Graph Markov Neural Networks

et classification dans des graphes de documents

Séminaire TALia du 5 novembre 2021

[Source principale]

arXiv.org > cs > arXiv:1905.06214

Computer Science > Machine Learning

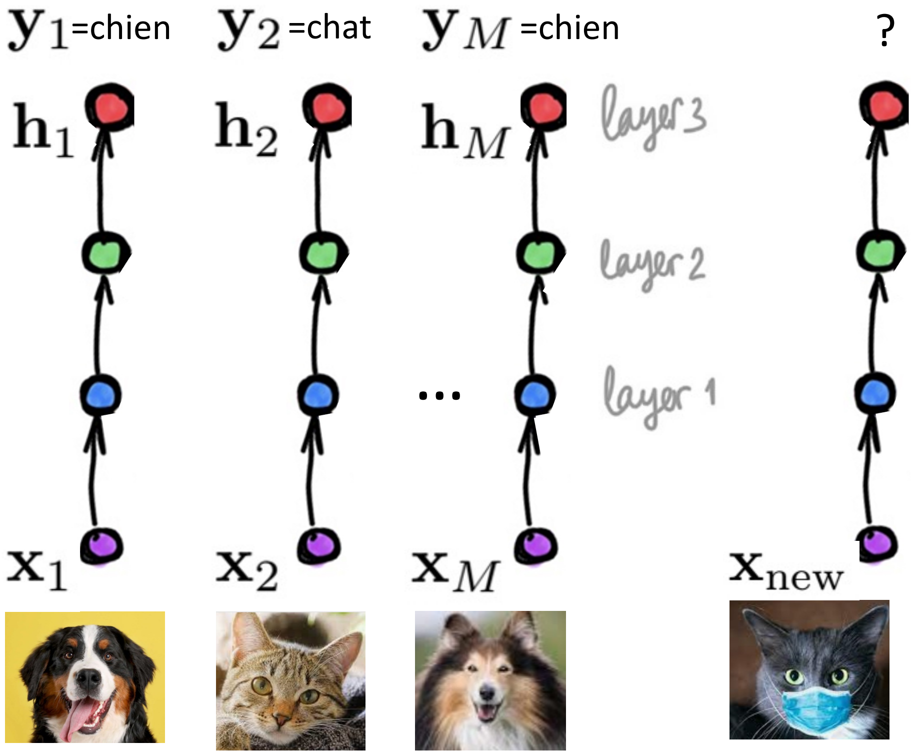
[Submitted on 15 May 2019 (v1), last revised 23 Jul 2020 (this version, v3)]

