

تحلیل یک دهه از نمرات دانشجویان دانشکده مهندسی برق دانشگاه صنعتی شریف، با استفاده از پردازش سیگنال‌های گرافی

امیرحسین گل شیرازی^۱، رضا پرهیزکار^۲، آرش امینی^۳ و محمد مهدی امتی^۴

تاریخ دریافت: ۱۴۰۳/۰۵/۰۴، تاریخ پذیرش: ۱۴۰۳/۰۹/۲۴

DOI: 10.22047/ijee.2024.469633.2098

DOR: 20.1001.1.16072316.1404.27.105.3.3

چکیده: در پژوهش حاضر، به بررسی عملکرد تحصیلی دانشجویان و شناسایی الگوهای مؤثر بر موفقیت یا افت تحصیلی آنان، با استفاده از ابزارهای پردازش سیگنال گرافی، می‌پردازیم. جامعه آماری بررسی شده شامل دانشجویان کارشناسی دانشکده مهندسی برق دانشگاه صنعتی شریف، در بازه زمانی ۱۴۰۰-۱۳۹۰، و داده‌های پژوهشی شامل نمرات دانشجویان در دروس گوناگون، شاخه‌های تخصصی (گرایش) و زمان پذیرش هر درس است. شایان ذکر است که از تمامی داده‌ها استفاده و از نمونه‌گیری پرهیز شده است. در روش استفاده شده، هر دانشجو را یک گره در یک گراف در نظر گرفته و گره‌ها را، براساس مشابهت عملکرد تحصیلی، به صورت وزن دار متصل کرده‌ایم. سپس، با استفاده از اتصالات گراف، تطابق نمرات دروس خاص را با عملکرد کلی دانشجویان بررسی کرده‌ایم. نتایج نشان دهنده تطابق نداشتن نمرات برخی از دروس با عملکرد کلی دانشجویان است که ممکن است به تنوع سیاست‌های نمره‌دهی و سبک‌های آموزشی مرتبط باشد. این بازخورد به بهبود شیوه نمره‌دهی کمک می‌کند. همچنین ارزیابی‌ها نشان می‌دهد انتخاب گرایش دانشجویان تنها در ۴۴ درصد از موارد با توانایی آنان همخوانی دارد. در تحلیلی دیگر، مشخص شد نمرات ۵ درصد از دانشجویان در دوران همه‌گیری کرونا رشد چشمگیری داشته که ممکن است به تقلب نظام‌مند در امتحانات مجازی اشاره داشته باشد.

واژگان کلیدی: آشکارسازی تقلب در آزمون، افت تحصیلی، آموزش عالی، پردازش سیگنال‌های گرافی، رشد تحصیلی

- ۱- دانشجوی کارشناسی ارشد، گروه سیستم‌ها و شبکه‌های مخابراتی، دانشکده مهندسی برق، دانشگاه صنعتی شریف، تهران، ایران. amirgol3000@yahoo.com و amir.golshirazi@ee.sharif.edu
- ۲- متخصص ارشد علم داده، زوریخ، سوئیس. reza.parhizkar@gmail.com
- ۳- دانشیار گروه سیستم‌ها و شبکه‌های مخابراتی دانشکده مهندسی برق، دانشگاه صنعتی شریف، تهران، ایران (نویسنده مسئول). aamini@sharif.edu
- ۴- فارغ‌التحصیل کارشناسی ارشد گروه سیستم‌ها و شبکه‌های مخابراتی دانشکده مهندسی برق، دانشگاه صنعتی شریف، تهران، ایران. mohammad.omati@yahoo.com

۱. مقدمه

رقابت بین دانشگاه‌ها، تغییرات ملی و جهانی، و تحولات سیاسی و اجتماعی در آموزش عالی روزافزون است. موفقیت در این وضعیت، نیازمند پویایی، برآورده ساختن نیازهای دانشجویان، به عنوان مشارکت‌کننده، و بهبود کارایی مالی و عملیاتی خود در برابر آنان است (van Barneveld et al., 2012). به همین دلیل، در سال‌های اخیر، بسیاری از دانشگاه‌ها به روش‌های تحلیل یادگیری، به منظور دستیابی به یافته‌های مربوط به پیشرفت تحصیلی دانشجویان، پیش‌بینی رفتارهای آینده و شناسایی مشکلات احتمالی در مراحل اولیه، روی آورده‌اند. تحلیل یادگیری در آموزش عالی رفتار یادگیری دانشجویان را بازتاب می‌دهد و کمکی خوب از سوی استادان یا مشاوران است. این پشتیبانی فردی/گروهی روش‌های جدید آموزش را نشان می‌دهد و امکان بازتاب رفتار یادگیری دانشجویان را فراهم می‌کند. همچنین تحلیل یادگیری همکاری‌های بین‌مؤسسه‌ای و توسعه دستورکار جامعه بزرگ دانشجویان و استادان را تسهیل می‌کند (Atif et al., 2013).

در سال‌های اخیر، با استفاده از اطلاعات تحصیلی دانشجویان و به منظور تحلیل و تخمین عملکرد آنان، تحقیقات بسیاری شده است. میگوئیس و همکاران در ۲۰۱۸ (Miguéis et al., 2018) داده‌های آموزشی ۲۴۵۹ دانشجو را جمع‌آوری کردند تا مدل‌های پیش‌بینی عملکرد تحصیلی دانشجویان را، با استفاده از الگوریتم‌های یادگیری ماشین^۱، مانند ماشین‌های بُردار پشتیبان^۲ (Guido et al., 2024; Hearst et al., 1998)، درخت تصمیم‌گیری^۳ (Song & Lu, 2015; Hastie et al., 2009)، جنگل تصادفی^۴ (Sarker, 2021; Breiman, 2001) و دسته‌بندی‌کننده بیز ساده^۵ (Rish, 2001) طراحی کنند. مدل پیشنهادی جنگل تصادفی با دقت بیش از ۹۵ درصد نشان داد نمرات ترم‌های قبلی نقشی مهم در ساخت مدل دارند.

در ۲۰۲۱، مینگیو و همکاران (Mingyu et al., 2021) از گروه‌بندی ویژگی‌های مربوط به تحصیل، زندگی، فعالیت‌های اینترنتی و اطلاعات پایه دانشجویان برای حل مسئله پیش‌بینی عملکرد تحصیلی استفاده کردند. مدل‌های پیش‌بینی از الگوریتم‌های یادگیری ماشین، مانند رگرسیون لجستیک^۶ (Sperandei, 2014)، درخت تصمیم‌گیری، ماشین بُردار پشتیبان، جنگل تصادفی، درخت تصمیم‌گیری تقویت‌شده گرادیان^۷ (Friedman, Chen & Guestrin, 2016)، XGBoost (Ke et al., 2017 و LightGBM (Lundberg & Lee, 2017; Catboost-SHAP Prokhorenkova et al., 2018)، با بهترین درصد دقت برحسب متریک‌های ضریب تعیین^۸، خطای میانگین مربعات^۹ و میانگین قدرمطلق خطا^{۱۰}، بود.

1- Machine learning

4- Random forest

7- Gradient boosting decision tree

10- Mean absolute error (MAE)

2- Support vector machine (SVM)

5- Naive Bayes classifier

8- R-Squared (R2)

3- Decision tree

6- Logistic regression

9- Mean squared error (MSE)

در ۲۰۲۲، حامیم و همکاران (Hamim et al., 2022) نتایج یادگیری دانشجویان را براساس جنبه‌هایی، مانند رفتار شخصی، رفتار اجتماعی، عادات تحصیلی، خانواده‌ها، رشته‌های تحصیلی و نمرات، طبقه‌بندی کردند. این پژوهش بر آزمایش و مقایسه الگوریتم‌های تقویتی، مانند تقویت سازگار، تقویت گرادیان^۲، تقویت گرادیان شدید^۳، CatBoost و LightGBM، برای یافتن مؤثرترین الگوریتم، تمرکز داشت. آنان در نتایج خود نشان دادند که LightGBM، با بهره‌گیری از ۱۴ ویژگی^۴ انتخاب شده مبتنی بر روش انتخاب ویژگی^۵، بیشترین دقت را، در مقایسه با سایر روش‌ها، از خود نشان می‌دهد. این ویژگی‌ها عمدتاً مربوط به رفتار یادگیری و نمرات دانشجویان است.

