Relational inductive biases, deep learning, and graph networks

Battaglia et al.

https://arxiv.org/abs/1806.01261

Robert Różański
ML Gdańsk 29.10.2018
Deep Learning challenges

complex language and scene understanding
reasoning about structured data
transfer learning beyond the learning conditions
learning from small amount of experience
unstructured approaches

weak assumptions

low inductive bias

high data requirements

high computation requirements

transfer learning (limited)
structured approaches

- strong a priori assumptions about data structures and computation
- high inductive bias
- low data requirements
- low computation requirements
- combinatorial generalisation

unstructured approaches

- weak assumptions
- low inductive bias
- high data requirements
- high computation requirements
- transfer learning (limited)
<table>
<thead>
<tr>
<th>Structured Approaches</th>
<th>Unstructured Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong a priori assumptions about data structures and computation</td>
<td>Weak assumptions</td>
</tr>
<tr>
<td>High inductive bias</td>
<td>Low inductive bias</td>
</tr>
<tr>
<td>Low data requirements</td>
<td>High data requirements</td>
</tr>
<tr>
<td>Low computation requirements</td>
<td>High computation requirements</td>
</tr>
<tr>
<td>Combinatorial generalisation</td>
<td>Transfer learning (limited)</td>
</tr>
</tbody>
</table>
Which structures to choose?
Which structures to choose?

entities

relations

properties
(a) Molecule

(b) Mass-Spring System

(c) $n$-body System

(d) Rigid Body System

(e) Sentence and Parse Tree

(f) Image and Fully-Connected Scene Graph

The brown dog jumped.
order invariance:
predicting center of mass
order invariance: predicting center of mass

e.g. for MLP order matters (n! permutations)
order invariance: predicting center of mass

e.g. for MLP order matters (n! permutations)

Forcing invariance:

1. compute features (per planet)

2. aggregate using order-invariant function
pairwise relations: predicting planet's position

planet's future position depends on parameters of all other planets:
pairwise relations:
predicting planet's position

planet's future position depends on parameters of all other planets:

\[ x'_i = f(x_i, \sum_j g(x_i, x_j)) \]
no connections

fully connected
no connections

fully connected

\[ x'_i = f(x_i, \sum_j g(x_i, x_j)) \]
no connections

intermediate

$\mathbf{x}'_i = f(\mathbf{x}_i, \sum_{j \in \delta(i)} g(\mathbf{x}_i, \mathbf{x}_j))$

fully connected

$\mathbf{x}'_i = f(\mathbf{x}_i, \sum_j g(\mathbf{x}_i, \mathbf{x}_j))$
(a) Edge update

\[ e'_k = \phi^e (e_k, v_{rk}, v_{sk}, u) \]
\[
\begin{align*}
\mathbf{e}_k' &= \phi^e (\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}) \\
\mathbf{v}_i' &= \phi^v (\mathbf{e}_i', \mathbf{v}_i, \mathbf{u}) \\
\mathbf{e}_i' &= \rho^{e\rightarrow v} (E_i')
\end{align*}
\]
(a) Edge update

\[ \mathbf{e}_k' = \phi^e \left( \mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u} \right) \]
\[ \mathbf{v}_i' = \phi^v \left( \overline{\mathbf{e}}_i', \mathbf{v}_i, \mathbf{u} \right) \]
\[ \mathbf{u}' = \phi^u \left( \overline{\mathbf{e}}', \overline{\mathbf{v}}', \mathbf{u} \right) \]

(b) Node update

\[ \overline{\mathbf{e}}_i' = \rho^{e\rightarrow v} \left( E_i' \right) \]
\[ \overline{\mathbf{e}}' = \rho^{e\rightarrow u} \left( E' \right) \]
\[ \overline{\mathbf{v}}' = \rho^{v\rightarrow u} \left( V' \right) \]