GMNN: Graph Markov Neural Networks

Meng Qu, Yoshua Bengio, Jian Tang

Intuitions sur les réseaux de neurones de graphe

Modèle inductif

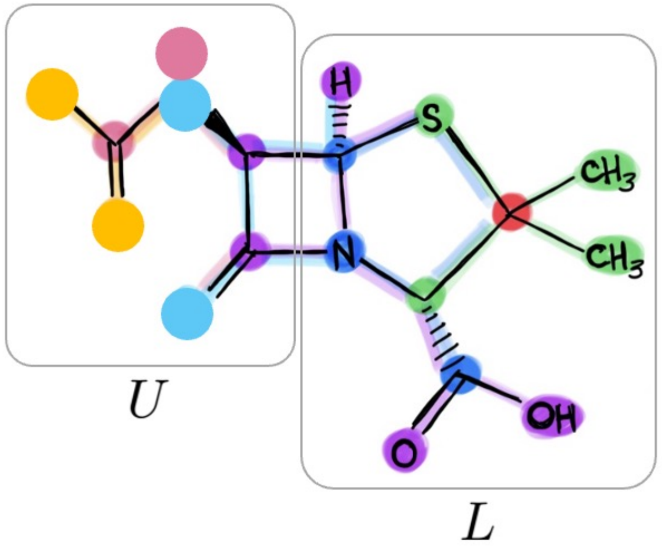


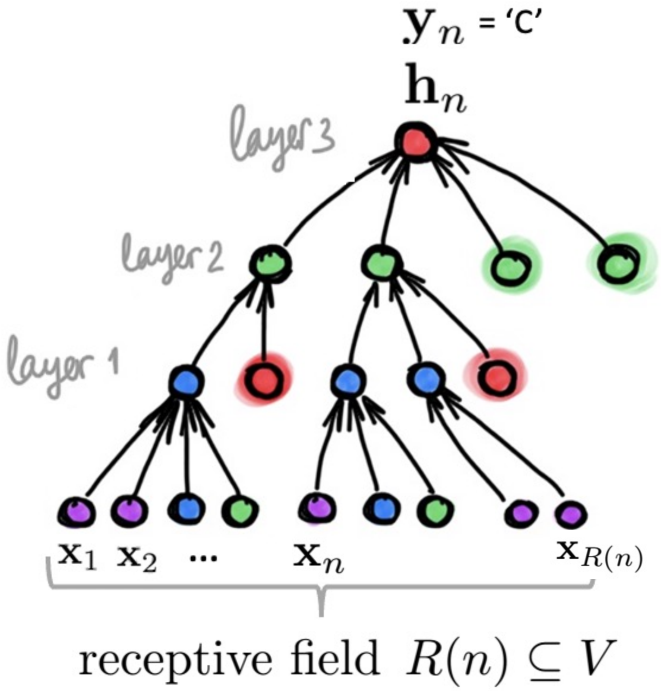
$$p(\mathbf{y}_i|\mathbf{x}_i) = \text{Cat}(\mathbf{y}_i|\text{softmax}(W\mathbf{h}_i)), \ \ \mathbf{h}_i = \mathbf{h}(\mathbf{x}_i))$$

$(\mathbf{x}_i, \mathbf{y}_i)$ sont i.i.d

Modèle transductif

$$G = (V, E, \mathbf{x}_V) \quad V = L \cup U$$





$$p(\mathbf{y}_n|\mathbf{x}_V) = \text{Cat}(\mathbf{y}_n|\text{softmax}(W\mathbf{h}_n)), \ \ \mathbf{h}_n = \mathbf{h}(\mathbf{x}_{R(n)}))$$

$(\mathbf{x}_i, \mathbf{y}_i)$ ne sont pas i.i.d

Construction itérative des représentations dans un GCN

motivation théorique

Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering

Michaël Defferrard Xavier Bresson Pierre Vandergheynst
EPFL, Lausanne, Switzerland
{michael.defferrard,xavier.bresson,pierre.vandergheynst}@epfl.ch

article

Un **schéma de propagation** simple pour les GCN

$$H^{(l+1)} = \sigma\left(\underbrace{\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}}_{\text{matrice d'adjacence normalisée avec self-connexion } \in R^{|V|\times|V|}} H^{(l)}\underbrace{W^{(l)}}_{\text{matrice de poids partagés entre tous les nœuds } \in R^{d_l \times d_{l+1}}}\right)$$

où

$$H^{(l)} := \begin{bmatrix} \left(\mathbf{h}_1^{(l)}\right)^\top \\ \left(\mathbf{h}_2^{(l)}\right)^\top \\ \vdots \\ \left(\mathbf{h}_{|V|}^{(l)}\right)^\top \end{bmatrix} \in \mathbb{R}^{|V|\times d_l}$$

Indépendance conditionnelle des étiquettes $\mathbf{y}_V$ étant donnés les attributs $\mathbf{x}_V$

$$p(\mathbf{y}_V|\mathbf{x}_V) = \prod_{n\in V} p(\mathbf{y}_n|\mathbf{x}_n)$$

Un GNN calcule des représentations des nœuds mais ignore complètement les dépendances entre labels !

Comment tenir compte des corrélations entre labels ?

L'expérience montre que la prise en compte des corrélations entre les étiquettes est **bénéfique** pour améliorer la précision de la classification !

Classifying Wikipedia in a fine-grained hierarchy : what graphs can contribute

Tiphaine Viard[1,2*], **Thomas McLachlan**[2], **Hamidreza Ghader**[2,3] and **Satoshi Sekine**[2]
[1]Laboratoire d'Informatique de Paris Nord (LIPN), Villetaneuse, France
[2]Language Information Access Team, Riken AIP, Tokyo, Japan
[3]ILPS, Amsterdam University
{thomas.mclachlan, hamidreza.ghader, satoshi.sekine}@riken.jp, viard@lipn.univ-paris13.fr

Modélisation « classique » par des **Conditional Random Fields** (CRF) :

