

به نام ایزد منان

تمرین سوم درس بازیابی اطلاعات، «سامانه‌های توصیه‌گر»



استاد درس: دکتر ممتازی

پاییز ۱۴۰۱ - دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر



نکاتی در مورد این تمرین که نیاز به توجه و دقت دوستان دارد:

۱- در جدول زیر نحوه اعمال جریمه تاخیر در ارسال تمرین‌ها ذکر شده است.

میزان جریمه	میزان تاخیر (روز)
هر روز ۵٪	۱ الی ۲ روز
هر روز ۱۰٪	۲ الی ۶ روز

در صورتی که بین ۷ تا ۱۴ روز تاخیر داشته باشید، نمره شما از ۵۰٪ محاسبه می‌شود و پس از این بازه با توجه به سایر تمرین‌ها و زمان تحویل، به تمرین ارسالی نمره‌ای تعلق نمی‌گیرد.

۲- هرگونه کپی کردن باعث عدم تعلق نمره به تمامی افراد مشارکت کننده در آن می‌شود.

۳- آخرین مهلت ارسال تمرین، ساعت ۲۳:۵۵ دقیقه روز ۲۰ بهمن می‌باشد. این زمان با توجه به جمع‌بندی‌های صورت گرفته، شرایط و با توجه به سایر تمرین‌ها در نظر گرفته شده است و قابل تمدید نمی‌باشد.

۴- دوستان فایل ارسالی خود را به صورت فشرده و به صورت «شماره دانشجویی_HW01» مانند HW01_99131123 نام گذاری کنید. در این فایل باید مواردی نظیر کدها، فایل گزارش و سایر موارد مورد نیاز در هنگام بررسی وجود داشته باشد و صرفاً این فایل در روز ارائه در نظر گرفته می‌شود.

۵- این تمرین دارای تحویل در محیط گوگل میت می‌باشد. زمان آن پس از یک هفته از پایان مهلت تمرین از طریق مودل درس اعلام می‌شود.

۶- زبان برنامه‌نویسی این تمرین می‌تواند پایتون، سی‌پلاس‌پلاس و یا جاوا باشد. (پیشنهاد ما پایتون است).

۷- کدهای خود را به صورت مناسب کامنت گذاری کنید. به صورتی که بتوان حداقل روال اجرا و موارد مورد نیاز را درک کرد.

۸- سعی کنید ابتدا تمامی سوالات و بخش‌ها را مطالعه کنید.

۹- استفاده از هیچ کتابخانه آماده‌ای به جز موارد مطرح شده در تمرین مجاز نمی‌باشد و شما باید تمامی موارد را پیاده‌سازی کنید.

۱۰- در صورت هرگونه سوال یا مشکل می‌توانید با تدریس‌یار درس از طریق ایمیل در ارتباط باشید.

aliaz2197@gmail.com

بخش اول – معرفی دادگان

دادگان^۱ ارائه شده در این تمرین شامل ۳ بخش *train*, *validation*, *test* می‌باشد. این دادگان مربوط به نظرات ثبت شده توسط کاربران برای رستوران‌های مختلف در وبسایت گوگل است. هدف ما در این تمرین است تا با استفاده از اطلاعات موجود، به هر کاربر ۵ رستوران برتر که به سلیقه کاربر نزدیک‌تر هستند را پیشنهاد دهیم.

مجموعه دادگان

ویژگی	توضیحات
<i>business_id</i>	شماره‌ی یکتای هر رستوران
<i>user_id</i>	شماره‌ی یکتای هر کاربر
<i>rating</i>	امتیاز کاربر به رستوران
<i>review_text</i>	متن نظر ثبت شده توسط کاربر
<i>pics</i>	مسیر تصاویر ثبت شده توسط کاربر برای نظر
<i>history_reviews</i>	متن نظرات قبلی ثبت شده توسط کاربر برای رستوران‌های دیگر

بخش دوم – سامانه توصیه‌گر مبتنی بر شبکه عصبی (۳۰ امتیاز)

با استفاده از قسمت *train* دادگان، سامانه توصیه‌گر را صرفاً با استفاده از امتیاز داده شده توسط هر کاربر به رستوران ایجاد کنید. برای ساخت سامانه توصیه‌گر از مدل‌های مبتنی بر شبکه عصبی استفاده کرده و با بکارگیری بخش *validation* و *test* دادگان، نتایج را گزارش کنید.

*در انتخاب و طراحی نوع شبکه آزاد هستید و می‌توانید از شبکه‌هایی با ساختار ساده هم استفاده کنید ولی باید نحوه عملکرد مدل را به طور کامل در فایل گزارش توضیح دهید.

بخش سوم – سامانه توصیه‌گر با استفاده از اطلاعات جانبی (۴۰ امتیاز)

در این بخش سعی کنید با استفاده از اطلاعات جانبی موجود برای هر رستوران و هر کاربر، سامانه توصیه‌گری با استفاده از اطلاعات جانبی و امتیاز کاربر توسعه دهید.

^۱ Dataset

^۲ Side Information

*می‌توانید به سامانه بخش دوم اطلاعات جانبی را اضافه کنید و یا مدل جدیدی را پیاده‌سازی کنید.
*برای اطلاعات جانبی از متن نظرات موجود استفاده کنید.

بخش چهارم – کار با روش‌های ارزیابی و گزارش نتایج (۳۰ امتیاز)

هر دو روش مطرح شده در بالا را با استفاده از معیارهای ارزیابی $P@5$ ، $Recall@5$ و با استفاده از بخش *test* دادگان، ارزیابی و گزارش کنید.
در فایل گزارش روش‌های پیاده‌سازی شده برای سامانه‌های توصیه‌گر بخش دوم و سوم را به طور کامل توضیح داده و عملکرد آن‌ها را با یکدیگر مقایسه کنید.

بخش آخر - برخی نکات در مورد گزارش و تمرین

- دادگان مطرح شده در این تمرین و تمامی بخش‌ها همراه با صورت تمرین در سایت درس قرار داده شده است.
- در این تمرین شما مجاز به استفاده کتابخانه‌های زیر و موارد مشابه و هم‌کاربرد با آن‌ها می‌باشد:
`pytorch, tensorflow, numpy, scipy, pandas, genism, pickle, keras`
- در این تمرین سعی شده است علاوه بر آشنایی شما با کاربرد مباحث ارائه شده در کلاس و لمس بهتر آن، خلاقیت و حل چالش شما نیز ارزیابی شود. لذا در صورتی که در این تمرین چالشی وجود دارد که شما راه حلی برای آن ارائه دادید و استفاده کردید، آن را در گزارش بیان کنید. اما اگر مشکلی بزرگ وجود دارد که نیاز به بررسی مجدد دارد، آن را از طریق ایمیل با تدریس‌یاران درس مطرح کنید.
- در صورتی که هر گونه پیش‌پردازش بر روی دادگان انجام دادید آن را در گزارش خود بیان کنید.
- این تمرین ۱ نمره از بارم کلی شما از تمرینات را با توجه به پوشش مباحث و حجم تمرین دارد. امتیاز این تمرین از ۱۰۰ محاسبه می‌شود که بارم هر بخش مشخص شده است.
- در تمامی بخش‌ها، میزان نتایج شما در ارزیابی شما تاثیر چندانی ندارند (مگر اینکه بسیار دور باشد). بلکه میزان تسلط، دیدگاه و پیاده‌سازی، تحلیل‌ها و خلاقیت شماست که در نمره شما تاثیر مستقیم دارد و بر اساس این موارد مورد ارزیابی قرار می‌گیرد.

موفق باشد - عزیزی