علاوه بر پژوهش‌های یادشده، شاخه‌ای وسیع به نام «تحلیل یادگیری»، در دانشگاه‌های ایالات متحده آمریکا، بر زمینه‌هایی تمرکز دارد که به استادان و تصمیم‌گیران دانشگاه‌ها امکان می‌دهند عملکرد کلی دانشجویان را بررسی و از آن در بازبینی دروس دانشگاهی یا بهبود عملکرد کلی دانشجویان استفاده کنند (Dietz-Uhler & Hurn, 2013; Campbell et al., 2007; Dyckhoff et al., 2012). برای مثال، استفاده از داده‌هایی که بیشتر سامانه‌های مدیریت یادگیری^۶ به صورت خودکار جمع‌آوری می‌کنند به استادان اجازه می‌دهد نحوه پیشرفت دانشجویان را در یک دوره تحصیلی بررسی کنند. اسمیت و همکاران (Smith et al., 2012) نشان دادند که تعداد دفعات ورود دانشجویان به سامانه‌های مدیریت یادگیری، میزان تعامل آنان با مباحث آموزشی، سرعت پیشرفت و نمرات تکالیفشان عملکرد آنان را در دوره تحصیلی به طور موفقیت‌آمیزی پیش‌بینی می‌کند. همان‌گونه که وبگاه‌های خرید برخط، مثل آمازون و دیجی‌کالا، از داده‌های تاریخچه خرید ما برای پیشنهاد خریدهای آینده استفاده می‌کنند، تحلیل یادگیری نیز به ما کمک می‌کند فرصت‌های یادگیری جدید یا اقدامات مربوط را به دانشجویان خود پیشنهاد دهیم (Campbell et al., 2007).

هدف اصلی ما در مقاله حاضر ارائه روش‌های مبتنی بر پردازش سیگنال گرافی (Sandryhaila & Shuman et al., 2013; Moura, 2013, 2014) به منظور بررسی سه رویکرد ذیل است:

۱. مقایسه عملکرد دانشجویان، به صورت عمومی و در دروس خاص؛ بررسی تطابق توزیع نمرات دانشجویان در دروس با توزیع سطح علمی آنان، همراه با شناسایی دروسی که بیشترین و کمترین تطابق‌پذیری را با عملکرد عمومی دانشجویان دارند.
۲. تحلیل صحت انتخاب شاخه (گرایش) دانشجویان، با توجه به عملکردشان؛ بررسی صحت انتخاب شاخه دانشجویان و تحلیل اینکه آیا شاخه مناسب‌تری برای برخی از دانشجویان وجود دارد یا خیر. بدین منظور، ابزاری برای پیش‌بینی معدل نمرات تخصصی هر دانشجو در شاخه‌های انتخاب نشده معرفی شده است که موفقیت دانشجو را در شاخه‌های گوناگون ارزیابی می‌کند.

1- Adaptive boosting

2- Gradient boosting

3- Extreme gradient boosting

4- Feature

5- Feature selection technique

6- Learning management systems (LMS)

۳. بررسی عملکرد تحصیلی دانشجویان طی زمان و تحلیل تأثیر همه‌گیری کرونا بر عملکرد افراد: ارزیابی روند رشد/افت تحصیلی دانشجویان در طول زمان و بررسی تأثیر همه‌گیری کرونا بر این روند. برای هر دانشجو در هر ترم شاخصی تعریف می‌کنیم که نشان‌دهنده عملکرد تحصیلی وی در آن ترم است و در طول تحصیل هر دانشجو رصد می‌شود. نتایج نشان می‌دهد ورودی‌هایی که نیمه ابتدایی تحصیلی آنان قبل از همه‌گیری کرونا و نیمه دوم تحصیلی‌شان در دوران همه‌گیری بوده است تفاوت‌های درخور توجهی با سایر ورودی‌ها دارند.

در پژوهش حاضر، به‌منظور کارآمدی روش‌های مطرح‌شده، از مجموعه داده‌ای جامع، متعلق به دانشکده مهندسی برق دانشگاه صنعتی شریف، استفاده کرده‌ایم که طی ده سال تحصیلی، از ۱۳۹۰-۱۳۹۱ تا ۱۳۹۹-۱۴۰۰، به‌صورت سری زمانی، گردآوری شده است. این داده‌ها شامل شماره ترم، شماره دانشجویی کدگذاری شده (منحصر به هر دانشجو)، دروس اخذشده و نمرات دانشجویان در هر نیمسال تحصیلی است. بخش ۲ مقاله به معرفی گراف، نحوه تعریف رئوس و یال‌ها در آن، سیگنال گرافی در حوزه گره و حوزه فوریه و در نهایت مفهوم یادگیری گراف می‌پردازد. در بخش ۳، داده‌های جمع‌آوری شده و روش پیشنهادی مبتنی بر پردازش گرافی معرفی می‌شود. در بخش پایانی، ۴، نتایج مبتنی بر سه رویکرد اشاره‌شده تحلیل می‌شود.

۲. مقدمات پردازش سیگنال‌های گرافی

در بیشتر مسائل مربوط به پردازش سیگنال، با سیگنال‌هایی مواجهیم که متغیر مستقل آنها از جنس زمان/مکان است. برای مثال، سیگنال اندازه‌گیری شده حسگر دما در اتاق چگونگی تغییر دما را در طول زمان نشان می‌دهد. چنین متغیرهایی دارای نوعی توالی و ترتیب ذاتی هستند، یعنی ابتدا لحظه ۰.۹ ثانیه، سپس لحظه ۱ ثانیه و پس از آن لحظه ۱.۱ ثانیه محقق شده و هر لحظه میانی (مثل ثانیه ۱ در مثال یادشده) رابط لحظه‌های قبل و بعد خودش است. به‌علاوه، همبستگی هر لحظه به لحظه‌های مجاورش بیش از زمان‌های دیگر است. برای مثال، در سیگنال تصویری، متغیر مستقل مکان دو بُعدی است و دارای توالی معنادار.

فرض کنید می‌خواهیم دمای هوای امروز را در ایستگاه‌های هواشناسی تهران بزرگ به‌صورت سیگنال نمایش دهیم به‌طوری‌که متغیر مستقل معرف ایستگاه‌های هواشناسی و متغیر وابسته معرف دمای اندازه‌گیری شده در هر ایستگاه باشد. چنانچه بخواهیم این داده را به‌عنوان سیگنال مشابه سیگنال زمانی (یا مکانی) پردازش کنیم باید نوعی توالی معنادار میان ایستگاه‌های هواشناسی برقرار باشد به‌صورتی‌که به‌همان ترتیب ایستگاه‌ها را روی محوری افقی قرار و دمای اندازه‌گیری شده هر ایستگاه را روی محوری عمودی نمایش دهیم. انتخاب توالی معنادار از مهم‌ترین مراحل تحلیل و پردازش سیگنال به‌شمار می‌آید. فرض کنید ایستگاه‌های اندازه‌گیری دما را به‌ترتیب الفبای نام

محل قرارگیری آنها مرتب کنیم. چنین توالی‌ای معیار مناسبی برای نشان دادن همبستگی و مشابهت ایستگاه‌های مجاور روی محور متغیر مستقل نیست چراکه ممکن است دمای اندازه‌گیری شده در دو ایستگاه بسیار دور که صرفاً به دلیل نزدیکی نام خود در مجاورت هم و روی محور متغیر مستقل قرار گرفته‌اند بسیار متفاوت باشد. در مورد مطالعه شده، می‌توان بررسی کرد که تعیین متغیر مستقل تک‌بعدی، چنان‌که هر ایستگاه تنها دو همسایه داشته باشد، انتخاب بهینه‌ای نیست.

تعدد داده‌های مشابه مثال بالا و اهمیت تحلیل چنین سیگنال‌هایی منجر به ابداع پردازش سیگنال گرافی^۱ در دنیای امروز شده است (Ortega, 2022). در این شاخه از پردازش سیگنال، هر متغیر مستقل به مثابه یک گره از گراف انتخاب می‌شود و یال‌های وزن دار/بی‌وزن گره‌ها را به گره‌های دیگر متصل می‌کنند. وزن یال بین دو گره نشان‌دهنده میزان همبستگی دو گره است. حال، می‌توان سیگنال موردنظر را روی این گراف نشان داد و متناظر با هر گره مقداری عددی را (برای مثال، دمای هوای امروز تهران در گره متناظر با آن ایستگاه) تعریف کرد.

در ادامه فصل پیش رو، به معرفی کوتاه مفاهیم پایه‌ای گراف، سیگنال گرافی و پردازش سیگنال گرافی می‌پردازیم و تئوری گراف و نشانه‌گذاری‌های^۲ این حوزه را مرور می‌کنیم.

۲-۱. اجزای گراف

هر گراف متشکل از تعدادی گره و یال است. در بیشتر مسائل کاربردی، یال‌ها وزن دار هستند و مقدار وزن یال بین دو گره معرف میزان همبستگی آنها با یکدیگر است. گراف‌ها شامل انواعی، از جمله جهت دار و طوقه دار، هستند اما، به دلیل اهمیت گراف‌های ساده و وزن دار در تحلیل نمرات دانشجویان، در مقاله حاضر درخصوص این نوع از گراف‌ها بحث می‌کنیم.

