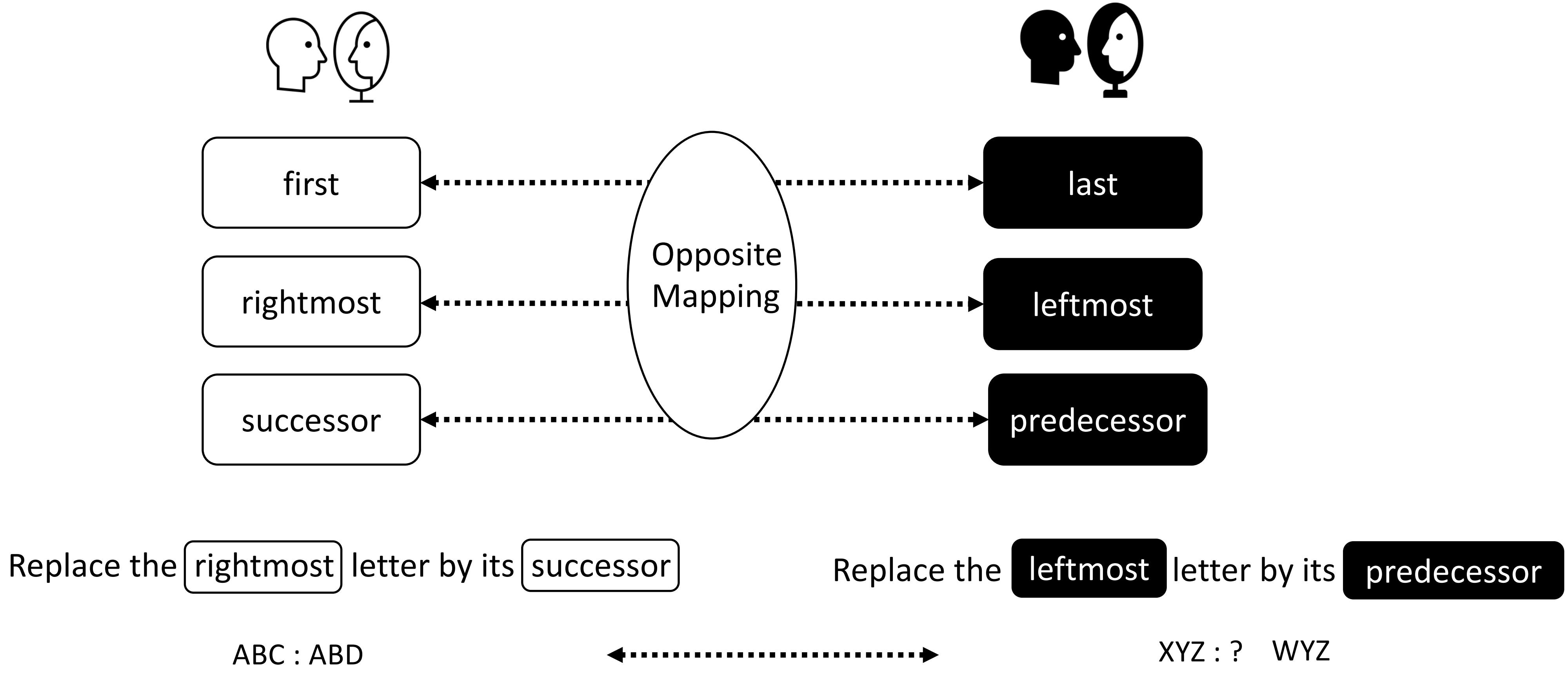


Cognitive Reasoning

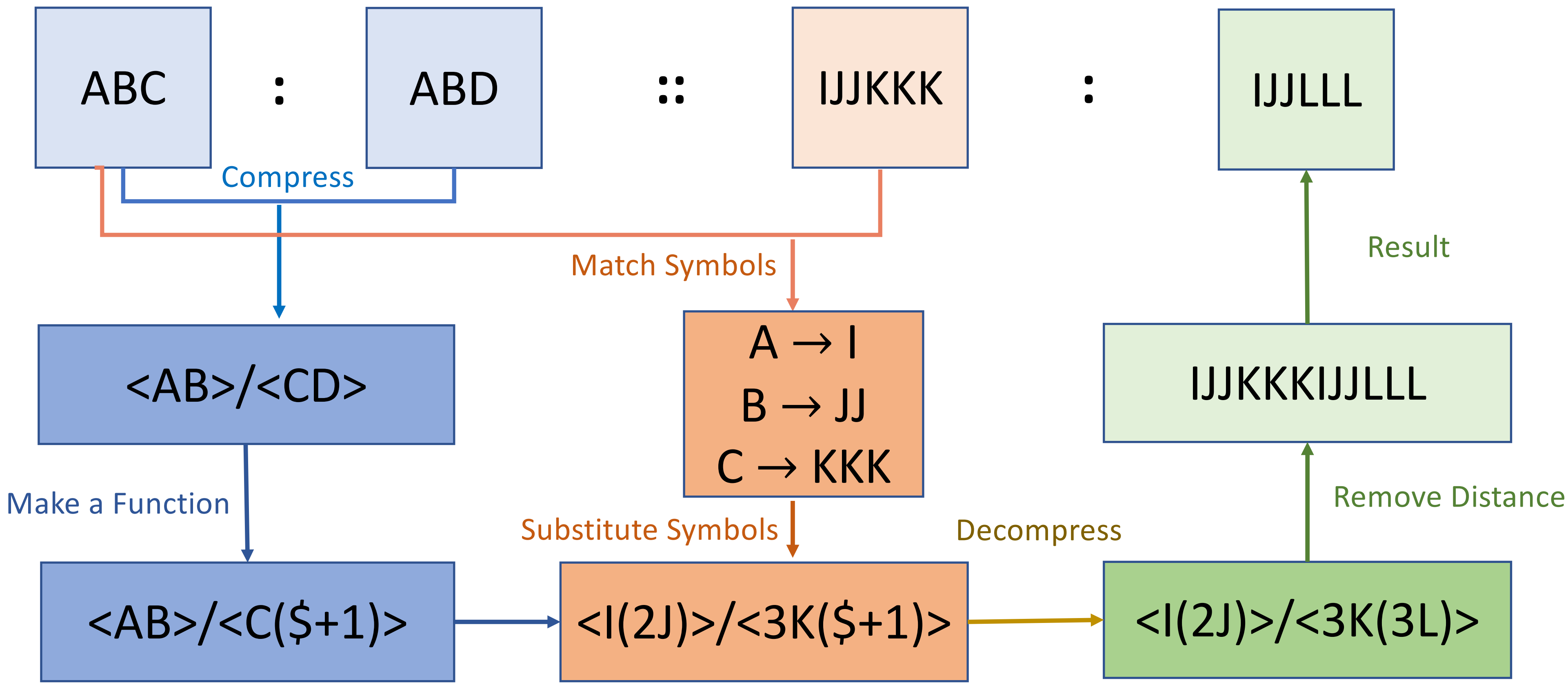
Related Methods



Copycat & Metacat

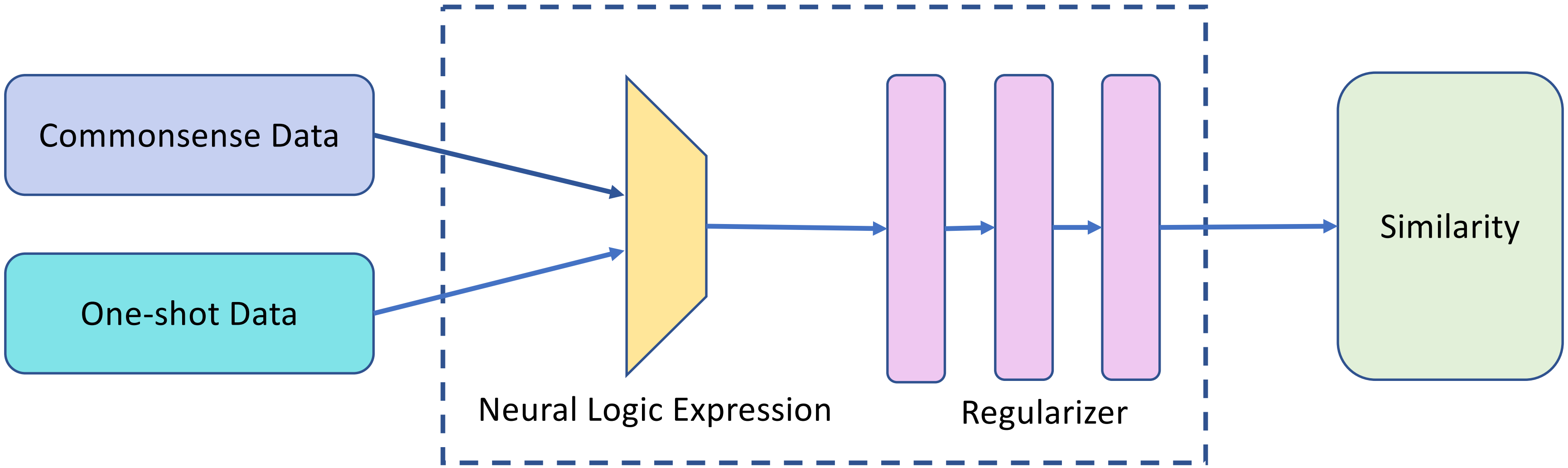


Pisa


ABC
:
ABD
::
IJJKKK
:
IJJLLL
Compress
Match Symbols
Result
<AB>/<CD>
A → I
B → JJ
C → KKK
IJJKKKIJJLLL
Make a Function
Substitute Symbols
Decompress
Remove Distance
<AB>/<C($+1)>
<I(2J)>/<3K($+1)>
<I(2J)>/<3K(3L)>

Neural Logic Analogy Learning (Noan)

