بسم الله الرحمن الرحيم



تحليل هوشمند تصاوير زيست پزشكي

نیم سال اول ۲۰-۳۰ مدرس: محمد حسین رهبان

دانشگاه صنعتی شریف دانشکدهی مهندسی کامییوتر

تمرين چهارم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین ها بدون کسر نمره تا سقف ۱۲ روز وجود دارد. محل بارگزاری جواب تمرین ها بعد از ۴ روز بسته خواهد شد و پس از گذشت این مدت، پاسخهای ارسال شده پذیرفته نخواهند شد.
- توجه داشته باشید که نوت بوکهای شما باید قابلیت بازاجرای ۱۰۰ درصد داشته باشند و در صورت نیاز به نصب یک کتابخانه یا دسترسی به یک فایل، مراحل نصب و دانلود (از یک محل عمومی) در نوت بوک وجود داشته باشد.
- همفکری در انجام تمرین مانعی ندارد، فقط توجه داشته باشید که پاسخ تمرین حتما باید توسط خود شخص نوشته شده باشد. همچنین در صورت همفکری در هر تمرین، در ابتدای جواب تمرین نام افرادی که با آنها همفکری کرده اید را حتما ذکر کنید.
- برای پاسخ به سوالات نظری در صورتی که از برگه خود عکس تهیه میکنید، حتما توجه داشته باشید که تصویر کاملا واضح و خوانا باشد. درصورتی که خوانایی کافی را نداشته باشد، تصحیح نخواهد شد.
- محل بارگذاری سوالات نظری و عملی در هر تمرین مجزا خواهد بود. به منظور بارگذاری بایستی تمارین تئوری در یک فایل زیپ با نام IABI_Theo_hw4_[First-Name]_[Last-Name]_[Student-Id].zip و تمارین عملی نیز در یک فایل مجزای زیپ با نام IABI_Prac_hw4_[First-Name]_[Last-Name] بارگذاری شوند.
- در صورت وجود هرگونه ابهام یا مشکل، در کوئرای درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.

بخش نظری (۵۰ نمره)

سوال اول: تابع Aggregation در ۱۰ نمره)

- (الف) توضیح دهید که چرا انتخاب توابع aggregation با خاصیت permutation invariance برای مسائل MIL ضروری است.
- (ب) در مسئله یادگیری Multiple Instance Learning(MIL)، سه تابع aggregation برای ترکیب ویژگیهای پچها در نظر گرفته شدهاند:

$$f_{\max}(x_1, x_2, \dots, x_n) = \max(x_1, x_2, \dots, x_n)$$
 (T

$$f_{\text{prod}}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n x_i$$
 (ب

$$f_{\text{concat}}(x_1, x_2, \dots, x_n) = [x_1, x_2, \dots, x_n]$$
 (Expression)

برای هر یک از این توابع توضیح دهید که آیا مناسب برای مسائل MIL هستند یا نه و چرا.

(الف) در مسائل Multiple، Instance Learning (MIL) یک نمونه معمولاً شامل مجموعهای از پچها یا نمونههای فرعی است. هدف این است که مدل از ویژگیهای این پچها برای پیش بینیهای کلی استفاده کند. در اینجا ترتیب پچها اهمیت ندارد، زیرا هیچگونه اطلاعات اضافی در مورد ترتیب پچها وجود ندارد. بنابراین، تغییر در ترتیب پچها نباید بر خروجی مدل تاثیر بگذارد.

اگر تابع aggregation انتخابی خاصیت permutation invariance نداشته باشد، تغییر در ترتیب پچها ممکن است باعث تغییرات غیر منطقی در پیش بینیهای مدل شود. این امر می تواند مدل را از دقت صحیح باز دارد و پیش بینیهایی که تنها به ترتیب پچها وابسته هستند، به دست دهد. به همین دلیل، انتخاب توابع aggregation با خاصیت permutation invariance برای مسائل MIL ضروری است.