۲-۱-۱. ماتریس وزن، درجه و لاپلاسیان گراف

فرض کنید گراف وزن دار G دارای n گره است؛ در این صورت: $w_{n \times n}$ ماتریس وزن گراف G در نظر گرفته می‌شود به طوری که $w_{(i,j)}$ برابر با وزن بین دو رأس با شماره i و j است؛ $D_{n \times n}$ ماتریس درجه گراف یک ماتریس قطری است که درایه i نام قطر اصلی آن برابر با مجموع وزن یال‌های متصل به گره i نام است.

$$D_{i,i} = \sum_j w_{i,j} \quad (1)$$

$L_{n \times n}$ نیز ماتریس لاپلاسیان گراف نامیده و به صورت زیر تعریف می‌شود:

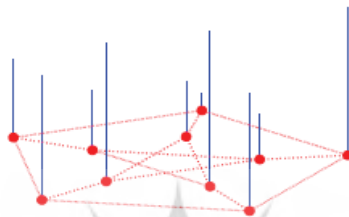
$$L = D - W \quad (2)$$

با توجه به تعاریف مذکور، جمع اعداد هر سطر ماتریس لاپلاسیان برابر صفر است و، به عبارتی، بردار

$[1, 1, \dots, 1]^T$ بردار ویژه متناظر با مقدار ویژه صفر برای ماتریس لاپلاسیان. داشتن مقدار ویژه صفر نتیجه می‌دهد که ماتریس لاپلاسیان رتبه کامل^۱ نیست و همواره وارون ناپذیر است.

۲-۲. سیگنال‌های گرافی

با اختصاص دادن یک مقدار عددی به هر گره گراف وزن دار G می‌توان یک سیگنال گرافی (شکل ۱) را مشابه سیگنال‌های کلاسیک در حوزه زمان/مکان تعریف کرد.



شکل ۱. سیگنال گرافی

به منظور نمایش ساده سیگنال‌های گرافی، ابتدا گره‌های گراف را از 1 تا n شماره‌گذاری می‌کنیم و سپس سیگنال گرافی را با بردار $n \times 1$ \mathbf{x} نمایش می‌دهیم به طوری که مؤلفه x_i نشان دهنده مقدار سیگنال گرافی در گره i ام باشد.

سیگنال‌های بُرداری (در مقابل سیگنال‌های عددی) سیگنال‌هایی هستند که در آنها به هر گره، به جای یک مقدار عددی، یک بردار نسبت داده می‌شود و هر مشخصه بُرداری مربوط به گره i مقدار عددی یک ویژگی از این گره را نشان می‌دهد. در چنین سیگنال‌هایی، بردارهای نظیر گره‌های متفاوت هم بُعد هستند.

۲-۳. سیگنال گرافی هموار

در گراف G ، با ماتریس لاپلاسیان L و سیگنال بُرداری $n \times 1$ \mathbf{x} ، انرژی سیگنال به صورت

$$E = \mathbf{x}^T \cdot L \cdot \mathbf{x} \quad (3)$$

تعریف می‌شود. این مقدار برابر است با

$$E = \frac{1}{2} \sum_{i,j} W_{ij} |x_i - x_j|^2 \quad (4)$$

درخصوص سیگنال‌های بُرداری، این رابطه به صورت زیر نمایش داده می‌شود:

$$E = \frac{1}{2} \sum_{i,j} W_{ij} Z_{ij} \quad (5)$$

که Z_{ij} فاصله بُرداری دو بُردار x_i و x_j است. در ادامه، خواهیم دید که فاصله تعریف‌های دیگری نیز دارد.

هرچه انرژی گراف G کمتر باشد سیگنال گرافی هموارتر است؛ به عبارتی دیگر، گره‌های همبسته‌ای که بین خود یال‌هایی با وزن زیاد دارند مقادیر بُرداری تقریباً یکسانی دارند.

شایان ذکر است که اگر وزن تمامی یال‌ها مقداری نامنفی باشد، با توجه به مفاهیم جبر خطی و معادلات ریاضی مذکور، نتیجه می‌گیریم که ماتریس لاپلاسیان مثبت نیمه‌معین است و تمامی مقادیر ویژه آن مثبت هستند. حال، فرض کنید مقادیر ویژه ماتریس لاپلاسیان به ترتیب $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ و بُردارهای ویژه یک‌ه آن‌ها به‌طور متناظر U_1, U_2, \dots, U_n باشد. در این صورت، انرژی دیریکله بُردار ویژه i ام به شکل زیر محاسبه می‌شود:

$$U_i^T . L . U_i = \lambda_i \quad (۶)$$

رابطه بالا نشان می‌دهد که بُردارهای ویژه متناظر با مقادیرهای ویژه کوچک‌تر همواری بیشتری روی گراف دارند.

۲-۴. تبدیل فوریه گرافی

برای تعریف تبدیل فوریه سیگنال x که با \hat{x} نمایش داده می‌شود به تعدادی بُردار پایه متعام در فضای سیگنال‌های گرافی نیاز داریم. ماتریس متعام $U_{n \times n}$ را از کنار هم قراردادن بُردارهای ویژه ماتریس لاپلاسیان تشکیل می‌دهیم:

$$U_{n \times n} = [U_1, \dots, U_i, \dots, U_n] \Rightarrow L = U^T . A . U \quad (۷)$$

به صورتی که A نشان‌دهنده یک ماتریس قطری با مؤلفه قطری i ام برابر λ_i است. در نتیجه، ضرایب فوریه x از مقادیر ضرب داخلی سیگنال با بُردارهای پایه به دست می‌آید.

$$\hat{x} = U^T . x \quad , \quad x = U . \hat{x} \quad (۸)$$

تبدیل فوریه \hat{x} بُرداری است که مؤلفه‌های ابتدایی آن مقدار فرکانس‌های هموار (پایین‌گذر) و مؤلفه‌های انتهایی آن مقدار فرکانس‌های ناهموار (بالاگذر) در سیگنال را نشان می‌دهند.

۲-۵. درون‌یابی سیگنال گرافی

فرض کنید مقدار سیگنال گرافی x تنها در گره‌های i_1, i_2, \dots, i_m مشخص باشد و هدف پیدا کردن مقادیر نامعلوم سیگنال روی گره‌های دیگر (درون‌یابی سیگنال)، اگر بدانیم که تنها k مؤلفه اول تبدیل فوریه \hat{x} می‌توانند ناصفر باشند می‌توانیم این مسئله را به مسئله معکوس خطی تبدیل کنیم؛

کافی است تبدیل فوریه سیگنال \mathbf{x} را به دست آوریم:

$$\begin{aligned} [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]^T &= [\mathbf{U}_1, \dots, \mathbf{U}_k, \dots, \mathbf{U}_n] \cdot [\hat{\mathbf{x}}_1 \ \dots \ \hat{\mathbf{x}}_k \ \vec{0}^T]^T \Rightarrow \\ [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]^T &= [\mathbf{U}_1, \dots, \mathbf{U}_k] \cdot [\hat{\mathbf{x}}_1 \ \dots \ \hat{\mathbf{x}}_k]^T \Rightarrow \\ [\mathbf{x}_{i_1} \ \mathbf{x}_{i_2} \ \dots \ \mathbf{x}_{i_m}]^T &= [\mathbf{U}_{i,j}]_{i \in \{i_1, \dots, i_m\}, j \in \{1, \dots, k\}} \cdot [\hat{\mathbf{x}}_1 \ \dots \ \hat{\mathbf{x}}_k]^T \Rightarrow \\ [\hat{\mathbf{x}}_1 \ \dots \ \hat{\mathbf{x}}_k]^T &= [\mathbf{U}_{i,j}]_{i \in \{i_1, \dots, i_m\}, j \in \{1, \dots, k\}}^\dagger \cdot [\mathbf{x}_{i_1} \ \mathbf{x}_{i_2} \ \dots \ \mathbf{x}_{i_m}]^T \end{aligned} \quad (9)$$

با داشتن شرط $k \geq m$ و با جای‌گذاری k مؤلفه اول تبدیل فوریه، می‌توان سیگنال اولیه \mathbf{x} را به صورت زیر به دست آورد:

$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]^T = [\mathbf{U}_{i,j}]_{i \in \{1, \dots, n\}, j \in \{1, \dots, k\}} \cdot [\mathbf{U}_{i,j}]_{i \in \{i_1, \dots, i_m\}, j \in \{1, \dots, k\}}^\dagger \cdot [\mathbf{x}_{i_1} \ \mathbf{x}_{i_2} \ \dots \ \mathbf{x}_{i_m}]^T \quad (10)$$

۲-۶. یادگیری گراف

تاکنون فرض بر این بوده است که، از قبل، از شکل گراف مطلعیم اما در بسیاری از مسائل، همچون موضوع پژوهش حاضر، گراف اولیه مشخص نیست و شکل و ارتباطات گراف بخشی از مجهول‌های مسئله است. در چنین وضعیتی، گراف را می‌توان با توجه به شباهت اجزا ساخت.