- A dynamic neural architecture to solve analogies based on Logic-Integrated Neural Networks (LINN)



Model Preliminaries

Basic Operations

AND

OR

NOT

Basic Laws

$$\neg(x \land y) \iff \neg x \lor \neg y$$

$$\neg(x \lor y) \iff \neg x \land \neg y$$

Material Implication

$$x \to y \iff \neg x \lor y$$

Reasoning with Commonsense Data

Human level understanding about the alphabetical order

Positive

Negative

- A → A, B → B, C → C, …
- A → B, B → C, C → D, …
- AA → AA, BB → BB, CC → CC, …

- A → C, B → D, C → E, …
- C → A, D → B, E → C, …
- AA→ CC, BB → DD, CC → EE, …

Commonsense dataset in Noan

- Positive Sample
  - One−order
    - A → A, B → B, C → C, …
    - A → B, B → C, C → D, …
    - B → A, C → B, D → C, …
  - Two−order
    - AA → AA, BB → BB, CC → CC, …
    - AB → BC, BC → CD, CD → DE, …
    - BC → AB, CD → BC, DE → CD, …
- Negative Sample: Any case does not belong to One-order & Two-order

Reasoning with One-shot Data

If $\boxed{\text{ABC} \rightarrow \text{ABD}}$, IJK $\rightarrow$?

One-shot Data - - - Specific piece of data given to the user

III $\rightarrow$?

If given AAA $\rightarrow$ BBB, the answer should be JJJ

If given AAA $\rightarrow$ A, the answer should be I

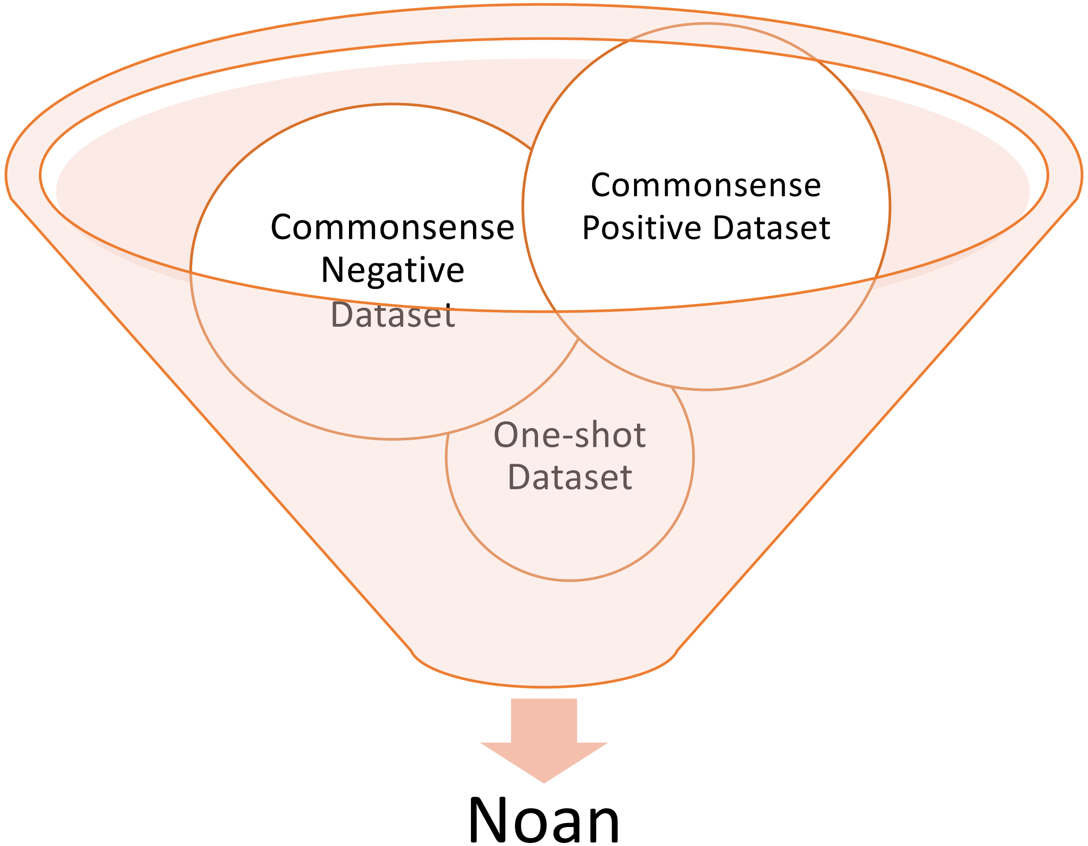
Reasoning with One-shot Data

If ABC → ABD, IJK → ?

