Risi Kondor.

papers by Risi Kondor. These ideas will be really helpful for our setup. One is "Covariant Compositional Networks for Learning Graphs", where it introduces permutation covariance tensor for different message passing scheme.

Covariant Compositional networks ~~for~~ learning graphs.

Limitat^ns

→ fixed length repr^n regardless of size of graph.

→ represent^n invariant to permutations of graph's vertices

→ graph kernels, ⇒ to get multi-level repr^n of a graph through a series of message passing & hashing st

→ permutation invariant vertex aggregation operation performed by most GNN limit their expressive power

---

* Covariant compositional Network (CCN)

→ Represent vertex features with higher covariance tensors

→ structured tensor representation

→ Permutation covariance

Compositional Networks

↳ Represent complex objects as comb^n of their parts

object G : ⟶ n elementary parts (atoms)

$\mathcal{E} = \{e_1 \cdots e_n\}$

Composit^n scheme for G is DAG $\mathcal{M}$ in which each node v is associated with subsets P_v of $\mathcal{E}$

subsets are called " parts of G "

* Comp-Net (N)

① Bottom level :- n leaf nodes
each leaf node v is associated with elem. atom e
P_v contains a single atom e.

② $\mathcal{M} \rightarrow$ unique root node v_r corresponds to entire set $\{e_1 \cdots e_n\}$

③ two nodes v, v' :- if v is a descendent of v'
$\Rightarrow$ P_v is a subset of $P_{v'}$

Message passing NN $G = (V, E)$ in $L+1$ layer network

[set of vertices V is also the set of elementary $\mathcal{E}$

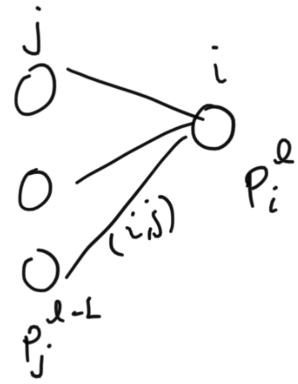
v f feature tensors for each vertex of G

layer $l = 0$:- $v_i^0 \rightarrow$ single vertex $P_i^0 = \{i\}$ for $i \in V$
$f_i^0 \rightarrow l_i$ (vertex label)

$l = 1 \cdots L$ v_i^l — connected to all nodes from prev. level

children of $v_i^l = \{v_j^{l-1} \mid j = (i,j) \in E\}$

$$p_i^{l} = \bigcup_{j:(i,j)\in E} p_j^{l-1}$$



feature tensor, $f_i^{l} = \underbrace{\phi}_{\text{aggregat}^n\ f^n.}\left(\{ f_j^{l-1} \mid j \in p_i^{l} \}\right)$

<u>Layer $l = L+1$</u> : single node v_r – repr^n entire graph &

$p_r \equiv V$

$f_r = \phi\left(\{ f_i^{L} \mid i \in p_r \}\right)$ — collects info^m from all nodes at level L

$*$ $\begin{cases} v : & \text{neuron} \\ p : & \text{receptive field} \\ f : & \text{activation} \end{cases}$

3) <u>Covariance</u> :-

Standard message passing NN → Summation / averaging as ϕ

↳ can't capture any infor^m abt the connectivity of vertices' neighborhood.

"Permutation Covariance" (desirable for our neural activations f

G, comp-Net $\mathcal{N}$ | isomorphic G', $\mathcal{N}'$

v · neuron | v' - correspond neuron

$P_v = (e_{p_1} \cdots e_{p_m})$ | $P_{v'} = (e_{q_1} \cdots e_{q_m})$

$\pi \in S_m$ be "permutation" that aligns the orderings of the two receptive fields.

$$\boxed{e_{q_{\pi(a)}} = e_{p_a}}$$

$\mathcal{N}$ is covariant to permutations, if ~~for~~ any π

$\exists$ corresponding R_π s.t $f_{v'} = R_\pi(f_v)$

permuting (π) vertices of graph G $\Rightarrow$ change activations by some fixed f^n R_π

depend on permutat^n

Message Passing :-

zeroth order MP $\rightarrow$ each vertex · feature length of c

sum together vertex features

(lose identity infor^n on where certain vertex features originated from.)

1st order · each vertex . v, $f_v^l \in \mathbb{R}^{|P_v^l| \times c}$

each row corresponds to a vertex in neighborhood of v.

1st order covariant node →

Permut^n of receptive field P_v by any $\pi \in S_{|P_v|}$

activat^n transforms as $\boxed{f_v \mapsto P_\pi f_v}$

Tensor magnitude of components, m

basis vector, n n^m

* Coordinate transformation :- magnitude change by L
invariant under (L) basis vector transform a/c L^{-1}
(tensors remain same)

* **Einstein Notation** a^i – a super i

$$\sum_{i=1}^{3} a_i x_i \xrightarrow{E.N} a_i x_i$$

$$a_{ij} b_j = \sum_j a_{ij} b_j$$

j : dummy index. (summed)

i : free index (Not summed or occurs once single.