فرض کنید سیگنال نظیر گره i ام گراف یک بردار k تایی $\mathbf{x}_i \in \mathbf{R}^k$ باشد (برای مثال، نمرات دانشجویی i ام در k درس مشترک بین دانشجویان). فرض رایج برای حل مسئله یادگیری گراف این است که سیگنال گرافی روی گراف هموار است. قبلاً دیدیم که هرچه انرژی دیریکله کمتر باشد سیگنال روی گراف هموارتر است. پس، می‌توان با حل یک مسئله بهینه‌سازی شکل گرافی را (که معادل با ماتریس \mathbf{W} است) طوری پیدا کرد که انرژی دیریکله کمترین مقدار خود را بگیرد. از سوی دیگر، به منظور جلوگیری از جواب‌های بدیهی ناخواسته، از جمله $\mathbf{W}=\mathbf{0}$ ، می‌بایست عبارت‌های دیگری نیز به مسئله بهینه‌سازی اضافه شود. کالوفولیاس (Kalofolias & Perraudin, 2019; Kalofolias, 2016) با طرح مسئله زیر و با از بین بردن امکان جواب‌های بدیهی میزان \mathbf{W} را کنترل می‌کند.

$$\mathbf{W}^* = \min_{\mathbf{W}} \|\mathbf{W} \circ \mathbf{Z}\|_1 - aI^T \log(\mathbf{W}1) + \frac{\beta}{2} \|\mathbf{W}\|_F \quad (11)$$

در عبارت بالا، مؤلفه دوم اجازه نمی‌دهد مقادیر \mathbf{W} خیلی کوچک شوند و، از سوی دیگر، مؤلفه سوم اجازه نمی‌دهد مقادیر \mathbf{W} خیلی بزرگ شوند و گراف را \mathbf{W} تنگ نگه می‌دارد. کالوفولیاس نشان می‌دهد که حل این مسئله به جواب‌هایی بهتر از روش‌های دیگر یادگیری گراف منتهی می‌شود.

در مقاله پیش رو، از روش بالا برای یادگیری شکل گراف از روی نمرات دانشجویان استفاده می‌کنیم. پس از تخمین ماتریس وزن گراف، برای اینکه گراف بیشتر تُنک باشد، می‌توان یک آستانه^۱ تجربی D تعیین کرد و وزن یال‌هایی را که از این مقدار کمتر شده‌اند صفر کرد:

$$w_{i,j} = \{w_{i,j}^* \text{ if } w_{i,j}^* \geq D, \quad 0 \text{ if } w_{i,j}^* < D\} \quad (۱۲)$$

۳. معرفی داده‌ها و روش پیشنهادی

در پژوهش حاضر، به منظور تحلیل نمرات و عملکرد دانشجویان، گرافی را برای هر دوره دانشجویان هم‌ورودی دانشکده مهندسی برق دانشگاه صنعتی شریف در نظر می‌گیریم. در این گراف، گرهی متناظر با هر دانشجو وجود دارد. در صورتی که دو دانشجو به لحاظ عملکرد تحصیلی وضعیتی نسبتاً مشابه داشته باشند میان گره‌های متناظر آنان یک یال در نظر می‌گیریم که وزن آن نشان‌دهنده میزان شباهت وضعیت تحصیلی دو دانشجوست. در مرحله اول پژوهش، گرافی برای دانشجویان هر ورودی تشکیل می‌دهیم و سپس روی آن تحلیل‌های سه‌گانه می‌کنیم. در ادامه، به شرح مراحل و بررسی نتایج می‌پردازیم.

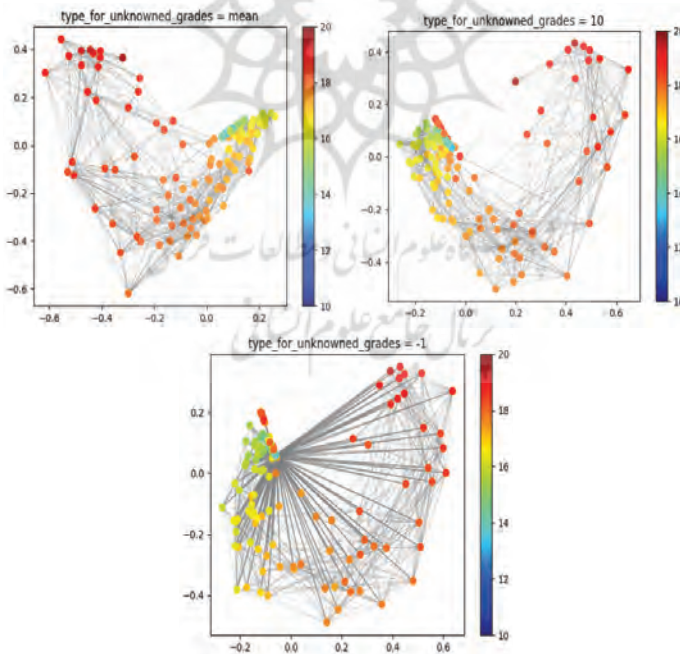
۱-۳. معرفی داده‌ها و سبک ارائه دروس در دانشکده مهندسی برق در پژوهش حاضر، از یک مجموعه داده جامع مربوط به دانشجویان دانشکده مهندسی برق دانشگاه شریف استفاده کرده‌ایم که شامل ۵ مشخصه است: مشخصه اول شماره ترم تحصیلی است. برای مثال، ترم پاییز (ترم اول) سال تحصیلی ۱۳۹۲-۱۳۹۳ را با عدد ۱۳۹۲۱ نمایش داده‌ایم. مشخصه دوم شماره درس، مشخصه سوم شماره دانشجویی رمز شده (برای حفظ حریم شخصی دانشجو)، مشخصه چهارم سال ورود دانشجو به دانشگاه، و مشخصه نهایی نمره اکتسابی است. این مجموعه داده تمامی نمرات دانشجویان کارشناسی مهندسی برق را در همه دروسی که در سال‌های تحصیلی ۱۳۹۰-۱۳۹۱ تا ۱۳۹۹-۱۴۰۰ گذرانده‌اند در بر می‌گیرد. بدیهی است که تعدادی از نمرات مربوط به دانشجویان ورودی نیمه دوم دهه ۸۰ خورشیدی است که از آنها استفاده نکرده‌ایم.

شایان ذکر است که در دانشکده مهندسی برق دانشگاه صنعتی شریف دانشجویان دوره کارشناسی را بدون شاخه‌ای مشخص شروع می‌کنند و این وضعیت تا پنج ترم ادامه دارد و در طول این پنج ترم دروس پایه مهندسی برق را (از تمامی شاخه‌ها) فرامی‌گیرند. در پایان ترم پنجم، بیشتر دانشجویان اولویت‌های تحصیلی خود را از شاخه‌های شش‌گانه مهندسی برق اعلام می‌کنند. سپس، معاونت آموزشی، با توجه به ظرفیت محدود شاخه‌ها، با در اولویت قراردادن دانشجویانی که عملکرد تحصیلی

بهتری در پنج ترم ابتدایی داشته‌اند، دانشجویان را در شش شاخه تقسیم‌بندی می‌کند. در طی پژوهش حاضر، شاخه‌ها به صورت A تا F نام‌گذاری شده‌اند.

۲-۳. روش تحلیل داده‌ها

راهبرد کلی پژوهش پیش رو این است که برای هر دوره از ورودی‌های دانشگاه یک گراف یاد گرفته شود. بدین منظور، برای هر دوره ماتریسی تشکیل داده‌ایم که یکی از ابعاد آن مربوط به دانشجویان دوره و بُعد دیگر مربوط به دروس ایشان است. سپس، دروسی را که کمتر از ۸۵ درصد از دانشجویان در آن مشارکت داشته‌اند حذف کرده‌ایم زیرا این دروس معیار مناسب ساخت گراف نیستند. در دروس باقی‌مانده، همچنان دانشجویانی هستند که یک/چند درس از این دسته را هیچ‌گاه نگذرانده‌اند. برای این افراد چه نمره‌ای در نظر گرفته شود؟ چند حالت را بررسی کرده‌ایم: نمره ۱-، نمره ۱۰، میانگین سایر دانشجویان در درس مذکور، و میانگین نسبی نمرات. در نهایت، روش چهارم را برگزیده‌ایم که درباره آن توضیح می‌دهیم. قبل از توضیح روش چهارم، در شکل ۲ به بررسی نتایج سه روش اول برای دانشجویان ورودی ۱۳۹۵ می‌پردازیم. (رنگ هر دانشجو نشانه معدل کل آن دانشجو است).



شکل ۲. گراف دانشجویان ورودی ۱۳۹۵ با سه فرض متفاوت برای نمرات نامعلوم: نمره ۱-، نمره ۱۰، میانگین [سایر دانشجویان در آن درس]، به ترتیب از چپ به راست.

توجه کنید که پس از تشکیل گراف، برای رسم، می‌بایست به هر فرد (گره) یک مشخصه دو بُعدی نسبت دهیم. بدین منظور، از الگوریتم MDS^۱ استفاده کرده‌ایم (Parhizkar, 2013) که گره نظیر دانشجویانی را که در بُردار نمرات خود مشابه هستند در نزدیکی یکدیگر قرار می‌دهد. به عبارتی، اگر برای هر دانشجو یک بُردار از نمرات با بُعد تعداد دروس مشترک دانشجویان (دست‌کم ۸۵ درصد مشارکت) در نظر بگیریم فاصله گره‌ها در نمایش به نوعی فاصله اقلیدسی بُردارها را از طریق وزن یال بین گره‌ها نمایان می‌کند.