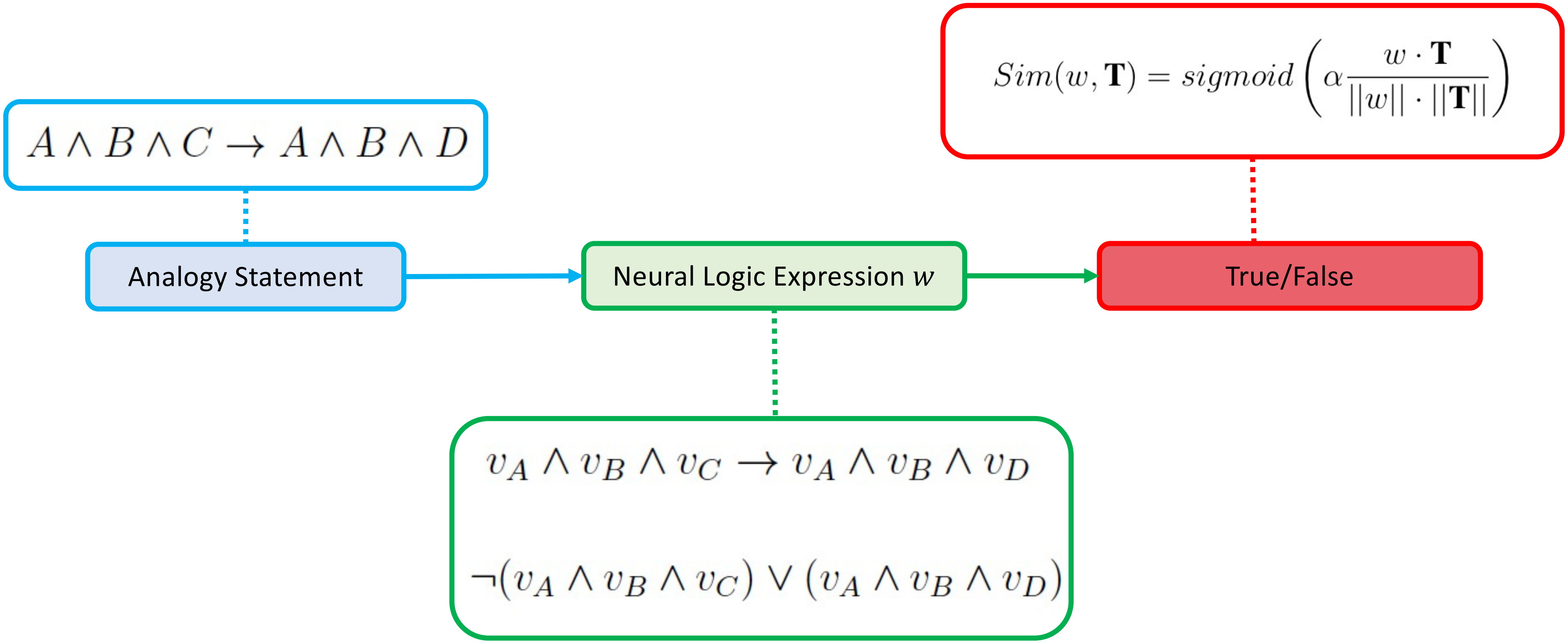
- Expand one-shot dataset by repeating this one piece of data

- One-shot Dataset: $\begin{cases} \text{ABC} \rightarrow \text{ABD} \\ \text{ABC} \rightarrow \text{ABD} \\ \quad\quad \ldots \end{cases}$

