

به نام خدا

تشخیص Frailty در سالمندان با استفاده از قابلیت‌های هوش مصنوعی

تهیه کننده: الهام شفیعی

اردیبهشت 1403

مقدمه

هدف این پروژه، استفاده از هوش مصنوعی برای تحلیل داده‌ها به منظور کمک به تشخیص frailty در سالمندان است. این پروژه مشخص می‌کند که چه ویژگی‌های فیزیولوژیکی و رفتاری می‌توانند به عنوان نشانگرهایی برای تشخیص frailty عمل کنند. هوش مصنوعی می‌تواند در توسعه سامانه‌های غربالگری frailty در سالمندان موثر باشد و برای ارتقای مراقبت‌های بهداشتی و کیفیت زندگی آنها به کار رود. جزئیات نیازمندی‌های پروژه در فایل FrailtyDetectionWithAI.pdf آمده است. تمامی مستندات و کدهای پروژه در مسیر زیر در دسترس هستند:

<https://github.com/elhamsha/health-frailty>

توصیف

این پروژه دو دیتاست ورودی دارد:

- اولین دیتاست با نام sensors_dataset.csv شامل اطلاعات تحرک و حضور سالمندان در محل زندگی آنهاست که با استفاده از حسگرهای موجود در محل گردآوری شده است. این دیتاست 4 فیلد و 58633 رکورد دارد. فیلدهای آن عبارتند از: شناسه کاربر، تاریخ ثبت ورود، زمان ثبت ورود، و اتاقی که سالمند وارد آن شده است. فرض می‌شود که اگر سالمند در تاریخ و ساعت مشخصی وارد اتاقی شده، تا ثبت رکورد بعدی همان روز، در آن اتاق باقی مانده است. توصیف بیشتر فیلدهای این دیتاست در جدول زیر آمده است:

Parameter	Description
part_id	The user ID, which should be a 4-digit number
ts_date	The recording date, which follows the "YYYYMMDD" format (e.g., 14 September 2017, is formatted as 20170914).
ts_time	The recording time, which follows the "hh:mm:ss" format.
room	The room which the person entered on the specific date and time. We assume that the person remained in the room till the next recording of the same day.

- دومین دیتاست با نام clinical_data.csv شامل اطلاعات بالینی سالمندان بوده و 55 فیلد و 540 رکورد دارد. فیلدهای آن شامل موارد زیر است:

Parameter	Description
part_id	The user ID
fried	Categorization by Fried
hospitalization_one_year	Number of hospitalizations in the last year
hospitalization_three_years	Number of hospitalizations in the last three years
ortho_hypotension	Orthostatic hypotension detection
vision	Vision
audition	Audition
weight_loss	Unintentional weight loss
exhaustion_score	Self-reported exhaustion
raise_chair_time	Lower limb strength
balance_single	Single foot station (Balance)
gait_get_up	Timed Get Up And Go Test
gait_speed_4m	Speed for 4 meters' straight walk
gait_optional_binary	Gait optional evaluation
gait_speed_slower	Slowed walking speed
grip_strength_abnormal	Grip strength outside the norms
low_physical_activity	Low physical activity
falls_one_year	Number of falls in the last year
fractures_three_years	Number of fractures during the last 3 years
fried_clinician	Fried's categorization according to clinician's estimation
bmi_score	Body Mass Index
bmi_body_fat	Body Fat (%)
waist	Waist circumference
lean_body_mass	Lean Body Mass
screening_score	Mini Nutritional Assessment (MNA) screening score
cognitive_total_score	Montreal Cognitive Assessment (MoCA) test score
memory_complain	Memory complain
mmse_total_score	Folstein Mini-Mental State Exam score
sleep	Reported sleeping problems
depression_total_score	15-item Geriatric Depression Scale (GDS-15)
anxiety_perception	Anxiety auto-evaluation
living_alone	Living Conditions
leisure_out	Leisure activities
leisure_club	Membership of a club
social_visits	Number of visits and social interactions per week
social_calls	Number of telephone calls exchanged per week
social_phone	Approximate time spent on phone per week
social_skype	Approximate time spent on videoconference per week
social_text	Number of written messages sent by the participant per week

house_suitable_participant	Subjective suitability of the housing environment according to participant's evaluation
house_suitable_professional	Subjective suitability of the housing environment according to investigator's evaluation
stairs_number	Number of steps to access house
life_quality	Quality of life self-rating
health_rate	Self-rated health status
health_rate_comparison	Self-assessed change since last year
pain_perception	Self-rated pain
activity_regular	Regular physical activity
smoking	Smoking
alcohol_units	Alcohol Use
katz_index	Katz Index of ADL
iadl_grade	Instrumental Activities of Daily Living
comorbidities_count	Number of comorbidities
comorbidities_significant_count	Number of comorbidities which affect significantly the person's functional status
medication_count	Number of medication

پارامتر مهم "fried" ، جمعیت سالمندان را به سه گروه زیر دسته‌بندی می‌کند:

- **Frail** : افراد سالمندی که آسیب‌پذیر هستند و خطر بروز یک رویداد مهم (مضر) در زندگی آنها افزایش یافته است.
- **Pre-frail** : افراد مسنی که به سمت frailty حرکت می‌کنند.
- **Non-frail** : سالمندانی که frailty ندارند.