همان‌طور که در شکل ۳ می‌بینید، در وضعیتی که نمرات نامعلوم را برابر ۱- قرار دهیم رُئوس پُردرجه تشکیل می‌شوند که نامطلوب است و با تصورات ما از مسئله اصلی همخوانی ندارد. در آزمایش‌های ما، بهترین روش بدین شرح به دست آمد که نمرات نامعلوم با روش میانگین نسبی (به منظور کاهش اثر بزرگی مطلق داده‌ها و تبدیل آنها به مقیاسی یکسان) محاسبه شوند:

$$a = \text{نمره نامعلوم دانشجوی } x \text{ در درس } x \text{ (مجهول)}$$

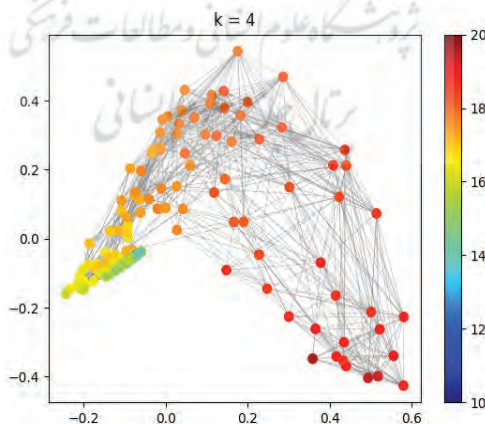
$$b = \text{متوسط نمره دانشجویانی که درس } x \text{ را داشته‌اند در درس } x$$

$$c = \text{متوسط معدل کل همه دانشجویان}$$

$$d = \text{معدل کل دانشجوی } x$$

$$a = b \times \frac{c}{d} \quad (13)$$

حاصل یادگیری گراف بدین نحو را در شکل ۳ مشاهده می‌کنید. این گراف مشکلات روش‌های سابق را، از جمله «رُئوس پُردرجه» و «وجود یال بین دانشجویان با معدل‌های بسیار متفاوت»، ندارد.

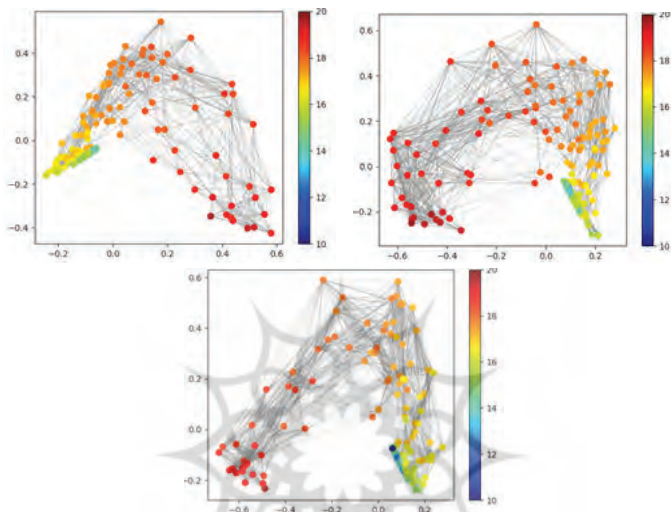


شکل ۳. گراف یادگرفته شده برای دانشجویان ورودی ۱۳۹۵، با استفاده از روش میانگین نسبی برای تخمین نمرات نامعلوم.

۴. نتایج تحلیل

۴-۱. نمایش گراف و تعبیر آن

در تصاویر زیر، گراف آموخته شده برای سه دوره از دانشجویان هم‌ورودی دانشکده مهندسی برق را نمایش داده‌ایم.



شکل ۴. گراف یادگرفته شده دانشجویان ورودی ۱۳۹۴، ۱۳۹۵ و ۱۳۹۷ (به ترتیب از راست به چپ).

شایان ذکر است که گراف‌ها با استفاده از الگوریتم MDS رسم شده‌اند که در بخش تحلیل داده‌ها به مزیت استفاده از آن اشاره کردیم. از شکل ۴ دو نتیجه حاصل می‌شود: اول آنکه، همان‌طور که می‌بینید، دانشجویان با معدل پایین در همه دوره‌ها فاصله بسیار نزدیکی به یکدیگر دارند اما فواصل دانشجویان با معدل بالا نسبتاً دورتر است. توجه کنید که اگر دو دانشجوی هم‌معدل از هم فاصله زیادی داشته باشند بدین معناست که هرچند این دو فرد به صورت متوسط عملکرد مشابهی دارند در دروس متفاوت نمرات بسیار متفاوتی کسب کرده‌اند؛ به بیان دیگر، دروسی که نقطه قوت یکی است نقطه ضعف دیگری است و برعکس. به علاوه، تمایز با افراد هم‌معدل را می‌توان تفسیر به شکوفایی تحصیلی کرد. نتیجه موارد یادشده این است که متأسفانه، در نظام آموزشی فعلی، افراد با متوسط عملکرد ضعیف هیچ‌گونه شکوفایی‌ای در برابر هم‌نوعان خود ندارند و در بیشتر دروس مشابه یکدیگر عمل می‌کنند.

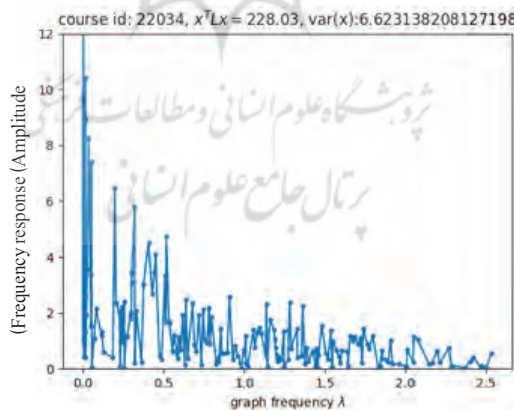
دومین نکته این است که افراد با معدل بالا در ورودی‌های ۱۳۹۴ و ۱۳۹۵ فواصل اقلیدسی نسبتاً زیادی با یکدیگر دارند اما این فاصله برای دانشجویان ورودی ۱۳۹۷ (که بیش از نیمی از تحصیل ایشان در دوره همه‌گیری ویروس کرونا گذشت) به وضوح کاهش یافته است بدین معنی که افراد با معدل

بالا مشابهت زیادی در عملکرد تحصیلی خود با ورودی‌های قبل پیدا کرده‌اند. این افراد به صورت هماهنگ در تعدادی از دروس عملکرد خوبی دارند و در تعدادی دیگر عملکرد بد. تفاوت مذکور در شکل گراف متناظر این ورودی را می‌توان ناشی از تقلب در آزمون‌های مجازی دانست.

۴-۲. عملکرد دانشجویان در دروس و توجیه آن

یکی از تحلیل‌های داده‌ها بررسی همواری/ناهمواری نمرات دروس روی گراف یادگیری شده است. در این تحلیل، مشخص می‌شود که نمرات کدام دروس روی گراف بسیار هموار و نمرات کدام دروس بسیار ناهموار است. به عبارتی، این تحلیل بازخوردی برای استادان دانشکده است که نمره‌دهی و ارزش‌یابی ایشان (بدین شرط که عملکرد کلی دانشجویان در آن درس تغییر چشمگیری نکرده باشد) تا چه حد با عملکرد کلی دانشجویان منطبق است. برای مثال، برای دانشجویان ورودی ۱۳۹۲ این ارزیابی صورت گرفته و انرژی سیگنال دروس مشترک روی گراف دانشجویان محاسبه شده است. سپس، دروس بسیار هموار و دروس بسیار ناهموار استخراج شده‌اند.

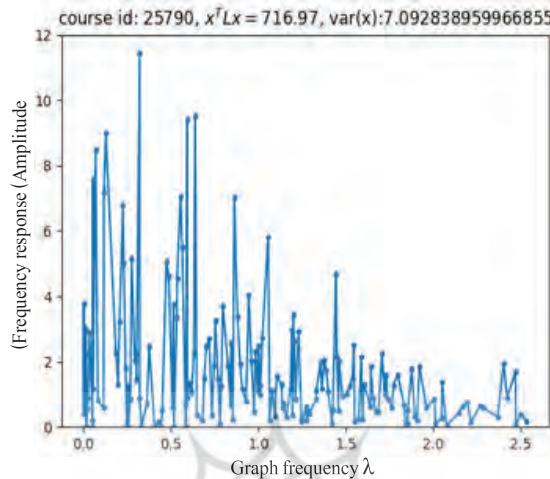
دروس با نمرات هموار بر گراف دانشجویان ورودی ۱۳۹۲ عبارت‌اند از: معادلات دیفرانسیل، آزمایشگاه فیزیک ۱، اصول مهندسی برق و کارگاه عمومی. برای نمونه، تبدیل فوریه سیگنال درس معادلات دیفرانسیل را در تصویر ۵ مشاهده می‌کنید. همان‌طور که می‌بینید، این خروجی بسیار پایین‌گذر است و در فرکانس‌های پایین مقادیر بزرگ دارد.



