Compositional generalization through meta sequence-to-sequence learning
Brenden M. Lake, NYU

Rafael Elberg

Departamento de Ciencias de la Computación, PUC Chile

September 26, 2023

Contenidos

Composicionalidad

- Composicionalidad: "The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined."(Partee, 2004)
- "If I told you that a schmister was a sister over the age of 10 but under the age of 21, perhaps giving you a single example, you could immediately infer whether you had any schmisters, whether your best friend had a schmister, whether your children or parents had any schmisters, and so forth."(Marcus, 2018).
- Es muy complejo de obtener este tipo de razonamiento con modelos de DL.Lake, 2019

Método

- Idea: Usar memoria externa y meta-training para aprender templates y variables.
- Transformar el problema a templates y variables, recordando la información útil del espacio de soporte.

Método

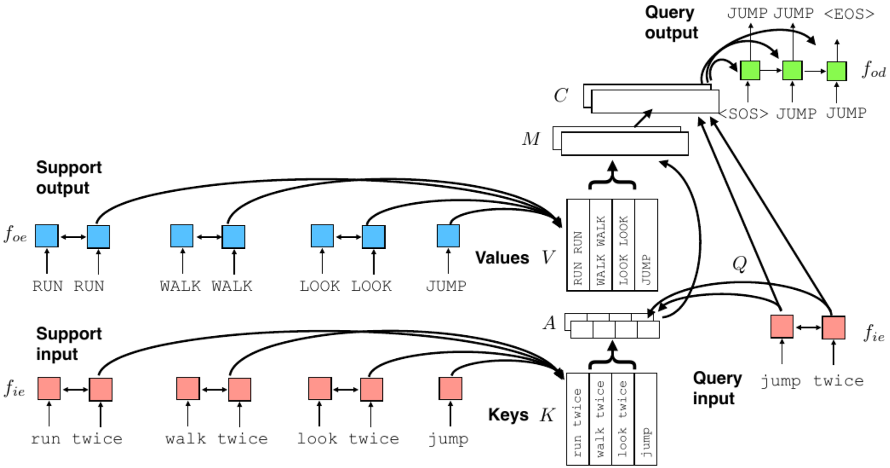

Figure: Modelo Meta seq2seq Learner

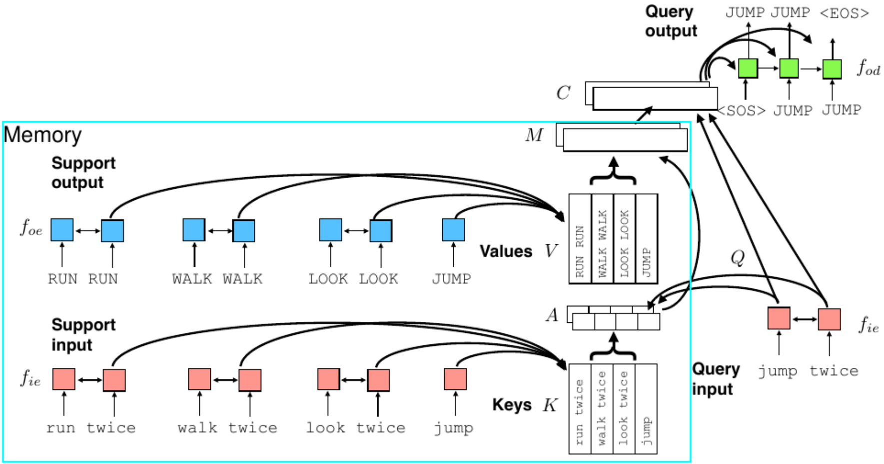

Figure: Memory module

Método

- BiLSTM encoder f_{ie} transforma support input a embeddings en Keys y query input a embeddings en Query.
- BiLSTM encoder f_{io} transforma support output a embeddings en Values.
- LSTM decoder f_{do} transforma Contexto a espacio de outputs, con información recibida de f_{ie} con un mecanismo de atención.
- Framework de entrenamiento meta-training con support loss.

Experimentos

Exclusividad mutua

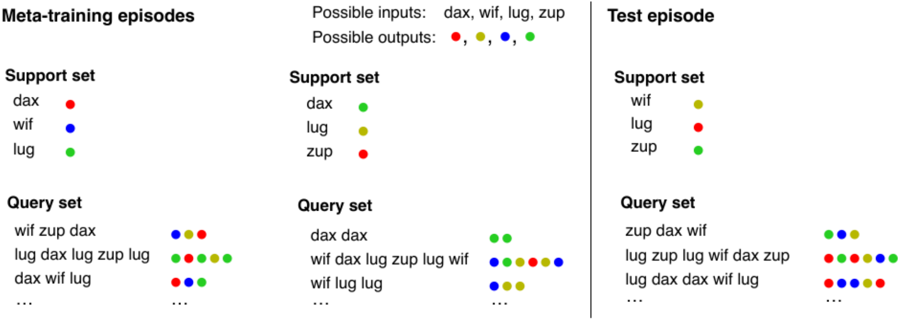

Figure: Problema de concatenación y razonamiento por exclusividad mutua.