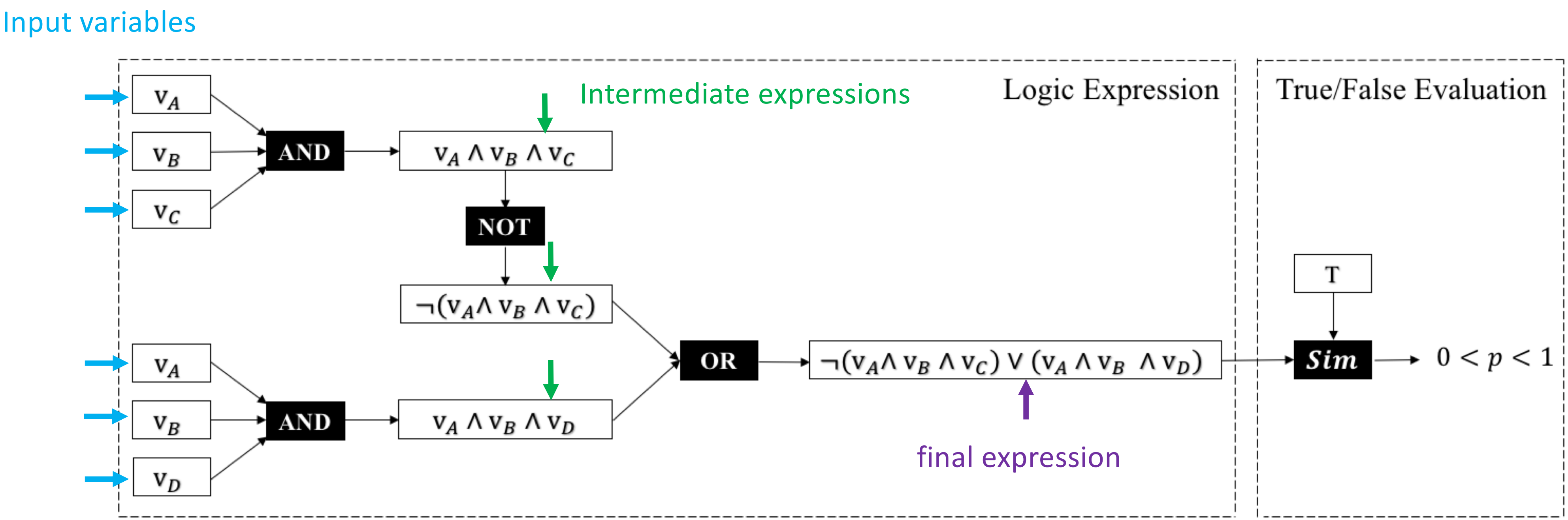
Dataset in Noan



Neural Modules



Neural Modules



Neural Modules

- Consider True/False Prediction as a classification problem, and the according loss function is:

$$E = \{e_i\}_{i=1}^{m}$$

A set of expressions

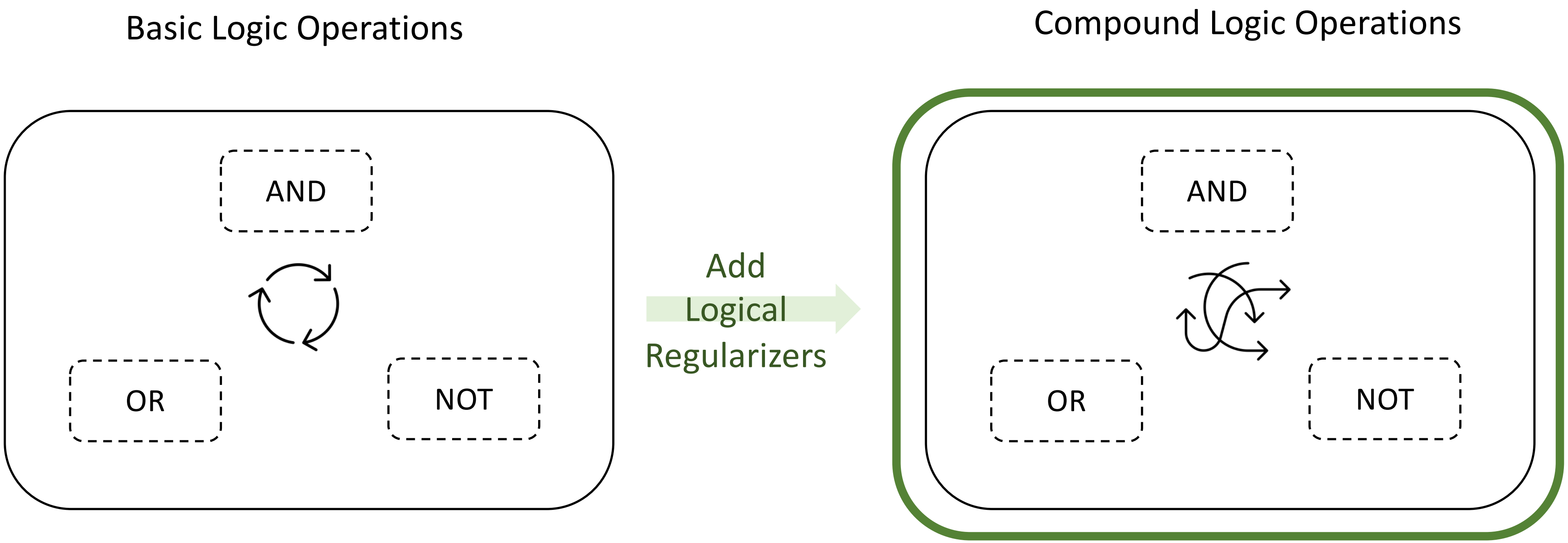
$$\mathrm{Y} = \{y_i\}_{y=1}^{m}$$

Corresponding True/False values

$$L_{loss} = L_{ce} = -\sum_{e_i \in E} y_i log(p_i) + (1 - y_i) log(1 - p_i)$$

Neural Modules

Not Enough!