(ب) در مسئله ،MIL سه تابع aggregation به صورت زیر تعریف شدهاند:

$$f_{\max}(x_1, x_2, \dots, x_n) = \max(x_1, x_2, \dots, x_n)$$
 (\(T\)

$$f_{\text{prod}}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n x_i$$
 (ب

$$f_{\text{concat}}(x_1, x_2, \dots, x_n) = [x_1, x_2, \dots, x_n]$$
 (7

تابع به این تابع تنها بزرگترین مقدار از مجموعه پچها را انتخاب میکند. در مسائل ،MIL این نوع aggregation میتواند مفید باشد زیرا اگر یکی از پچها ویژگی مهم و برجسته ای داشته باشد، این ویژگی به طور مؤثر در پیش بینی نهایی مدل تاثیر میگذارد. به طور مثال، اگر یکی از پچها در یک تصویر حاوی اطلاعات حیاتی باشد (مثلاً وجود یک شیء خاص)، این تابع میتواند به درستی آن اطلاعات را در پیش بینی نهایی وارد کند. همچنین این تابع خاصیت permutation invariance دارد.

نتیجه: تابع $f_{
m max}$ معمولاً برای MIL نتیجه:

قابع f_{prod} این تابع ضرب ویژگیهای پچها را محاسبه میکند. در حالی که این تابع از نظر ریاضی خاصیت permutation invariance دارد، معمولاً برای مسائل MIL مناسب نیست. دلیل این امر این است که اگر یکی از پچها ویژگی کوچکی (مثلاً نزدیک به صفر) داشته باشد، ضرب تمام ویژگیها ممکن است منجر به صفر یا مقدار بسیار کوچکی شود که پیش بینی مدل را به شدت تحت تأثیر قرار می دهد. همچنین، این نوع aggregation ممکن است برای داده هایی که ویژگی هایشان مقادیر پراکنده و متفاوت دارند، مناسب نباشد.

نتیجه: تابع f_{prod} معمولاً برای MIL مناسب نیست، مگر در موارد خاص که ویژگیها مقادیر بزرگ یا مساوی یک داشته باشند و از نظر مدل این نوع aggregation منطقی باشد.

تابع feoncat این تابع ویژگیهای پچها را به صورت یک بردار متصل میکند. از آنجا که تابع concat ترتیب پچها را در نظر میگیرد، خاصیت permutation invariance ندارد و با تغییر ترتیب پچها، خروجی تغییر میکند. به همین دلیل، استفاده از این تابع در مسائل MIL معمولاً توصیه نمی شود، زیرا تغییرات در ترتیب پچها می تواند پیش بینی های مدل را تحت تاثیر قرار دهد. این تابع بیشتر برای استفاده در مدل هایی که ترتیب ورودی ها اهمیت دارد (مثل شبکه های عصبی بازگشتی) مناسب است.

نتیجه: تابع $f_{\rm concat}$ برای MIL مناسب نیست، زیرا خاصیت permutation invariance ندارد و ترتیب پچها را در نظر می گیرد.

سوال دوم: طبقه بندی رابطه ای در گراف (۶ نمره)

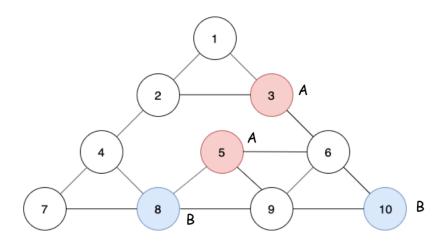
میخواهیم از طبقهبندی رابطهای برای پیش بینی برچسبهای گرهها استفاده کنیم. گراف G را در نظر بگیرید که در شکل ۱ نشان داده شده است. میخواهیم گرهها را به دو کلاس A و B طبقه بندی کنیم. برچسبهای گرههای A، A و A د داده شده اند.

استفاده از یک طبقه بند رابطه ای احتمالاتی برای پیش بینی Y_i برای گره i به صورت زیر تعریف می شود:

$$P(Y_i = c) = \frac{1}{|N_i|} \sum_{(i,j) \in E} W(i,j) P(Y_j = c)$$

که در آن $|N_i|$ تعداد همسایههای گره i است.

