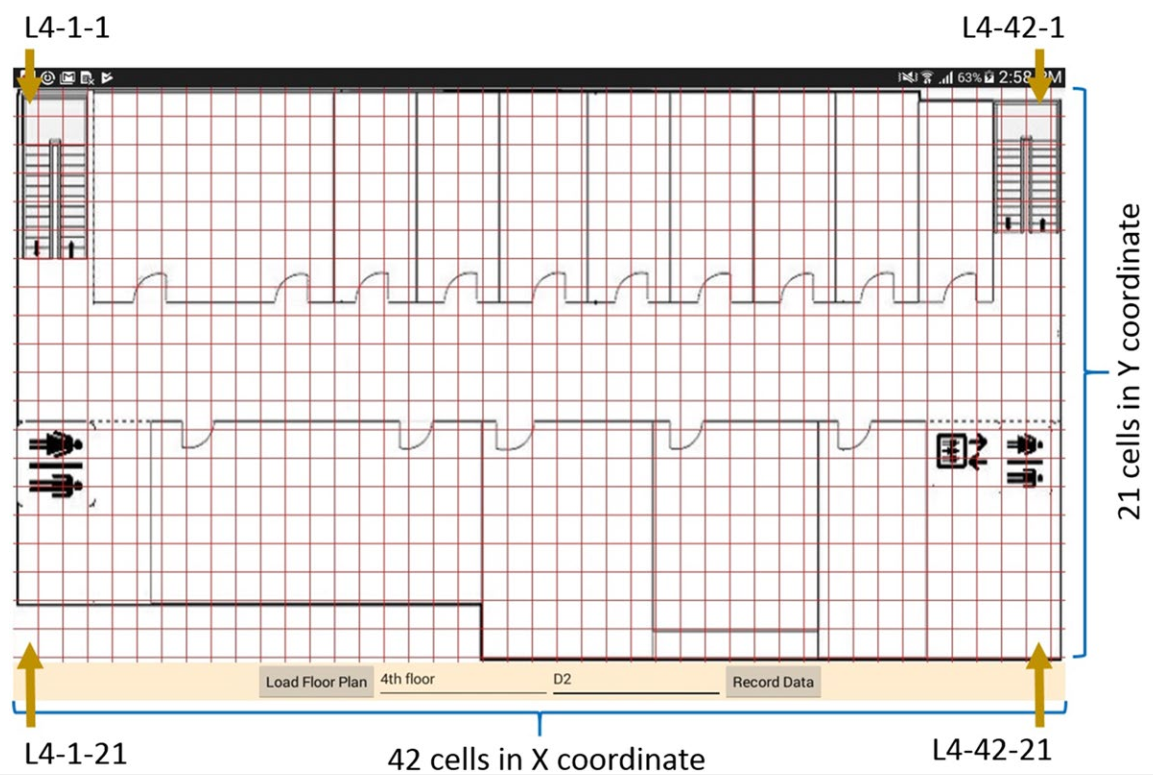
1. شبیه سازی و ارزیابی نتایج
   1. مقدمه

ابتدا مجموعه دادگان موقعیت‌یابی مبتنی بر RSS که اطلاعات اساسی برای ارزیابی الگوریتم‌ها و روش‌های موقعیت‌یابی می‌باشند، معرفی می‌شود. سپس، به بررسی نتایج شبیه‌سازی روش‌های آموزش مدل موقعیت‌یاب با استفاده از الگوریتم‌های یادگیری گروهی که راهکار اصلی در بهبود دقت موقعیت‌یابی هستند پرداخته می شود. پس از آن، پیاده‌سازی دو روش پیشنهادی برای بهبود مدل یادگیری گروهی؛ یعنی روش‌های کاهش ابعاد و الگوریتم ژنتیک بررسی می‌شود. روش کاهش ابعاد برای کاهش پیچیدگی محاسباتی و بهبود کارایی الگوریتم‌های موقعیت‌یاب ارزیابی می شود و روش‌های تنظیم ابرپارامترها با استفاده از الگوریتم ژنتیک نیز به عنوان یک ابزار قدرتمند برای بهینه‌سازی پارامترها در این الگوریتم‌ها، بررسی می‌شود. در نهایت، با جمع‌بندی نتایج به دست آمده از تمامی این روش‌ها، عملکرد و کارایی هر یک از آن‌ها در تعیین موقعیت به صورت کامل ارزیابی می گردد و با نتایج مشابه مقایسه می‌شود.