Markov car indépendant des nœuds qui ne sont pas des voisins

$$p(\mathbf{y}_V|\mathbf{x}_V) = \frac{1}{Z(\mathbf{x}_V)} \prod_{(n,m)\in E} \psi(\mathbf{y}_m, \mathbf{y}_n; \mathbf{x}_V) \Rightarrow p(\mathbf{y}_n|\mathbf{y}_{V\setminus n};\mathbf{x}_V) = p(\mathbf{y}_n|\mathbf{y}_{\mathrm{NB}(n)};\mathbf{x}_V)$$

construit avec des features élaborées par des experts

On aimerait calculer $p(\mathbf{y}_U|\mathbf{y}_L;\mathbf{x}_V)$ … hélas **problème d'inférence difficile** à cause de $\frac{1}{Z(\mathbf{x}_V)}$

Approche variationelle par optimisation de ELBO (1)

Objectif des GMNN : Faire la synthèse entre les GNN's et les CRF's pour modéliser des graphes d'objets dont les labels sont corrélés.

« evidence »
variables latentes

1. On définit un modèle prédictif paramétrique $p_\phi(\mathbf{y}_L, \mathbf{y}_U | \mathbf{x}_V)$ (à choisir) et on essaie de maximiser la vraisemblance pour la distribution marginale $p_\phi(\mathbf{y}_L | \mathbf{x}_V)$ sur les labels observés.

2. On utilise la borne inférieure variationelle ELBO (Evidence Lower Bound)

distribution marginale sur les observations

L'écart = KL[q||p]>0

distribution conjointe

distribution variationelle

ELBO[q,θ]

$$\log p_\phi(\mathbf{y}_L | \mathbf{x}_V) \geq \mathbb{E}_{q_\theta(\mathbf{y}_U | \mathbf{x}_V)}[\log p_\phi(\underbrace{\mathbf{y}_L, \mathbf{y}_U}_{:= \mathbf{y}_V} | \mathbf{x}_V) - \log q_\theta(\mathbf{y}_U | \mathbf{x}_V)], \quad \forall q_\theta(\mathbf{y}_U | \mathbf{x}_V)$$

On a l'égalité si et seulement si $q_\theta(\mathbf{y}_U | \mathbf{x}_V) = p_\phi(\mathbf{y}_U | \mathbf{y}_L, \mathbf{x}_V)$ qu'on cherche justement à calculer !

3. On choisi un modèle judicieux pour la distribution variationelle $q_\theta(\mathbf{y}_U | \mathbf{x}_V)$.

Approche variationelle par optimisation de ELBO (2)

4. On alterne une étape d'inférence (E = Expectation) et une étape d'apprentissage (M = Maximization)

(E) Expectation

On fixe p_ϕ et on optimise les paramètres θ pour rapprocher $q_\theta(\mathbf{y}_U|\mathbf{x}_V)$ de $p_\phi(\mathbf{y}_U|\mathbf{y}_L, \mathbf{x}_V)$ c.à.d. minimiser $\mathrm{KL}(p_\phi \| q_\theta)$

Distribution cible

(M) Maximization

On fixe q_θ et on optimise les paramètres ϕ pour maximiser ELBO donc la log vraisemblance

$$\ell(\phi) := \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\log p_\phi(\mathbf{y}_L, \mathbf{y}_U|\mathbf{x}_V)]$$

Choix pour la distribution variationelle qui prédira les labels à partir des arguments :

$$q_\theta(\mathbf{y}_U|\mathbf{x}_V) := \prod_{n \in U} q_\theta(\mathbf{y}_n|\mathbf{x}_V)$$

Approximation du champ moyen

$$q_\theta(\mathbf{y}_n|\mathbf{x}_V) := \mathrm{Cat}(\mathbf{y}_n|\mathrm{softmax}(W_\theta \mathbf{h}_{\theta,n}(\mathbf{x}_V)))$$

Calculé avec un GNN choisi avec soin !

Choix pour la distribution CRF qui prédira les corrélations entre les labels:

$$p_\phi(\mathbf{y}_n|\mathbf{y}_{\mathrm{NB(n)}}; \mathbf{x}_V) = \mathrm{Cat}(\mathbf{y}_n|\mathrm{softmax}(W_\phi \mathbf{h}_{\phi,n}))$$

On montre que ça suffit !

Calculé avec un GCN basique

Les GMNN en une figure !

1 $q_\theta(\mathbf{y}_U|\mathbf{x}_V) := \prod_{n \in U} q_\theta(\mathbf{y}_n|\mathbf{x}_V)$ est initialisée avec un « bon » GNN comme p.ex. GCN avec réduction de variance

2