Graph-Driven Generative Models for Heterogeneous Multi-Task Learning

Wenlin Wang[1], Hongteng Xu[2], Zhe Gan[3], Bai Li[1], Guoyin Wang[1]
Liqun Chen[1], Qian Yang[1], Wenqi Wang[4] and Lawrence Carin[1]

[1]Duke University, [2]Infinia ML, [3]Microsoft Dynamics 365 AI Research, [4]Facebook

CONTRIBUTION

Propose a graph-driven generative model that unifies heterogeneous learning tasks into a single framework.

- A single graph structure shared by different generative processes.
- A GCN that embeds the nodes of the graph and specializes the organization and usage to different tasks.
- A universal solution for healthcare tasks, including clinical topic model, procedure recommendation and admission type prediction.

GRAPH-DRIVEN VAEs

Heterogeneous multi-task learning: We model heterogeneous multi-tasks with multiple generative processes

$$\boldsymbol{y}_k \sim p_{\theta_k}(\boldsymbol{y}_k|\boldsymbol{z}_k)\,, \quad \boldsymbol{z}_k \sim p(\boldsymbol{z}_k)\,, \quad k = 1, \dots, K. \qquad (1)$$

with corresponding inference networks specified as

$$\boldsymbol{z}_k \sim q_{\psi_k}(\boldsymbol{z}_k|f_\phi(\boldsymbol{x}_k))\,, \quad k = 1, \dots, K. \qquad (2)$$

The goal is to maximize the following evidence lower bound (ELBO)

$$\mathcal{L}(\theta_{1:K}, \psi_{1:K}, \phi) = \Sigma_k \Big[\mathbb{E}_{q_{\psi_k}(\boldsymbol{z}_k|f_\phi(\boldsymbol{x}_k))}[\log p_{\theta_k}(\boldsymbol{y}_k|\boldsymbol{z}_k)] - \mathrm{KL}(q_{\psi_k}(\boldsymbol{z}_k|f_\phi(\boldsymbol{x}_k)) \,\|\, p(\boldsymbol{z}_k))\Big]. \qquad (3)$$

Model tasks as sub-graphs: Since features are organized in different views and the interactions between observed entities can be different, we use a data graph $G(\mathcal{V}, \mathcal{X}, \boldsymbol{A})$ to model such data. A typical specification for healthcare tasks are summarized as follows

Task	G_k		$\boldsymbol{y}_k$
	$\mathcal{V}_k$	$x^k_{v^a_n}$ in $\mathcal{X}_k$	
Topic Modeling	$\mathcal{V}$	MaxPooling($\{\boldsymbol{x}_v\}_{v\in\mathcal{V}^d_n\cup\mathcal{V}^p_n}$)	Bi-term ICD codes
Procedure Recommendation	$\mathcal{V}^d \cup \mathcal{V}^a$	MaxPooling($\{\boldsymbol{x}_v\}_{v\in\mathcal{V}^d_n}$)	List of procedures
Admission-type Prediction	$\mathcal{V}$	MaxPooling($\{\boldsymbol{x}_v\}_{v\in\mathcal{V}^d_n\cup\mathcal{V}^p_n}$)	Admission type, $c \in \mathcal{C}$

Construction of Edges: Two statistics are considered as edges.
Edges between ICD codes Point-wise mutural information (PMI) is deployed as weight between each pair of ICD codes.
Edges between ICD codes and admissions TF-IDF is used as weight between admissions and ICD codes for each task.