* 1. دادگان موقعیت‌یابی مبتنی بر RSS

تعداد زیادی از مجموعه دادگان موقعیت یابی مبتنی بر RSS وجود دارد. در میان آن ها، چندین مجموعه دادگان به سبب ویژگی‌ها و اطلاعات غنی‌تر، مجبوب تر هستند و بیشتر مورد استفاده قرار می‌گیرند. از جمله معروف‌ترین مجموعه دادگان، مجموعه دادگان UJIndoorLoc که در سال 2014 میلادی معرفی شد برای شبکه غیر بی سیم Wi-Fi معرفی شد [1]. گرچه این مجموعه دادگان به عنوان معیار سنجش الگوریتم‌های موقعیت‌یاب به کار می‌رود، اما شامل شرایط مختلف محیطی نمی‌شود. مجموعه دادگان IPIN در سال 2016 تا 2022 میلادی برای مسابقات ردیابی ربات‌ها معرفی شدند که عمده اطلاعات آن مربوط به سنسورهای ربات است که مبتنی بر موقعیت‌یابی آفلاین است [2].

دادگان استفاده شده در این پژوهش، مجموعه دادگان JUIndoorLoc است که در سال 2019 میلادی ارائه شده است [3] و به خاطر ضبط داده در شرایط مختلف محیطی، اعتبار صحت این مجموعه دادگان در طول زمان بیشتر است. داده ها از طبقه سوم، چهارم و پنجم ساختمانی در دانشگاه جوادپور گردآوری شده اند. در این ساختمان، هر طبقه 882 متر مربع به طول 42 متر و عرض 21 متر را پوشش می دهد و از اتاق های اساتید، کلاس های درس، اتاق های سمینار، آزمایشگاه های تحقیقاتی و راهرو تشکیل شده است. این ناحیه به سلول‌های 1 متر × 1 متر تقسیم می‌شود. بنابراین، مختصات و هر طبقه به ترتیب به 42 و 21 سلول تقسیم می شود که در شکل 2 نشان داده شده است.



نمایشی از تقسیم‌بندی اثر انگشت طبقه چهارم محیط مجموعه دادگان استفاده شده [3]

گردآوری این مجموعه دادگان، با استفاده از نرم افزار WDC که بر روی دستگاه‌های اندرویدی نصب می‌شود، انجام شده است و برای هر بلوک از موقعیت مقادیر RSSI به مدت 120 ثانیه ضبط می‌شود و در پایگاه داده سیستم موقعیت یاب ذخیره می شود. با توجه به مشاهدات، مقادیر RSSI با سخت‌افزارهای مختلف و شرایط مختلف محیطی متفاوت است. لذا این مجموه دادگان برای استحکام الگوریتم‌ها در شرایط گوناگون جمع شده است. از جمله آن، جمع آوری داده از نقاط مختلف یک بلوک به جای جمع آوری از مرکز بلوک است. همچنین، داده‌ها در زمان‌های مختلف روز جمع‌آوری شده‌اند تا ماهیت‌های مختلف را به دلیل اثرات محیطی مختلف درک کنند. به علاوه در هنگام جمع آوری داده ها، ناهمگونی های زمینه ای مختلف مانند بسته یا باز بودن در اتاق ها و حضور و عدم حضور انسان ها در نظر گرفته شده است. برای درنظرگیری سخت افزارهای مختلف نیز از چهار دستگاه اندرویدی متفاوت استفاده شده است.

مجموعه دادگان شامل 25364 نمونه است و هر نمونه دارای 177 ویژگی است که شامل 172 مقدار RSSI از نقاط دسترسی مختلف ، شناسه منحصر به فرد هر بلوک ، باز و بسته بودن درب اتاق، حضور یا عدم حضور انسان، شناسه دستگاه ثبت کننده و زمان جمع آوری داده بر حسب میلی ثانیه است.