- فرض كنيد همهى يالها وزن 1 دارند.
- برای گرههای دارای برچسب، با مقدار واقعی Y مقداردهی اولیه کنید و برای گرههای بدون برچسب، از مقداردهی اولیه بی طرفانه یعنی $P(Y_i = A) = 0.5$ استفاده کنید.
 - گرهها را به ترتیب شماره گره بهروزرسانی کنید.



شكل 1: گراف G

- (آ) پس از مرحله(تکرار) دوم، مقدار $P(Y_i=B)$ را برای i=2,4,6 بدست آورید.
- $P(Y_i = A) > 0.5$ به عنوان آستانه احتمال استفاده کنیم، یعنی گره i به کلاس i تعلق دارد اگر i به کلاس i تعلق خواهد داشت؟ و کدام به کلاس i تعلق خواهد داشت؟

پاسخ:

- (آ) مقادیر بهدستآمده: (به ترتیب ۲، ۴، ۶) 0.24، 0.67، 0.46
 - به صورت زیر خواهند بود: $P(Y_i=A)$ به صورت زیر خواهند بود:

(1:0.86), (2:0.71), (3:1.00), (4:0.29), (5:1.00), (6:0.52), (7:0.14), (8:0.00), (9:0.38), (10:0.00)

بنابراین گرههای با شمارههای [۱، ۲، ۳، ۵، ۶] به کلاس A تعلق دارند و گرههای با شمارههای $\{4, \, V, \, A\}$ بنابراین $\{4, \, V, \, A\}$ به کلاس $\{4, \, V, \, A\}$ تعلق دارند.

سوال سوم: تابع Aggregator در GNN (۱۲ نمره)

فرض کنید یک گراف G داریم. ماتریس A ماتریس مجاورت آن است که نشان می دهد کدام گرهها به هم متصل اند و همچنین ماتریس قطری D که به ابعاد $|V| \times |V|$ است، ماتریس درجه گرههای گراف است. بطوریکه عناصر قطر آن، درجه هر گره را نشان می دهد. فرض کنید $h_i^{(l)}$ نمایش دهنده Embedding گره i در لایه i باشد. همچنین از یک تجمیع کننده میانگین گیر (Mean Aggregator) برای آپدیت گرهها استفاده می کنیم به طوری که:

$$h_i^{(l+1)} = \frac{1}{|N_i|} \sum_{j \in N_i} h_j^{(l)}$$

اگر مقدار Embedding همه گرهها در لایه L را به شکل زیر Vectorize کنیم:

$$H^{(L)} = \begin{bmatrix} h_1^{(L)} \\ h_2^{(L)} \\ \vdots \\ h_{|V|}^{(L)} \end{bmatrix}$$

- (آ) یک رابطه ماتریسی پیدا کنید که با داشتن ماتریسهای A ، D و بردار $H^{(L+1)}$ ، بردار $H^{(L+1)}$ که هر گره با استفاده از تابع تجمیع کننده میانگین گیر آپدیت شده را محاسبه کند.
- (ب) حال فرض کنید در گراف G میخواهیم بصورت تصادفی با احتمال یکنواخت (Uniform) حرکت کنیم. یعنی اگر گره i به گرههای j و k متصل باشد، به احتمال مساوی از گره i به گره i یا k میتوانیم برویم. ماتریس انتقال (Transition) را برای این حرکت تصادفی با داشتن ماتریسهای i و i بدست بیاورید. با توجه به پاسخ بخش «آ» و این بخش چه نتیجهای میتوان گرفت؟ (راهنمایی: ماتریس Transition ماتریس i ماتریس i است که درایه i در آن احتمال رفتن از گره i به گره i با ناز این میدود.
- (ج) حال فرض کنید یک Skip Connection به تابع Aggregator اضافه کنیم که مقدار Embedding قبلی خودش نیز در مقدار Embedding لایه بعدش تاثیرگذار باشد، به طوریکه:

$$h_i^{(l+1)} = \frac{1}{3}h_i^{(l)} + \frac{2}{3}\frac{1}{|N_i|} \sum_{i \in N:} h_j^{(l)}$$

همانند بخش «آ» رابطه ماتریسی مربوط به این تابع تجمیع کننده را بنویسید.