MODEL ARCHITECTURE

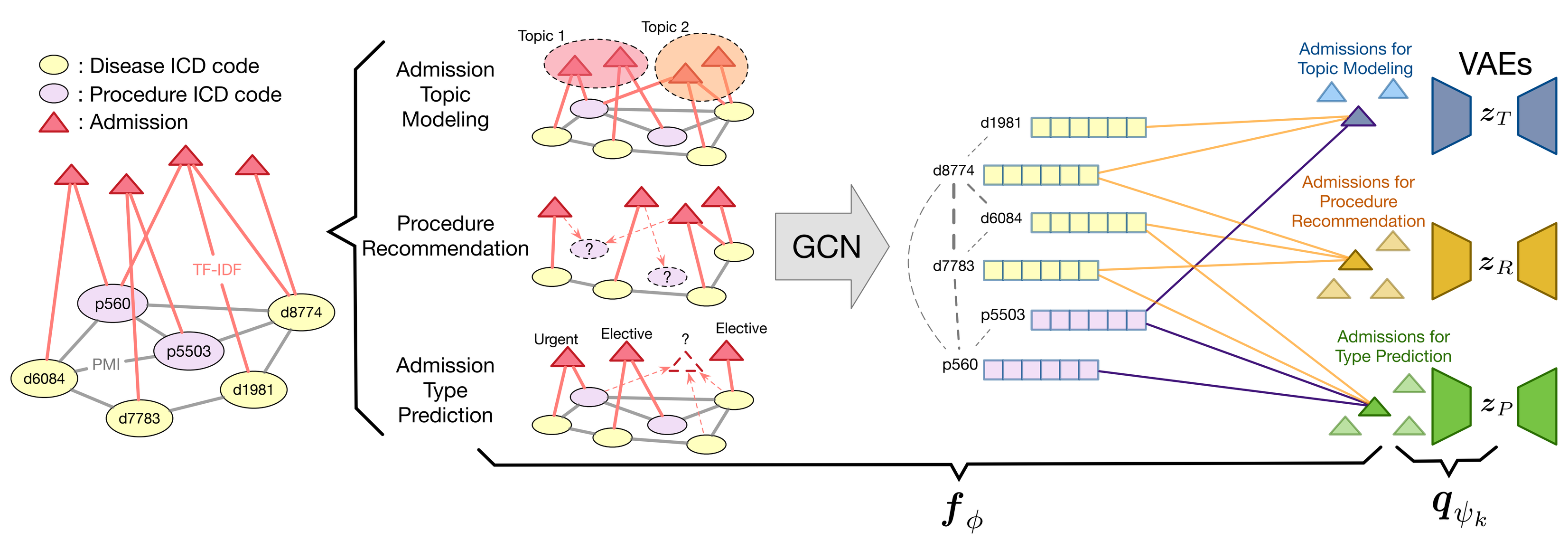

Figure: Illustration of the proposed model for healthcare tasks.

HETEROGENEOUS MULTI-TASKS

Topic Modeling of Admissions The generative process of our neural Bi-term Topic Model (NBTM) is

$$\boldsymbol{z}_T \sim \mathrm{Dir}(\boldsymbol{\alpha}), \quad l \sim \mathrm{Multi}(1, \boldsymbol{z}_T), \quad \boldsymbol{y}_T \sim \mathrm{Multi}(2, \boldsymbol{\beta}_l)\,, \qquad (4)$$

where $\boldsymbol{y}_T$ is the bi-term variable, v_i, v_j are two ICD codes, $\boldsymbol{z}_T$ is the topic distribution.

Procedure Recommendation We consider the following generative process for procedures $\boldsymbol{y}_R$

$$\boldsymbol{z}_R \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \pi_R \propto \exp\{g(\boldsymbol{z}_R)\}, \quad \boldsymbol{y}_R \sim \mathrm{Multi}(M, \pi_R)\,, \qquad (5)$$

where $\boldsymbol{y}_R$ is $|\mathcal{V}^p|$-dimensional variable and its instance is a list of M recommended procedures.

Admission-Type Prediction Given an admission, the goal is to predict the admission type given both its diseases and procedures. The corresponding generative process is:

$$\boldsymbol{z}_P \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \pi_P \propto \exp\{h(\boldsymbol{z}_P)\}, \quad \boldsymbol{y}_P \sim \mathrm{Multi}(1, \pi_P)\,, \qquad (6)$$

where $\boldsymbol{y}_P$ is a variable and its instance corresponds to an admission type in the set $\mathcal{C}$.

EXPERIMENTS

	Small	**Median**	**Large**
$\lvert\mathcal{V}^d\rvert$	247	874	2,765
$\lvert\mathcal{V}^p\rvert$	75	258	819
$\lvert\mathcal{V}^a\rvert$	28,315	30,535	31,213

Table: Statistics of the MIMIC-III dataset.