* + 1. پیش پردازش دادگان موقعیت‌یابی

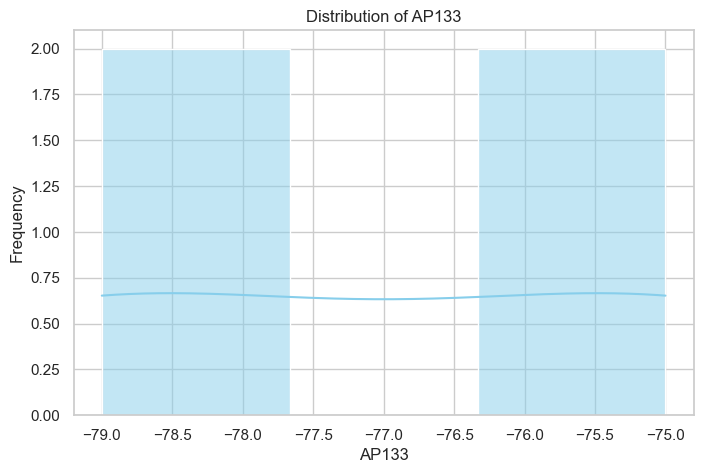
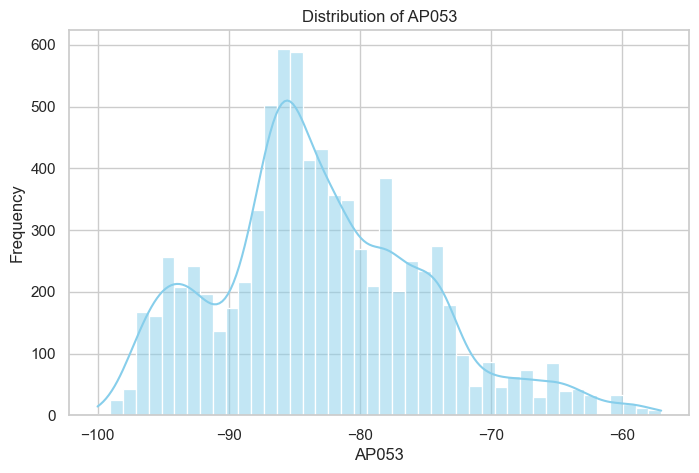
پیش پردازش دادگان، مجموعه اقداماتی است که پیش از استفاده از دادگان موقعیت یابی اجرا می ‌شود. این اقدامات در راستای پاکسازی و تقویت داده‌ها صورت می‌گیرد تا دقت و قابلیت استفاده از آن‌ها در مراحل بعدی افزایش یابد. اقدامات انجام شده به شرح زیر است:

* پاک‌سازی داده

این مرحله شامل حذف مقادیر تکراری و پاکسازی مقادیر گمشده است. با توجه به نتایج برآمده از پردازش مجموعه دادگان، مقادیر تکراری در آن وجود ندارد. همچنین در هر نمونه، مقادیر RSSI از نقاط دسترسی خارج از محدوده برابر با جای‌گذاری شده است.