(د) همانطور که تا اینجا با تابع تجمیعکننده Mean آشنا شده اید، توابع تجمیعکننده دیگری برای تجمیع Embedding همسایههای یک گره در گراف وجود دارند. دو تابع دیگر عبارت است از:

$$\begin{split} h_v^{l+1} &= \mathrm{AGGREGATE_{min}}\left(\left\{h_u^{(l)}, \, \forall u \in \mathcal{N}(v)\right\}\right) = \min_{u \in \mathcal{N}(v)} h_u^{(l)} \\ h_v^{l+1} &= \mathrm{AGGREGATE_{sum}}\left(\left\{h_u^{(l)}, \, \forall u \in \mathcal{N}(v)\right\}\right) = \sum_{u \in \mathcal{N}(v)} h_u^{(l)} \end{split}$$

حال یک گراف G(V,E) بیابید که دارای دو شرط زیر باشد:

- $|V| > 3 \bullet$
- $\forall v \in V, \deg(v) \geq 2 \bullet$

و همچنین مقدار اولیهای به عنوان Embedding (همان h_i^0) داشته باشد که پس از اعمال تابع تجمیع کننده Min در لایه اول و اعمال تابع تجمیع کننده Sum در لایه دوم، مقدار Embedding گرهها در لایه دوم با مقادیر اولیه یکی شوند(یعنی $\forall v \in V, h_v^0 = h_v^2$).

پاسخ:

$$H^{(L+1)} = D^{-1}AH^{(L)}$$
 (1)

(ب) مقدار ورودی (i,j) در ماتریس انتقال به صورت زیر است:

- j و i برای دو گره متصل و $\frac{1}{|D_{i.i}|}$
 - 0 در غیر این صورت،

که به صورت $D^{-1}A$ نوشته می شود.

i نتیجه میگیریم مقدار Embedding گره i در لایه بعدی GNN برابر است با امید ریاضی Embedding گره i در یک مرحله حرکت تصادفی یکنواخت با شروع از گره i در گراف G.

 (π,j) مقدار ورودی (i,j) در ماتریس انتقال به صورت زیر است:

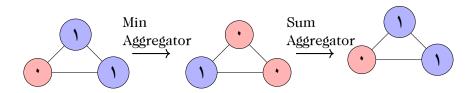
- i=j هنگامی که $rac{1}{3}$
- j برای دو گره متصل و $\frac{2}{3\cdot |D_{i,i}|}$
 - 0 در غیر این صورت،

که به صورت $I=\frac{2}{3}D^{-1}A+\frac{1}{3}$ نوشته میشود. با توجه به نتیجه بخش ب میتوان رابطه ماتریسی را بصورت زیر نوشت:

$$H^{(L+1)} = (\frac{2}{3}D^{-1}A + \frac{1}{3}I)H^{(L)}$$

(د)

Initial Embeddings Layer 1 Embeddings Layer 2 Embeddings



سوال چهارم: چالشهای GNNها (۸ نمره)

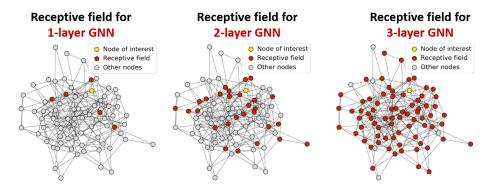
میدانیم یک لایه از شبکه GNN استاندارد از تابع زیر برای Aggregation و اعمال Activation Function استفاده می کند تا مقادیر گرهها را بروزرسانی کند.

$$\mathbf{h}_{v}^{(l)} = \sigma \left(\sum_{u \in \mathcal{N}(v)} \mathbf{W}^{(l)} \frac{\mathbf{h}_{u}^{(l-1)}}{|\mathcal{N}(v)|} \right)$$