Topic Modeling

Method	Small			Median			Large		
	T=10	T=30	T=50	T=10	T=30	T=50	T=10	T=30	T=50
LDA	0.110	0.106	0.098	0.123	0.102	0.107	0.101	0.106	0.103
AVITM	0.132	0.125	0.121	0.135	0.110	0.107	0.123	0.116	0.108
BTM	0.117	0.109	0.105	0.127	0.108	0.105	0.104	0.110	0.107
GD-VAE (T)	0.142	0.141	0.135	0.140	0.137	0.132	0.128	0.129	0.123
GD-VAE (TP)	0.142	0.138	0.136	0.143	0.137	0.134	0.129	0.127	0.125
GD-VAE (TR)	0.147	0.147	0.144	0.146	0.141	0.137	**0.136**	0.133	0.127
GD-VAE	**0.151**	**0.149**	**0.145**	**0.148**	**0.144**	**0.140**	**0.136**	**0.137**	**0.131**

Table: Results on topic coherence for different models.

Procedure Recommendation

Dataset	Method	Top-1 (%)			Top-3 (%)			Top-5 (%)			Top-10 (%)		
		R	P	F1	R	P	F1	R	P	F1	R	P	F1
Small	Word2Vec	19.5	47.8	24.7	35.4	34.9	30.8	47.1	29.6	32.0	62.3	21.1	28.5
	DWL	19.7	48.2	25.0	35.9	35.2	31.3	47.5	30.3	32.4	63.0	20.9	28.7
	BPR	23.5	57.6	29.8	44.8	43.5	38.7	56.8	35.7	38.8	73.1	24.8	33.6
	VAE-CF	24.0	57.8	30.7	46.0	43.5	39.3	57.8	35.2	39.1	74.0	24.2	33.8
	GD-VAE	**25.6**	**58.6**	**31.8**	**47.0**	**43.8**	**39.8**	**58.7**	**36.2**	**39.6**	**75.9**	**25.1**	**34.5**
Median	Word2Vec	7.8	27.6	10.9	27.7	30.5	25.1	38.3	26.9	27.7	52.8	20.1	26.1
	DWL	8.0	27.5	11.1	27.9	30.8	25.2	39.5	27.0	27.9	53.9	20.9	27.4
	BPR	10.2	35.8	14.9	38.6	40.2	34.3	49.3	33.3	34.9	65.2	23.8	31.4
	VAE-CF	21.2	52.9	26.2	41.2	42.0	36.0	53.4	35.3	37.3	68.2	24.9	32.9
	GD-VAE	**23.2**	**57.9**	**29.6**	**43.2**	**43.9**	**38.2**	**54.6**	**36.0**	**38.4**	**70.4**	**25.3**	**33.7**
Large	Word2Vec	5.3	22.9	8.7	14.6	21.1	15.3	24.8	21.0	20.1	41.1	17.7	22.2
	DWL	5.6	23.0	9.0	14.9	21.3	15.6	24.8	21.4	20.5	42.0	18.2	23.0
	BPR	7.3	26.7	10.2	23.0	27.1	21.2	38.4	27.6	27.9	56.6	21.7	28.0
	VAE-CF	17.8	50.1	23.5	35.2	37.9	33.4	47.9	32.4	34.6	63.0	21.7	30.2
	GD-VAE	**21.2**	**56.4**	**27.4**	40.9	**43.0**	**36.7**	**51.4**	**35.2**	**36.8**	**66.5**	**24.9**	**32.7**

Table: Comparison of various methods on procedure recommendation.

Admission-Type Prediction

Data	Small			Median			Large		
Method	P	R	F1	P	R	F1	P	R	F1
TF-IDF	84.26	87.19	85.18	86.12	88.61	87.22	88.45	89.10	87.76
Word2Vec	85.08	87.89	86.23	86.60	88.87	87.71	87.11	89.16	88.12
FastText	84.21	87.15	85.29	86.66	88.65	87.39	88.06	89.23	88.00
SWEM	85.56	88.10	86.77	87.01	89.28	88.12	87.55	89.88	88.67
LEAM	85.34	88.03	86.55	87.03	89.29	88.14	87.61	89.94	88.73
GD-VAE	**87.00**	**89.60**	**88.01**	**88.19**	**89.70**	**88.94**	**89.14**	**91.01**	**90.05**

Table: Results on admission-type prediction.