Neural Modules

Logical regularizers and the corresponding logical rules

	Logical Rule	Equation	Logic Regularizer r_i
NOT	Negation	$\neg T = F$	$r_1 = \sum_{w \in W \cup \{T\}} Sim(\text{NOT}(\mathbf{w}), \mathbf{w})$
	Double Negation	$\neg(\neg w) = w$	$r_2 = \sum_{w \in W} 1 - Sim(\text{NOT}(\text{NOT}(\mathbf{w})), \mathbf{w})$
AND	Identify	$w \wedge T = w$	$r_3 = \sum_{w \in W} 1 - Sim(\text{AND}(\mathbf{w}, \mathbf{T}), \mathbf{w})$
	Annihilator	$w \wedge F = F$	$r_4 = \sum_{w \in W} 1 - Sim(\text{AND}(\mathbf{w}, \mathbf{F}), \mathbf{F})$
	Idempotence	$w \wedge w = w$	$r_5 = \sum_{w \in W} 1 - Sim(\text{AND}(\mathbf{w}, \mathbf{w}), \mathbf{w})$
	Complementation	$w \wedge \neg w = F$	$r_6 = \sum_{w \in W} 1 - Sim(\text{AND}(\mathbf{w}, \text{NOT}(\mathbf{w})), \mathbf{F})$
OR	Identify	$w \vee F = w$	$r_7 = \sum_{w \in W} 1 - Sim(\text{OR}(\mathbf{w}, \mathbf{F}), \mathbf{w})$
	Annihilator	$w \vee T = T$	$r_8 = \sum_{w \in W} 1 - Sim(\text{OR}(\mathbf{w}, \mathbf{T}), \mathbf{T})$
	Idempotence	$w \vee w = w$	$r_9 = \sum_{w \in W} 1 - Sim(\text{OR}(\mathbf{w}, \mathbf{w}), \mathbf{w})$
	Complementation	$w \vee \neg = T$	$r_{10} = \sum_{w \in W} 1 - Sim(\text{OR}(\mathbf{w}, \text{NOT}(\mathbf{w})), \mathbf{T})$

Neural Modules with Logical Regularization

- Add Logical Regularizers

$$L_1 = L_{loss} + \boxed{\lambda_l R_l} = L_{loss} + \lambda_l \sum_i r_i$$

- Avoid overfitting on the length of vectors

$$\begin{aligned} L_2 &= L_{loss} + \lambda_l R_l + \boxed{\lambda_\ell R_\ell} \\ &= L_{loss} + \lambda_l \sum_i r_i + \lambda_\ell \sum_{w \in W} ||\mathbf{w}||_F^2 \end{aligned}$$

Neural Modules with Logical Regularization

- Avoid overfitting on the length of vectors

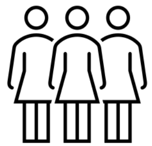
$$
\begin{aligned}
L_2 &= L_{loss} + \lambda_l R_l + \boxed{\lambda_\ell R_\ell} \\
&= L_{loss} + \lambda_l \sum_i r_i + \lambda_\ell \sum_{w \in W} ||\mathbf{w}||_F^2
\end{aligned}
$$

- Avoid overfitting on the number of parameters:

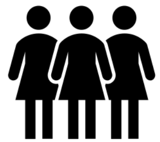
$$
\begin{aligned}
L &= L_{loss} + \lambda_l R_l + \lambda_\ell R_\ell + \boxed{\lambda_\theta R_\theta} \\
&= L_{loss} + \lambda_l \sum_i r_i + \lambda_\ell \sum_{w \in W} ||\mathbf{w}||_F^2 + \lambda_\theta ||\theta||_F^2
\end{aligned}
$$

Experiment Datasets

Murena’s Dataset

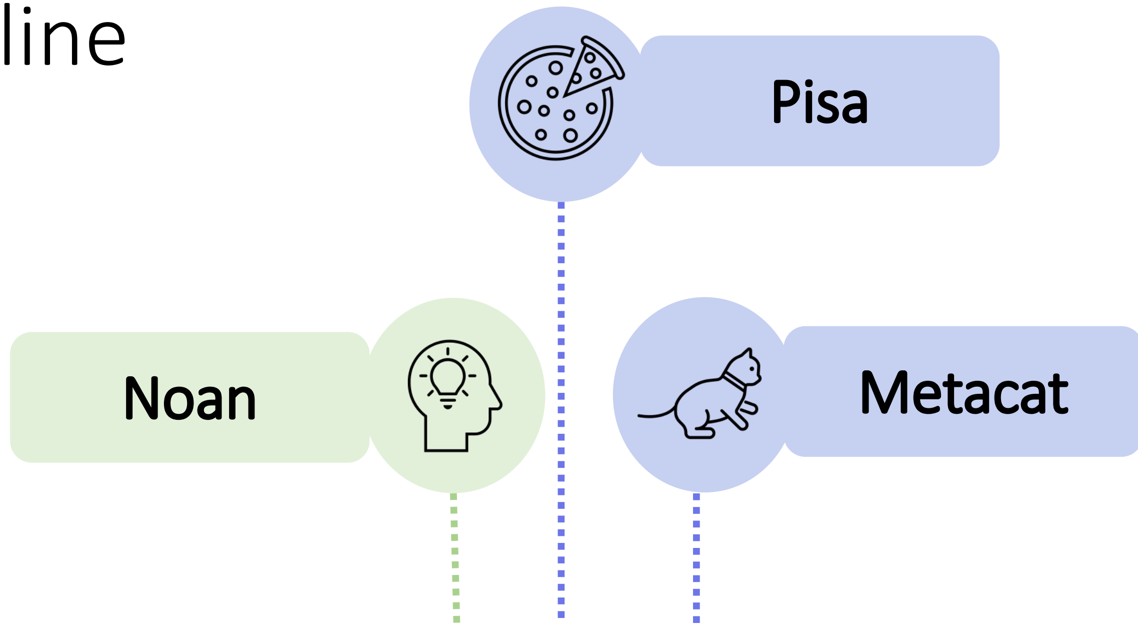


Rijsdijk’s Dataset



Year	2017	2020
Complexity	Easy	Hard
One-shot	ABC : ABD	Different
# Participants	68	35
# Questions	11	20

Experiment Baseline



Given X	Solutions	Selected	P_n	P_p	P_m
IJK	IJL	93%	1	1	1
	IJD	2.9%	2	∞	∞

Percentage

Rank

Results for Murena’s Dataset

- Given the same one-shot data ABC:ABD::X:?