شکل ۵. تبدیل فوریه درس معادلات دانشجویان ورودی ۱۳۹۲.

دروس با نمرات بسیار ناهموار روی گراف دانشجویان ورودی ۹۲ عبارت‌اند از: مدارهای آنالوگ و آزمایشگاه، تحلیل سیستم‌های انرژی الکتریکی ۱ و پروژه کارشناسی ۲. برای نمونه، تبدیل فوریه سیگنال درس پروژه کارشناسی ۲ را در شکل ۶ مشاهده می‌کنید. همان‌طور که می‌بینید، در فرکانس‌های بالا نیز

مقادیر بزرگ مشهود است.



شکل ۶. تبدیل فوریه درس پروژه کارشناسی ۲ دانشجویان ورودی ۱۳۹۲.

ازجمله حدسیات درخصوص ناهمواری دروس یادشده این است که همه این دروس در چندین گروه با استادان متفاوت گذرانده می‌شوند و احتمالاً سیاست نمره‌دهی یکسانی استفاده نمی‌شود. البته، در درس پروژه کارشناسی ۲ که در پایان دوره کارشناسی گذرانده می‌شود تعداد درخور توجهی از دانشجویان درگیر مسائل گرفتن پذیرش تحصیلی از خارج کشور یا آماده‌شدن برای شرکت در کنکور کارشناسی ارشد هستند و در نتیجه تعداد زیادی از هر دو دسته دانشجویان قوی/ضعیف این درس را جدی نمی‌گیرند و در آن عملکرد مطلوبی ندارند. در عوض، درس معادلات دیفرانسیل وضعیتی بسیار متفاوت با درس پروژه کارشناسی ۲ دارد: در این درس که در ترم‌های ابتدایی گذرانده می‌شود، همه دانشجویان به‌طور جدی مشغول تحصیل‌اند و نظام ارزش‌یابی و نمره‌دهی آن به سلیقه آموزگار درس ارتباطی ندارد و دانشجویان در وضعیتی مشابه ارزیابی می‌شوند.

۳-۴. بررسی میزان موفقیت دانشجویان در انتخاب شاخه

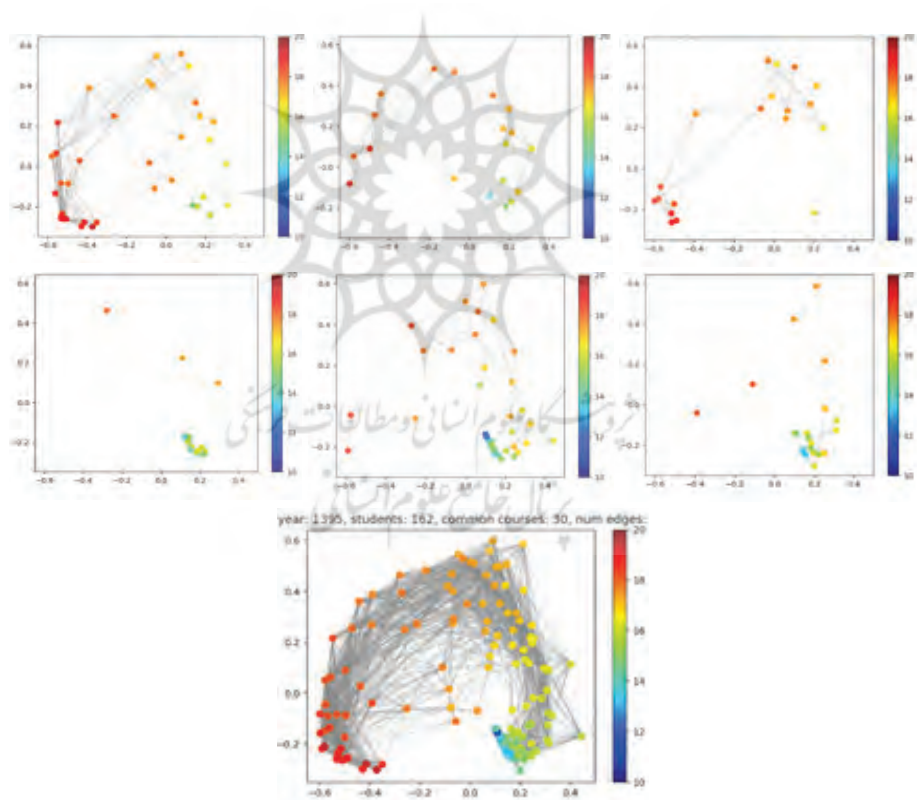
همان‌طور که اشاره کردیم، در رشته مهندسی برق دانشگاه صنعتی شریف، شاخه پس از ترم پنجم انتخاب می‌شود. در تحلیل حاضر، اولین گام مشخص کردن شاخه‌های دانشجویان است. در برنامه آموزشی دانشکده، هر شاخه شامل ۳ درس اجباری و ۵ درس اختیاری می‌شود که دانشجویان می‌بایست از میان آنها دو درس انتخاب کنند. برای تعیین شاخه دانشجویان، به هر دانشجو در هر شاخه عددی نسبت می‌دهیم که این عدد برابر است با:

تعداد دروس اجباری گذرانده شده دانشجو از شاخه مذکور + ۰٫۵ × تعداد دروس انتخابی گذرانده شده دانشجو از شاخه مذکور).

در اصل، در این روش وزن دروس اجباری دو برابر دروس انتخابی است.

حال، برای هر دانشجو ۶ عدد محاسبه می شود و شاخه هر دانشجو همان شاخه ای است که در آن بیشترین مقدار را کسب کرده است.

ممکن است سؤال مطرح شود که چرا صرفاً به دروس اجباری گذرانده شده دانشجویان توجه نمی کنیم. شایان ذکر است که هر دانشجو می تواند برخی از دروس تخصصی سایر شاخه ها را زیر عنوان «واحد اختیاری» اخذ کند. بنابراین، مجبور به استفاده از روش مذکور شده ایم. نتیجه گرایش بندی دانشجویان ورودی ۹۵ برای شش شاخه از A تا F و تعداد دانشجویان به ترتیب ۳۷، ۲۱، ۱۹، ۱۴، ۴۵ و ۲۳ به شرح زیر است:



شکل ۷. نتیجه شاخه بندی بر مبنای تعداد دروسی که دانشجویان از شاخه های گوناگون گذرانده اند، برای ورودی های ۱۳۹۵. تصویر سمت پایین گراف جامع دانشجویان ورودی ۱۳۹۵ است. سطر بالا، از چپ به راست، به ترتیب زیرگراف نظیر دانشجویان گرایش A و B را نمایش می دهد. سطر پایین، از چپ به راست، به ترتیب زیرگراف نظیر دانشجویان D، E و F را نمایش می دهد.

تحلیل بدین شرح است که معدل دانشجویان از شاخه‌های گوناگون دروس تخصصی محاسبه شده است. حال، معدل دانشجویان شاخه B را در دروس تخصصی شان در نظر بگیرید. این اعداد در اصل سیگنال‌هایی ناقص روی گراف دانشجویان هستند که مقدار آن فقط برای دانشجویان شاخه B مشخص است و برای سایر دانشجویان مشخص نیست؛ یعنی نمی‌دانیم دانشجویان سایر شاخه‌ها اگر دروس تخصصی شاخه B را می‌گذرانند چه معدلی کسب می‌کردند. برای حل این مسئله، می‌توان از درون‌یابی سیگنال‌های گرافی استفاده کرد تا معدل دانشجویان سایر شاخه‌ها را در دروس شاخه B به دست آورد. به همین ترتیب، می‌توان معدل هر دانشجو را در دروس تخصصی شاخه‌های دیگر محاسبه کرد. درحقیقت، براساس نمرات تخصصی هر دانشجو در شاخه خودش، مقایسه آن با سایر دانشجویان آن شاخه و همچنین نمرات همه دانشجویان در دروس اصلی مشترک، پیش‌بینی می‌کنیم که او در پنج شاخه دیگر چه معدلی در دروس تخصصی کسب می‌کرد. اکنون می‌توان این شش عدد را مقایسه و تعیین کرد که بهترین انتخاب برای این دانشجو کدام شاخه بوده و آیا شاخه خود را به‌درستی انتخاب کرده است یا خیر. بدین ترتیب، ماتریسی به دست می‌آید که در جدول ۱ مشاهده می‌کنید.

هر سطر متناظر با دانشجویان یکی از شاخه‌هاست. بنابراین، در هر سطر در سمت چپ، ابتدا اسم شاخه دانشجویان متناظر آن سطر را مشاهده می‌کنید. اعداد هر سطر نشان‌دهنده درصد دانشجویانی از این شاخه است که بهترین وضعیت تحصیلی را در شاخه متناظر با هر ستون می‌داشتند. بدیهی است که جمع اعداد هر سطر برابر ۱۰۰ درصد است.