→ No index may occur 3+ in a given single term.

$\cancel{a_{ii} b_{ij}}$

a_j^j : J is dummy index

a_i^{ij} – i = free, j = dummy

* free index same on both sides

$$x_i = a_{ij} b_j$$

$$a_i = A_{ki} B_{kj} x_j + C_{ik} u_k$$ free : i
dummy = j, k.

tensor component transform in a

$a^i \rightarrow$ contravariant manner $\rightarrow$ opposite manner as basis vectors.
$b_j \rightarrow$ covariant manner $\rightarrow$ as same " " " "

Contravariant :- $$\bar{V}^i = V^r \frac{\partial \bar{x}^i}{\partial x^r} \quad (1 \leq i \leq n)$$

Contraction

$$S = \left(S^{i_1 i_2 \cdots i_p}_{j_1 \cdots j_q} \right)$$ ← contravariant rank
← covariant rank

p+q = rank

↓ contract^n

$$S^{i_1 i_2 \cdots u \cdots i_p}_{j_1 j_2 \cdots u \cdots j_q}$$

repl $i_f = j_g = u$.

contract^n w.r.t
contravariant index i_f
covariant index j_g

↓

(p-1, q-1) rank tensor

Inner Product

$$S \cdot T = \left(S^{i_1 \cdots u, \cdots i_p}_{j_1 j_2 \cdots u \cdots j_q} \right) \cdot \left(T^{k_1 \cdots u \cdots k_r}_{l_1 \cdots u \cdots l_s} \right)$$

↓

tensor of rank $p + q + r + s \underbrace{-2}_{\text{summed over}}$

outer product
- prod

Tensor $S \otimes T$: No summed over like in ·

Kronecker Direct wikipedia.

rank $= (p+q+r+s)$ tensor.

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}.$$

$$a \otimes b = \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & \cdot & \cdot \\ a_3 b_1 & & a_3 b_3 \end{bmatrix}$$

Prev. models $\rightarrow$ Instances of zeroth order message passing, where each vertex repr^n is a vector of C-channels

$\Downarrow$

loss of structural info during message aggregation step.

1st order CCN / 2nd order CCN
(CCN 1D) (CCN 2D)

$G = (V, E)$

$L+1$ levels CCN.

$(0, \cdots L)$

0:- vertex $v \longrightarrow$ input feature vector $l_v \in \mathbb{R}^C$

receptive field of v at level l : P_v^l

$$P_v^l \triangleq \begin{cases} \{v\} & l = 0 \\ \bigcup\limits_{(u,v) \in E} P_u^l & l = 1 \cdots L \end{cases}$$

feature tensor F_v^{l} ← level ← vertex

for 0-th order CCN $F_v^l \in \mathbb{R}^c$

N : # of vertices in P_v^l.

1^{st} order CCN:- each vertex is represented by a matrix

$$F_v^l \in \mathbb{R}^{N \times C} \quad \left(2^{nd} \text{ order tensor}\right)$$

each row corresp to vertex in receptive field P_v^l

each channel is N-size vector (1^{st} order P-tensor)

2^{nd} order CCN:-

$$F_v^l \in \mathbb{R}^{N \times N \times c}$$

→ each channel is a second order P-tensor

The tensor F_v^l transforms in a covariant way w.r.t permutations of the vertices in P

* Constructing the aggregation function ϕ

(doing this in a way that preserves covariance is to "promote - stack - reduce"

Promote :- Not all vertices in a given receptive field have same sized tensor

index function χ : ensure all tensors are of same size by padding them zero

permut^{n} matrix, $\chi_\ell^{w \to v}$ $\quad |P_v^\ell| \times |P_w^{\ell-1}|$

$$\chi \cdot F_w^{\ell-1} \quad // \quad \chi \times F_w^{\ell-1} \times \chi^T$$

where $\chi = \chi_\ell^{w \to v}$.

Resized tensor denoted by $F_{w \to v}^{\ell}$

Stacked / Concatenated :-

- into a tensor one order higher
- Stacked index has same size as the receptive field.

Reduce :- Reduce higher order tensor down to expected size of vertex representat^{n}.

(Tensor Contraction)

learnable set of weights that reduces the

# of channels to C.

$\phi \rightarrow$ elementwise Non-linear fn γ (ReLU)

$F_v^l \rightarrow$ vectors of channel $\Theta(F_v^l) = F^l_{v\, i_1 \cdots i_p}$

Nonchannel inds

(single vector) sum up.