Match Rate	Noan	Pisa	Metacat
Top 1	72.7%	72.7%	72.7%
Top 2	$\boxed{100\%}$	90.9%	90.9%

Given X	Solutions	Selected	P_n	P_p	P_m
IJK	IJL	93%	1	1	1
	IJD	2.9%	2	∞	∞
BCA	BCB	49%	1	3	2
	BDA	43%	2	1	1
AABABC	AABABD	74%	1	1	1
	AACABD	12%	2	∞	∞
IJKLM	IJKLN	62%	1	1	1
	IJLLM	15%	2	∞	∞
KJI	KJJ	37%	1	1	1
	LJI	32%	2	∞	2
ACE	ACF	63%	1	1	1
	ACG	8.9%	7	∞	∞
BCD	BCE	81%	2	2	2
	BDE	5.9%	1	1	1
IJJKKK	IJJLLL	40%	1	1	1
	IJJKKL	25%	2	2	2
XYZ	XYA	85%	2	1	1
	IJD	4.4%	11	∞	∞
RSSTTT	RSSUUU	41%	1	1	1
	RSSTTU	31%	2	2	∞
MRRJJJ	MRRJJK	28%	2	2	1
	MRRKKK	19%	1	1	2

Results for Murena’s Dataset

Same Format
Different Choices from Human

Given X	Solutions	Selected	P_n	P_p	P_m
IJK	IJL	93%	1	1	1
	IJD	2.9%	2	∞	∞
BCA	BCB	49%	1	3	2
	BDA	43%	2	1	1
AABABC	AABABD	74%	1	1	1
	AACABD	12%	2	∞	∞
IJKLM	IJKLN	62%	1	1	1
	IJLLM	15%	2	∞	∞
KJI	KJJ	37%	1	1	1
	LJI	32%	2	∞	2
ACE	ACF	63%	1	1	1
	ACG	8.9%	7	∞	∞
BCD	BCE	81%	2	2	2
	BDE	5.9%	1	1	1
IJJKKK	IJJLLL	40%	1	1	1
	IJJKKL	25%	2	2	2
XYZ	XYA	85%	2	1	1
	IJD	4.4%	11	∞	∞
RSSTTT	RSSUUU	41%	1	1	1
	RSSTTU	31%	2	2	∞
MRRJJJ	MRRJJK	28%	2	2	1
	MRRKKK	19%	1	1	2

Results for Rijsdijk's Dataset

Match Rate	Noan	Pisa	Metacat
Top 1	90%	55%	30%
Top 2	95%	65%	40%

Given problem	Solutions	Selected	P_n	P_p	P_m
ABA:ACA::	AEA	97.1%	1	1	1
ADA:?	AFA	2.9%	2	∞	∞
ABAC:ADAE::	DAEA	60%	2	2	∞
BACA:?	BCCC	28.6%	1	21	∞
AE:BD::	DB	68.5%	1	3	1
CC:?	CC	17.1%	2	∞	2
ABBB:AAAB::	IIJJJ	57.1%	1	1	∞
IIIJJ:?	JJIII	14.3%	2	∞	∞
ABC:CBA::	IJKLM	88.6%	1	1	1
MLKJI:?	-	-	∞	∞	∞
ABCB:ABCB::	Q	100%	1	1	∞
Q:?	-	-	∞	∞	∞
ABC:BAC::	JIKL	54.3%	2	∞	∞
IJKL:?	KIJL	14.3%	3	2	∞
ABACA:BC::	AA	57.1%	1	1	∞
BACAD:?	BCD	31.4%	3	∞	∞
AB:ABC::	IJKLM	85.7%	1	1	1
IJKL:?	IJKLMN	11.4%	2	∞	∞