جدول ۱. تحلیل انتخاب شاخه دانشجویان ورودی ۱۳۹۵

شاخه مناسب شاخه فعلی	A	B	C	D	E	F
A	۵۷ درصد	۳۲ درصد	۰ درصد	۰ درصد	۰ درصد	۱۱ درصد
B	۱۹ درصد	۷۱ درصد	۰ درصد	۰ درصد	۰ درصد	۱۰ درصد
C	۳۲ درصد	۱۰ درصد	۴۷ درصد	۰ درصد	۰ درصد	۱۱ درصد
D	۶۴ درصد	۷ درصد	۰ درصد	۲۲ درصد	۰ درصد	۷ درصد
E	۴۹ درصد	۱۸ درصد	۰ درصد	۰ درصد	۲۷ درصد	۶ درصد
F	۳۵ درصد	۹ درصد	۰ درصد	۰ درصد	۴ درصد	۵۲ درصد

* نتایج محاسبه‌ها نشان می‌دهد در بین دانشجویان این دوره ۴۴٫۴۴ درصد از دانشجویان شاخه صحیح انتخاب کرده‌اند.
* اعداد بدون اعشار ثبت شده‌اند. (گرد شده‌اند.)

برای راهنمایی استفاده از جدول، فرض کنید می‌خواهیم ببینیم چند درصد از دانشجویان گرایش B شاخه مناسبشان (که متأسفانه آن را انتخاب نکرده‌اند) A بوده است. بدین منظور، می‌بایست اعداد سطر دوم را که متناظر با گرایش انتخابی B هستند در نظر بگیریم. اکنون، شاخه A، به‌عنوان شاخه مطلوب، متناظر اولین عدد از سمت چپ در این سطر است که حدود ۱۹ درصد است. بدیهی است اگر

همه دانشجویان انتخاب واحد درستی داشتند انتظار حصول ماتریس همانی را داشتیم. همان طور که می بینید، حدود ۴۴ درصد از دانشجویان انتخاب های مناسبی داشته اند. این داده ها در برنامه ریزی برای برنامه درسی دانشکده و تمهیدات انتخاب شاخه بسیار مؤثر و کاربردی هستند. درنهایت، محاسبه کرده ایم که اگر همه دانشجویان شاخه صحیح انتخاب می کردند (با فرض نبود محدودیت بر تعداد دانشجویان شاخه ها) هر شاخه میزبان چند درصد از کل دانشجویان ورودی می شد. نتیجه را در جدول ۲ مشاهده می کنید که تا حدی نشان دهنده وضعیت شاخه ها در دانشکده است.

جدول ۲. درصد صندلی های هر شاخه در صورت انتخاب شاخه صحیح توسط همه دانشجویان (ورودی ۱۳۹۵)

شاخه	درصد صندلی های دانشکده
A	۴۴
B	۲۵
C	۶
D	۲
E	۸
F	۱۵

* اعداد بدون اعشار ثبت شده اند. (گرد شده اند).

۴-۴. ارزیابی عملکرد در دوره کرونا

یکی از اهداف پژوهش پیش رو بررسی عملکرد دانشجویان در بازه های زمانی متفاوت بوده است. بدین منظور، معدل دانشجویان را در هر نیمسال آموزشی یک سیگنال گرافی در نظر می گیریم. برای مقایسه عملکرد تحصیلی هر دانشجو در هر نیمسال در مقایسه با دانشجویان همسایه در گراف (که در دروس مشترک عملکرد مشابه داشته اند). پارامتر q را مطابق زیر تعریف می کنیم. (چنانچه رابطه انرژی سیگنال را به یاد آورید متوجه خواهید شد که در حال محاسبه انرژی مدنظر به صورت محلی هستیم).

$$q_{student\ i, term\ j} = \frac{1}{\# neighbors\ of\ i} \times \sum_{k \in neighbors\ of\ i} \left((GPA_{i,j} - GPA_{k,j})^2 \times weight_{i,k} \right) \quad (14)$$

دقت کنید که این متغیر مقداری نامنفی است و میزان بزرگی آن نشان دهنده میزان تفاوت عملکرد دانشجو با همسایگان خود است اما مشخص نیست که این عملکرد با هدف بهبود و جهش تحصیلی

بوده است یا افت تحصیلی. برای مشخص شدن هدف، صرفاً می‌بایست برای متغیر q یک علامت تعیین شود. برای تعیین علامت، از مقدار زیر استفاده می‌کنیم:

$$flag_{student\ i, term\ j} = sign\left(\frac{1}{\# neighbors\ of\ i} \times \sum_{k \in neighbors\ of\ i} ((GPA_{i,j} - GPA_{k,j}) \times weight_{i,k})\right) \quad (15)$$

بنابراین، متغیر نهایی \mathcal{F} برای نمایش عملکرد علامت‌دار دانشجوی، در مقایسه با همسایگانش، به صورت زیر تعریف می‌شود:

$$\mathcal{F}_{student\ i, term\ j} = q_{student\ i, term\ j} \times flag_{student\ i, term\ j} \quad (16)$$

برای هر دانشجو و در هر نیمسال، پارامتر \mathcal{F} را محاسبه و سپس متوسط عملکرد وی را در نیمه اول مدت‌زمان تحلیل از متوسط عملکرد وی در نیمه دوم مدت‌زمان تحلیل کم می‌کنیم. اگر این مقدار مثبت باشد بدین معناست که دانشجو در مدت‌زمان تحلیل رشد تحصیلی داشته یا، به عبارتی، در نیمه دوم زمان تحلیل عملکردی بهتر از نیمه اول داشته است. هرچه این عدد بزرگ‌تر باشد یعنی رشد تحصیلی دانشجو بیشتر بوده است. اگر این مقدار منفی باشد بدین معناست که دانشجو در مدت‌زمان تحلیل افت تحصیلی داشته یا، به عبارتی، در نیمه دوم زمان تحلیل عملکردی بدتر از نیمه اول داشته است. به علاوه، هرچه این عدد کمتر باشد (اندازه آن بزرگ‌تر باشد)، دانشجو افت تحصیلی بیشتری داشته است.

با تعیین تجربی دو آستانه می‌توان دانشجویان را به سه دسته تقسیم کرد: دانشجویانی که در نیمه دوم زمان تحلیل رشدی درخور توجه‌تر از نیمه اول داشته‌اند؛ دانشجویانی که در نیمه دوم زمان تحلیل افت درخور توجه‌تر از نیمه اول داشته‌اند، و دانشجویانی که افت/رشد درخور توجهی نداشته‌اند.

این تحلیل برای دانشجویان دوره‌های متفاوت در شش نیمسال ابتدایی انجام گرفته است که نتیجه آن را در جدول ۳ مشاهده می‌کنید. یادآوری می‌کنیم که سه نیمسال تحصیلی دانشجویان ورودی ۱۳۹۷ قبل از همه‌گیری کرونا بوده است و سه نیمسال تحصیل مجازی داشته‌اند. بنابراین، نتایج این دوره درخور توجه است. (علت بررسی نکردن نیمسال هفتم و هشتم این دانشجویان اعتبار کافی نداشتن نمرات این دو نیمسال، به دلیل گرفتاری آنان به شرکت در کنکور کارشناسی ارشد یا مقدمات گرفتن پذیرش تحصیلی خارج از کشور است.)

همان‌طور که مشاهده می‌کنید، به طور عمومی در هر ورودی حدود ۶ درصد از دانشجویان رشد درخور توجه و ۶ درصد افت درخور توجه دارند و حدود ۹۰ درصد نه افت و نه رشد درخور توجهی دارند. در دوره ۹۷ که در میانه مسیر همه‌گیری کرونا آغاز شده است درصد افرادی که رشد چشمگیر داشته‌اند به ۱۱ رسیده است که می‌توان بخشی از آن را به ثقل در آزمون‌های مجازی نسبت داد. همچنین درصد افرادی که افت درخور توجه تحصیلی داشته‌اند به ۱۴ رسیده است که می‌توان آن را ناشی از افت کیفیت آموزش یا افسردگی دانست.

جدول ۳. تحلیل ترمی زمانی دانشجویان دوره‌های متفاوت

سال ورودی	درصد رشد تحصیلی	درصد افت تحصیلی
۱۳۹۳	۴ درصد	۵ درصد
۱۳۹۴	۵ درصد	۶ درصد
۱۳۹۵	۷ درصد	۹ درصد
۱۳۹۶	۷ درصد	۶ درصد
۱۳۹۷	۱۱ درصد	۱۴ درصد

۵. جمع‌بندی و پیشنهادها

در پژوهش حاضر، نمرات چندین سال اخیر دانشجویان دانشکده مهندسی برق دانشگاه صنعتی شریف را بررسی و بدین منظور از روش‌های پردازش سیگنال گرافی استفاده کرده‌ایم. در این بخش، صرفاً نتایج و تحلیل‌ها را مرور خواهیم کرد.