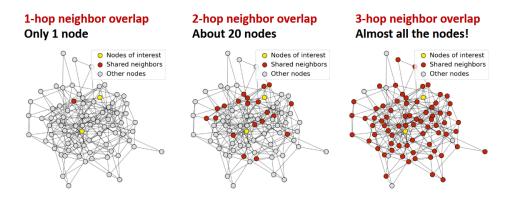
- (آ) یکی از چالشهای اصلی GNN مشکل Over-smoothing است. این مشکل را توضیح دهید. دلیل بوجود آمدن این مشکل چیست؟ بدلیل چه خاصیتی شبکههای گرافی این چالش را دارند؟ استدلال خود را برای دو گره که در فاصلهی دوری از هم در گراف قرار دارند، با مثالی ترسیمی بیان کنید.
- (ب) روشی ارائه کنید که اگر نیاز به تعداد لایههای زیادی از GNN داریم، بتواند از بروز این مشکل جلوگیری کند. تابع آپدیت متناظر با روشی که بیان کردید را بنویسید.

پاسخ:

(آ) این مشکل بیان میکند که در شبکههای گرافی، با افزایش تعداد لایههای GNN مقدار Embedding گرهها به یک مقدار یکسان همگرا میشود. دلیل بوجود آمدن این موضوع هم افزایش تعداد لایههای GNN است.



شکل ۲: خاصیت اصلی شبکه گرافی که در مشکل Over-smoothing دخیل است receptive field بالای گرههای این شبکه در تعداد لایههای بالاست.



شکل ۳: این مثال مشکل را ترسیم میکند که دو گره با فاصله نسبتا زیاد چرا به مقدار یکسانی بعد از گذر از تعداد لایههای زیاد میرسند. افزایش همسایههای shared آن دو گره در لایههای آخر دلیل این موضوع است.

(ب) استفاده از ایده Skip connection و داشتن بلاکهای residual که از مقدار Embedding لایه قبلیِ خودِ گره نیز استفاده میکند، کمک به جلوگیری از این مشکل خواهد کرد. تابع آپدیت به شکل زیر تغییر خواهد کرد:

$$\mathbf{h}_{v}^{(l)} = \sigma \left(\sum_{u \in \mathcal{N}(v)} \mathbf{W}^{(l)} \frac{\mathbf{h}_{u}^{(l-1)}}{|\mathcal{N}(v)|} + \mathbf{h}_{v}^{(l-1)} \right)$$

سوال پنجم: Image Registration (۱۰ نمره)

با توجه به مطالبی که در کلاس آموختید، برای اینکه بتوانیم یک moving image را با fixed image انطباق دهیم و سپس مدلی داشته باشیم که بتوانیم فرآیند **image registration** را انجام دهیم:

(آ) یکی از روشها برای این فرآیند استفاده از VoxelMorph است. ابتدا توضیح دهید ساختار VoxelMorph

به چه شکل است و با توجه به اینکه این روش Unsupervised است، چه مزیتی نسبت به روش Supervised که می توان برای آن استفاده کرد دارد؟

(ب) همان طور که می دانید، عبارت loss در VoxelMorph به این صورت تعریف می شود:

$$\mathcal{L} = \|m \circ \phi - f\| + \lambda \operatorname{Reg}(\phi)$$

که در آن بخش اول مربوط به تطبیق تصاویر (images match) و بخش دوم تضمین کننده یک میدان هموار (smooth field) است.

مهمترین مشکلی که در این loss وجود دارد چیست؟ و با چه روشی میتوان این مشکل را حل کرد؟ (راهنمایی: برای به دست آوردن اطلاعات بیشتر، به مقاله VoxelMorph مراجعه کنید.)

پاسخ:

(آ) در **VoxelMorph** دو تصویر fixed image و moving image به عنوان ورودی دریافت می شوند و سپس از طریق یک شبکه عصبی **CNN** به deformation می رسیم. این فرآیند به صورت unsupervised انجام می شود.

یک روش دیگر نیز وجود دارد که در آن از یک فرد خواسته می شود تمامی train dataset را بررسی کرده و deformation را برای هر pair به دست آورد. در این حالت صرفاً به یک شبکه عصبی CNN نیاز داریم تا deformation را یاد بگیرد.

همانطور که مشخص است، روش supervised به علت زمان زیادی که از منابع انسانی میگیرد و همچنین به دلیل احتمال وجود خطا توسط فردی که deformation را به دست میآورد، توصیه نمی شود.