Ability of Noan to Recognize Swaps

Given problem	Solutions	Selected	P_n	P_p	P_m
ABA:ACA:: ADA:?	AEA AFA	97.1% 2.9%	1 2	1 ∞	1 ∞
ABAC:ADAE:: BACA:?	DAEA BCCC	60% 28.6%	2 1	2 21	∞ ∞
AE:BD:: CC:?	DB CC	68.5% 17.1%	1 2	3 ∞	1 2
ABBB:AAAB:: IIIJJ:?	IIJJJ JJIII	57.1% 14.3%	1 2	1 ∞	∞ ∞
ABC:CBA:: MLKJI:?	IJKLM -	88.6% -	1 ∞	1 ∞	1 ∞
ABCB:ABCB:: Q:?	Q -	100% -	1 ∞	1 ∞	∞ ∞
ABC:BAC:: IJKL:?	JIKL KIJL	54.3% 14.3%	2 3	∞ 2	∞ ∞
ABACA:BC:: BACAD:?	AA BCD	57.1% 31.4%	1 3	1 ∞	∞ ∞
AB:ABC:: IJKL:?	IJKLM IJKLMN	85.7% 11.4%	1 2	1 ∞	1 ∞
ABC:ABBACCC:: FED:?	FEEFDDD -	91.4% -	1 ∞	2 ∞	1 ∞

ABC:BBC:: IKM:?	JKM KKM	57.1% 37.1%	1 2	7 2	∞ ∞
ABAC:ACAB:: DEFG:?	DGFE FGDE	68.6% 14.3%	1 2	2 1	∞ ∞
ABC:ABD:: CBA:?	DBA CBB	51.4% 45.7%	1 2	1 2	2 1
ABAC:ADAE:: FBFC:?	FDFE FDFA	94.3% 2.9%	1 6	1 ∞	∞ ∞
ABCD:CDAB:: IJKLMN:?	LMNIJK -	80.0% -	1 ∞	∞ ∞	∞ ∞
ABC:AAABBBCCC:: ABCD:?	AAABBBCCCDDD AAAABBBBCCCCDDDD	74.3% 17.1%	1 2	1 ∞	1 ∞
ABC:ABBCCC:: ABCD:?	ABBCCCDDDD ABBCCCDDD	85.7% 8.6%	1 2	∞ 1	∞ ∞
ABBCCC:DDDEEF:: AAABBC:?	DEEFFF DCCDDF	77.1% 8.6%	1 3	1 ∞	∞ ∞
A:AA:: AAA:?	AAAAAA AAAA	62.8% 25.7%	1 2	1 2	∞ 1
ABBA:BAAB:: IJKL:?	JILK JIJM	71.4% 11.4%	1 2	∞ 5	∞ ∞

Ability of Noan to Recognize Duplicates

Given problem	Solutions	Selected	P_n	P_p	P_m
ABA:ACA::	AEA	97.1%	1	1	1
ADA:?	AFA	2.9%	2	∞	∞
ABAC:ADAE::	DAEA	60%	2	2	∞
BACA:?	BCCC	28.6%	1	21	∞
AE:BD::	DB	68.5%	1	3	1
CC:?	CC	17.1%	2	∞	2
ABBB:AAAB::	IIJJJ	57.1%	1	1	∞
IIIJJ:?	JJIII	14.3%	2	∞	∞
ABC:CBA::	IJKLM	88.6%	1	1	1
MLKJI:?	-	-	∞	∞	∞
ABCB:ABCB::	Q	100%	1	1	∞
Q:?	-	-	∞	∞	∞
ABC:BAC::	JIKL	54.3%	2	∞	∞
IJKL:?	KIJL	14.3%	3	2	∞
ABACA:BC::	AA	57.1%	1	1	∞
BACAD:?	BCD	31.4%	3	∞	∞
AB:ABC::	IJKLM	85.7%	1	1	1
IJKL:?	IJKLMN	11.4%	2	∞	∞
ABC:ABBACCC::	FEEFDDD	91.4%	1	2	1
FED:?	-	-	∞	∞	∞

ABC:BBC::	JKM	57.1%	1	7	∞
IKM:?	KKM	37.1%	2	2	∞
ABAC:ACAB::	DGFE	68.6%	1	2	∞
DEFG:?	FGDE	14.3%	2	1	∞
ABC:ABD::	DBA	51.4%	1	1	2
CBA:?	CBB	45.7%	2	2	1
ABAC:ADAE::	FDFE	94.3%	1	1	∞
FBFC:?	FDFA	2.9%	6	∞	∞
ABCD:CDAB::	LMNIJK	80.0%	1	∞	∞
IJKLMN:?	-	-	∞	∞	∞
ABC:AAABBBCCC::	AAABBBCCCDDD	74.3%	1	1	1
ABCD:?	AAAABBBBCCCCDDDD	17.1%	2	∞	∞
ABC:ABBCCC::	ABBCCCDDDD	85.7%	1	∞	∞
ABCD:?	ABBCCCDDD	8.6%	2	1	∞
ABBCCC:DDDEEF::	DEEFFF	77.1%	1	1	∞
AAABBC:?	DCCDDF	8.6%	3	∞	∞
A:AA::	AAAAAA	62.8%	1	1	∞
AAA:?	AAAA	25.7%	2	2	1
ABBA:BAAB::	JILK	71.4%	1	∞	∞
IJKL:?	JIJM	11.4%	2	5	∞

Conclusion

- Experiments showed that our machine learning based approach Noan performs well on standard letter-string analogy datasets in different aspects.

- The Noan framework may also be used to solve more complex analogy problems such as word analogy in the future.