۱. تحلیل‌ها نشان می‌دهند نمرات تعدادی از دروس با وضعیت تحصیلی کلی دانشجویان تطابق ندارد. احتمالاً دلیل این امر ارائه دروس در گروه‌های متعدد و مشابهت نداشتن سیاست نمره‌دهی استادان است. بدیهی است که برای اصلاح این رویه می‌توان به آموزش و استادان مراجعه کرد.

۲. تحلیل‌ها نشان دادند که تنها ۴۰ تا ۵۰ درصد از دانشجویان شاخه تحصیلی خود را درست انتخاب می‌کنند. برای رفع این مشکل، ضروری است پیش از زمان انتخاب شاخه جلسات تخصصی معرفی شاخه‌ها برای دانشجویان برگزار شود. به‌طور تخصصی‌تر، نتایج نشان داد که دانشجویان برخی شاخه‌ها را نادیده می‌گیرند درحالی‌که بسیاری از دانشجویان، در صورت انتخاب این شاخه‌ها، عملکرد بسیار بهتری خواهند داشت. این تحلیل آموزش دانشکده را مجاب می‌کند که دروس بیشتری را از شاخه‌های مذکور در برنامه پنج نیمسال ابتدایی قرار دهد تا این احتمال که دانشجویان شاخه مناسبی انتخاب کنند افزایش یابد.

۳. در پژوهش حاضر، با استفاده از روش‌های پردازش سیگنال‌های گرافی نشان دادیم که در صورتی که همه دانشجویان شاخه درستی انتخاب کنند چه درصدی از صندلی‌های دانشکده می‌بایست به هر شاخه اختصاص یابد.

۴. در نهایت، آماری از درصد رشد/افت چشمگیر تحصیلی دانشجویان را در ورودی‌های متفاوت محاسبه کردیم که به خودی خود درخور توجه و تحلیل است و البته اثر بحران همه‌گیری کرونا را نیز بر این متغیرها مشاهده کردیم. همه‌گیری کرونا باعث هر دو پدیده رشد و افت چشمگیر تحصیلی بیش از دو برابر دانشجویان شده است. بدیهی است بخش آموزش کل دانشگاه می‌تواند درستی نمرات دانشجویانی را که در دوره همه‌گیری کرونا رشد تحصیلی درخور توجهی داشته‌اند بسنجد.

سپاسگزاری

از جناب آقای دکتر حمید ضرابی زاده، مدیر آموزش وقت دانشگاه صنعتی شریف، بابت اعتماد ایشان به در اختیار قراردادن دادگان مورد نیاز نویسندگان پژوهش، صمیمانه سپاسگزاریم.

References

- Atif, A., Richards, D., Bilgin, A., & Marrone, M. (2013). Learning analytics in higher education: A summary of tools and approaches. *30th Australasian Society for Computers in Learning in Tertiary Education Conference*, Sydney.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42(4), 40–57.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>.
- Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 12(1), 17–26.
- Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society*, 15(3), 58–76.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 235. <https://doi.org/10.3390/info1504023>.
- Hamim, T., Benabbou, F., & Sael, N. (2022). Student profile modeling using boosting algorithms. *International Journal of Web-Based Learning and Teaching Technologies*, 17(5), 1–13. <https://doi.org/10.4018/IJWLTT.293281>.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer. <https://doi.org/10.1007/978-0-387-84858-7>.
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, 13(4), 18–28. <https://doi.org/10.1109/5254.708428>.
- Kalofolias, V. (2016). How to learn a graph from smooth signals. *Journal of Machine Learning Research*, 17(1), 1–21.
- Kalofolias, V., & Perraudin, N. (2019). Large scale graph learning from smooth signals. *International Conference on Learning Representations*.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30).
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* (Vol. 30).
- Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51. <https://doi.org/10.1016/j.dss.2018.09.001>.
- Mingyu, Z., Sutong, W., Yanzhang, W., & Dujuan, W. (2021). An interpretable prediction method for university student academic crisis warning. *Complex & Intelligent Systems*, 8(1), 323–336. <https://doi.org/10.1007/s40747->

- 021-00566-7.
- Ortega, A. (2022). *Introduction to graph signal processing*. Cambridge University Press. <https://doi.org/10.1017/9781108889862>.
 - Parhizkar, R. (2013). *Euclidean distance matrices: Properties, algorithms, and applications* (PhD thesis). EPFL, Lausanne.
 - Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (Vol. 31) .
 - Rish, I. (2001). Empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence* (Vol. 3, No. 22, pp. 41-46) .
 - Sandryhaila, A., & Moura, J. M. F. (2013). Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*, 61(7), 1644-1656. <https://doi.org/10.1109/TSP.2013.2238935>.
 - Sandryhaila, A., & Moura, J. M. F. (2014). Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine*, 31(5), 80-90. <https://doi.org/10.1109/MSP.2014.2329213>.
 - Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications, and research directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>.
 - Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3), 83-98. <https://doi.org/10.1109/MSP.2012.2235192>.
 - Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, 16(3), 51-61.
 - Song, Y. Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130-135. <https://doi.org/10.11919/j.issn.1002-0829.215044>.
 - Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica (Zagreb)*, 24(1), 12-18. <https://doi.org/10.11613/BM.2014.003>.
 - van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*, 1, 1-11.



◀ امیرحسین گل شیرازی: در ۱۴۰۲ مدرک کارشناسی مهندسی برق را در شاخه سیستم‌ها و شبکه‌های مخابراتی از دانشگاه صنعتی شریف دریافت کرد. سپس، بلافاصله تحصیل در مقطع کارشناسی ارشد مهندسی برق را در همان شاخه و دانشگاه شروع کرد و هم‌اکنون در حال سپری کردن این دوره تحصیلی است. حیطه‌های پژوهشی موردعلاقه او شامل دو گروه می‌شود: گروه اول، تلاقی تئوری اطلاعات، آمار و یادگیری ماشین و، گروه دوم، مباحث تئوری بازی، تصمیم‌گیری بهینه و یادگیری تقویتی.

◀ آرش امینی: مدرک کارشناسی خود را در رشته مهندسی برق (سیستم‌های مخابراتی و پردازش سیگنال) و همچنین مهندسی



نفت (مخازن) و مدارک کارشناسی ارشد و دکترای خود را در مهندسی برق (سیستم‌های مخابراتی و پردازش سیگنال) از دانشگاه صنعتی شریف (تهران، ایران)، به ترتیب در ۱۳۸۴، ۱۳۸۵ و ۱۳۸۹، دریافت کرده است. وی از ۱۳۹۰ تا ۱۳۹۲، در مدرسه پلی‌تکنیک فدرال لوزان (EPFL) سوئیس، در مقام پژوهشگر فعالیت داشته است. در ۱۳۹۲، استادیار دانشگاه صنعتی شریف شد و از ۱۳۹۷ تاکنون دانشیار دانشگاه مذکور است.



◀ **رضا پرهیزکار:** مدرک کارشناسی خود را در مهندسی برق از دانشگاه صنعتی شریف تهران در ۱۳۸۶ و مدرک کارشناسی ارشد خود را در سیستم‌های مخابراتی از مؤسسه فدرال فناوری سوئیس (EPFL)، لوزان، در ۱۳۸۸ دریافت کرده است. در مقام کارآموز، شش ماه در مرکز تحقیقاتی نوکیا، لوزان، در ۱۳۸۸ و سه ماه در شرکت کوالکام (Qualcomm Inc.)، سن دیگو، در ۱۳۹۲ فعالیت داشته است. از ۱۳۸۸، در مقام دانشجوی دکتری در آزمایشگاه ارتباطات صوتی و تصویری (LCAV) در EPFL فعالیت می‌کند. پژوهش او در زمینه کالیبراسیون حسگرها برای دستگاه‌های توموگرافی اولتراسونیک جایزه بهترین مقاله دانشجویی را از کنفرانس ICASSP ۲۰۱۱ دریافت کرده است. علاقه‌مندی‌های پژوهشی وی شامل پردازش سیگنال ریاضی، مکان‌یابی حسگرها و مسائل معکوس است.



◀ **محمد مهدی امتی:** مدرک کارشناسی خود را در مهندسی برق (سیستم‌های مخابراتی و پردازش سیگنال) از دانشگاه علم و صنعت ایران، تهران، در ۱۳۹۹ و مدرک کارشناسی ارشد خود را در مهندسی برق (سیستم‌های مخابراتی و پردازش سیگنال) از دانشگاه صنعتی شریف، تهران، در ۱۴۰۱ دریافت کرده است. در حال حاضر، دستیار پژوهشی آزمایشگاه چندرسانه‌ای و پردازش سیگنال دانشگاه صنعتی شریف است. علاقه‌مندی‌های پژوهشی او شامل بهینه‌سازی (غیر)محدب، پردازش سیگنال‌های آماری و پردازش سیگنال روی گراف است.