این طبقه‌بندی توسط 5 پارامتر زیر اندازه‌گیری شده است. بنابراین می‌باید که این پارامترها در تحلیل مورد استفاده برای Classification یا دسته‌بندی frailty در نظر گرفته نشوند.

`grip_strength_abnormal, low_physical_activity, gait_speed_slower, weight_loss, exhaustion_score`

راه حل

فایل‌های کد این پروژه با عدد شماره‌گذاری شده و می‌باید به ترتیب اجرا شوند. توصیف عملکرد کدها در ادامه آمده است:

نام فایل	توصیف عملکرد کد
1_clinical_dataset_preprocess.ipynb	پیش‌پردازش و نرمال‌سازی داده‌های دیتاست بالینی و ذخیره آنها در فایل clinical_dataset_corrected.csv
2_clinical_dataset_classification.ipynb	اجرای الگوریتم‌های Classification بر دیتاست بالینی
3_sensors_dataset_preprocess.ipynb	پیش‌پردازش و نرمال‌سازی داده‌های دیتاست حسگرها و ذخیره آنها در فایل sensors_dataset_corrected.csv
4_percentage_in_rooms.ipynb	محاسبه درصد زمانی که هر فرد در اتاق‌های زیر سپری کرده: Bedroom, Bathroom, Livingroom, Kitchen و ذخیره آنها در percentage_in_rooms_dataset.csv
5_clustering_in_merged.ipynb	تلفیق دو دیتاست 'بالینی' و 'درصد زمانی حضور در اتاق‌ها' و اجرای الگوریتم Clustering

تحلیل اجرای Classification

دقت الگوریتم‌های مختلف Classification که بر دیتاست اطلاعات بالینی اجرا شده در جدول زیر آمده است:

Algorithm	Accuracy
Random Forest	0.63
Logistic Regression	0.58
Neural Network	0.56

با اجرای الگوریتم RandomForest ، پارامترهای مهم به ترتیب زیر محاسبه شده‌اند:

Parameter	Importance (↓)
gait_speed_4m	0.067
gait_get_up	0.056
cognitive_total_score	0.042
bmi_score	0.041
raise_chair_time	0.039
...	...
living_alone	0.003
gait_optional_binary	0.002

تحلیل اجرای Clustering

در این مرحله، ابتدا دو دیتاست 'بالینی' و 'درصد زمانی' تلفیق شدند و سپس الگوریتم Clustering اجرا شد. با اجرای الگوریتم KMeans ، مقدار ضریب Silhouette Score: 0.4 محاسبه شده است. با روش PCA مقدار این ضریب Silhouette Score: 0.6 گردید که بهبود یافته است.

ضریب Silhouette Coefficient

اگر Silhouette Coefficient نزدیک به 1 باشد، نشان‌دهنده یک خوشه‌بندی خوب است. ضریب نزدیک به -1، نشان‌دهنده خوشه‌بندی نادرست، و ضریب برابر با 0، نشان‌دهنده تداخل بین خوشه‌ها و نیاز به بهبود خوشه‌بندی است.

روش PCA (Principal Component Analysis)

PCA یا تحلیل مؤلفه‌های اصلی، یک روش رایج برای کاهش ابعاد داده‌ها و استخراج ویژگی‌های مهم است. این کاهش ابعاد باعث ساده‌تر شدن محاسبات و تجزیه و تحلیل داده‌ها می‌شود و می‌تواند بهبود قابل توجهی در عملکرد و دقت الگوریتم‌های یادگیری ماشین مانند خوشه‌بندی ایجاد کند.

تشخیص خوشه‌ها

پس از اجرای خوشه‌بندی، می‌باید تشخیص داد که هر یک از خوشه‌ها احتمالا به کدام یک از ۳ گروه (0) Non-frail, (1) Pre-frail, (2) Frail تعلق دارد. پس از اجرای الگوریتم، نتایج زیر حاصل شده و دیده می‌شود که احتمال تشخیص گروه Non-frail بیشتر از دو گروه دیگر است.

```
Cluster 0 is mostly comprised of: 1
```

```
Cluster 1 is mostly comprised of: 0
```

```
Cluster 2 is mostly comprised of: 1
```

```
Label distribution in Cluster 0:
```

```
fried
```

```
1    68
```

```
0    58
```

```
2    17
```

```
Name: count, dtype: int64
```

```
Label distribution in Cluster 1:
```

```
fried
```

```
0    21
```

```
1    10
```

```
2     6
```

```
Name: count, dtype: int64
```

```
Label distribution in Cluster 2:
```

```
fried
```

```
1    47
```

```
0    41
```

```
2    12
```

```
Name: count, dtype: int64
```