(ب) اصلی ترین مشکلی که این loss دارد، مشتق ناپذیر بودن آن است، زیرا هنگامی که میخواهیم -backpropa (ب) اصلی قطنان و gation انجام دهیم، ماتریس متناظر آن از یک سری indicator function تشکیل می شود که مشتق ناپذیر است.

برای حل این مشکل، یک متد دیگر معرفی شده است که از آن میتوان مشتق گرفت. که به صورت زیر می باشد:

$$m \circ \phi(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{Z}(\mathbf{p}')} m(\mathbf{q}) \prod_{d \in \{x, y, z\}} (1 - |p'_d - q_d|)$$

که در آن:

- است. voxel p' مجموعه همسایگیهای $\mathcal{Z}(\mathbf{p}')$
 - به ابعاد z, y, x اشاره دارد. d

به این ترتیب، چون این روش از linear interpolation استفاده میکند، امکان محاسبه گرادیانها و در نتیجه انجام back propagation طی بهینهسازی وجود دارد.

سوال ششم (۱۰ نمره)

همانطور که از قبل میدانید یکی از روشهای registration در تصاویر پزشکی VoxelMorph است. با توجه به سناریوی زیر به سوالات زیر پاسخ دهید: - تعداد کمی داده ی T۱-weighted داریم اما تصاویری که قرار است با آنها کار کنیم ممکن است از contrastهای مختلفی باشد (مثلاً تصاویر T۲).

(الف) به نظر شما استفاده از VoxelMorph برای چنین سناریویی مناسب است؟ چرا؟

(ب) یکی دیگر از روشهای register کردن تصاویر پزشکی **SynthMorph** است. برای سناریوی ذکرشده توضیح دهید استفاده از این روش مناسب است یا خیر؟ علت خود را نیز ذکر کنید. (برای آشنایی به روش - Synth Morph می توانید به این مقاله مراجعه کنید)

پاسخ:

- (الف) استفاده از VoxelMorph برای چنین سناریویی مناسب نیست. این روش بر پایه VoxelMorph برای چنین سناریویی مناسب نیست. این روش بر پایه voxelMorph است که نیاز به دادههای contrast_یکسان برای آموزش دارد. در این روش فرض بر این است که دادههای train و دادههایی که قرار است ثبت شوند از نظر ویژگیهای شدت و contrast به یکدیگر شباهت دارند. در این سناریو که تصاویر contrastهای مختلفی دارند (مثلاً ۲۲ در مقابل ۲۱)، مدل نمی تواند به خوبی تعمیم یابد زیرا تفاوتهای ساختاری ناشی از contrast باعث اختلال در عملکرد ثبت می شود. همچنین، تعداد کم دادههای T۱-weighted بنیز محدودیت دیگر این روش است که وقتی دیتای کمتری داریم عموما نتیجهی مظلوبی را VoxelMorph به ما نمی دهد
- (ب) SynthMorph برای این سناریو گزینه بسیار مناسبی است. این روش به جای تکیه بر دادههای واقعی، از دادههای مصنوعی که از نظر contrast، شدت، و نویز متنوع هستند برای آموزش استفاده میکند. در نتیجه، مدل نهایی توانایی تعمیمدهی به دادههای واقعی با contrastهای مختلف (مانند T۲) را خواهد داشت. به طور خاص، SynthMorph از یک فرآیند دادهافزایی استفاده میکند که طی آن تصاویر مصنوعی با sontrast و تغییرات متنوع تولید میشوند. این امر به مدل اجازه میدهد تا به تفاوتهای ناشی از contrast حساس نباشد و عملکرد قابل اعتمادی ارائه دهد. همچنین، این روش به دلیل عدم نیاز به دادههای واقعی زیاد، محدودیت تعداد کم دادههای اوائه دهد. همچنین، در دادههای ورودی وجود دارد (مانند این سناریو) عملکرد بسیار می وانند در محیطهایی که تغییرات زیادی در دادههای ورودی وجود دارد (مانند این سناریو) عملکرد بسیار خوبی داشته باشد.