Exclusividad mutua

- Cada episodio consiste en un mapeo aleatorio de cuatro pseudo palabras a cuatro colores.
- El set de soporte contiene el mapeo de tres de estas palabras.
- El set de queries contiene concatenaciones aleatorias de las cuatro palabras.
- El modelo obtiene un 100% de accuracy.

Exclusividad mutua

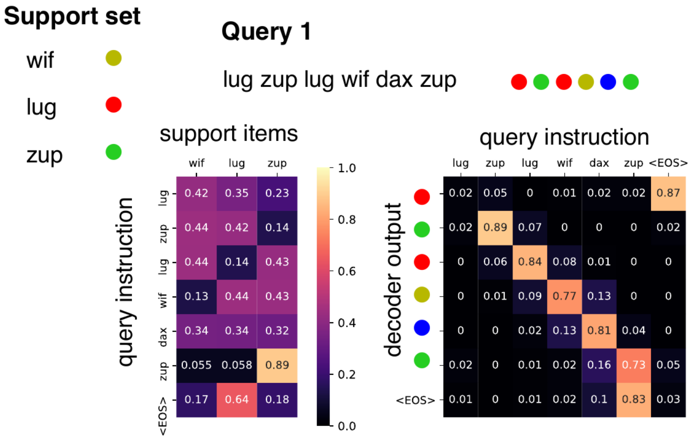

Figure: Mapas de atención ME.

SCAN add jump

jump	⇒	JUMP
jump left	⇒	LTURN JUMP
jump around right	⇒	RTURN JUMP RTURN JUMP RTURN JUMP RTURN JUMP
turn left twice	⇒	LTURN LTURN
jump thrice	⇒	JUMP JUMP JUMP
jump opposite left and walk thrice	⇒	LTURN LTURN JUMP WALK WALK WALK
jump opposite left after walk around left	⇒	LTURN WALK LTURN WALK LTURN WALK LTURN WALK LTURN LTURN JUMP

Figure: SCAN add jump test cases

SCAN add jump

- El objetivo es aprender el uso de la palabra jump.
- Seq2Seq clásico: Ve traducción jump JUMP y frases con los otros verbos.
- Seq2Seq meta train: Entrenamiento con permutaciones aleatorias de (‘jump’,‘run’, ‘walk’, ‘look’) y (‘JUMP’,‘RUN’,‘WALK’,‘LOOK’).
- Seq2Seq meta test: De soporte recibe sólo la permutación correcta de (‘jump’,‘run’, ‘walk’, ‘look’) y (‘JUMP’,‘RUN’,‘WALK’,‘LOOK’), y debe predecir un query de SCAN.
- Se hace el mismo experimento aumentando los datos con 20 nuevos priors y acciones.

SCAN add jump

Model	standard training	permutation meta-training	augmentation meta-training
meta seq2seq learning	—	**99.95%**	**98.71%**
-without support loss	—	5.43%	**99.48%**
-without decoder attention	—	10.32%	9.29%
standard seq2seq	0.03%	—	12.26%
syntactic attention [30]	78.4%	—	—

Figure: Resultados del test SCAN

SCAN length and around right

Model	around right	length
meta seq2seq learning	**99.96%**	16.64%
standard seq2seq	0.0%	7.71%
syntactic attention [30]	28.9%	15.2%

Figure: Resultados de tests largos y around right

Conclusiones

- Mediante meta training y testing, se el modelo logra generalizar composicionalmente, empleando lógica "algebraica" sobre el dominio.
- El método logra resolver problemas que anteriormente no eran atacables por métodos de NLP (Como la composicionalidad en SCAN), pero no logra generalizar sobre secuencias más largas.
- El método de memoria externa requiere conocimiento del dominio y recorrido, por lo que no logra extrapolar fuera del espacio de entrenamiento.

Críticas

- ✓ Una idea brillante y bien explicada.
- $\times$ Habla muy poco del SOTA.
- $\times$ No hay comparación en el problema de ME.
- $\times$ Poco contexto lingüístico, a diferencia de Russin et al., 2019.

Lake, B. M. (2019). Compositional generalization through meta sequence-to-sequence learning. *CoRR, abs/1906.05381*. http://arxiv.org/abs/1906.05381

Marcus, G. F. (2018). Deep learning: A critical appraisal. *ArXiv, abs/1801.00631*. https://api.semanticscholar.org/CorpusID:1872638

Partee, B. H. (2004). *Compositionality in formal semantics*. Blackwell Publishing Ltd. https://web.mit.edu/jda/www/teaching/6.884/readings/partee_1984.pdf

Russin, J., Jo, J., O'Reilly, R. C., & Bengio, Y. (2019). Compositional generalization in a deep seq2seq model by separating syntax and semantics. *CoRR, abs/1904.09708*. http://arxiv.org/abs/1904.09708