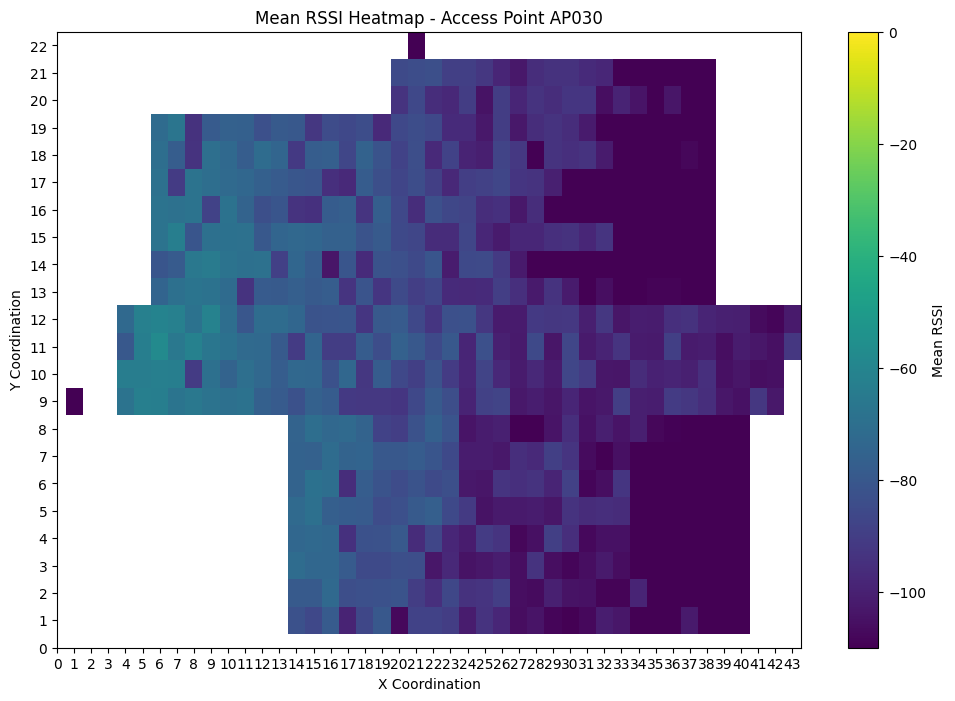
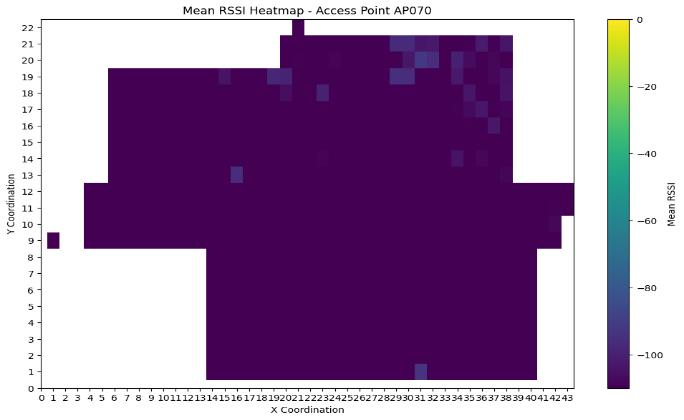
* *تجزیه و تحلیل اکتشافی دادگان[[1]](#footnote-2) (*EDA*)*

این مرحله از پیش پردازش برای بررسی توزیع هر ویژگی انجام می‌شود. به عنوان مثال، با توجه به ‏شکل (4˗2) که تعداد دفعات تکرار هر مقدار RSSI را برای هر نقطه دسترسی نشان می دهد، نقطه دسترسی AP053 نسبت به نقطه دسترسی AP133، از اطلاعات بیشتری برخوردار است. این تجزیه و تحلیل، ایده استفاده از دو الگوریتم پیشنهادی در جهت بهبود عملکرد مدل های یادگیری گروهی را تأیید می‌کند.

نمودار هیستوگرام تعداد دفعات تکرار هر مقدار RSSI در نقاط دسترسی AP053 و AP133

علاوه بر آن، با میانگین گیری از مقادیر RSSI در هر بلوک، نقشه ای از توزیع مقادیر RSSI در مختصات و حاصل می‌گردد که میزان اطلاعات دریافتی از هر نقطه دسترسی را به طور شهوری مصور می‌سازد. نمونه ای از این این نقشه در ‏شکل (4˗3) نشان داده شده است.

نقشه توزیع مقادیر RSSI در مختصات و از نقاط دسترسی AP030 و AP070 در طبقه چهارم

* نرمال سازی

نرمال‌سازی داده‌ها یک فرآیند مهم در پیش‌پردازش داده‌ها است که به مقیاس‌بندی داده‌ها به محدوده‌ای مشخص و یکنواخت اشاره دارد. در این فرآیند مقادیر هر ویژگی از مجموعه دادگان به یک بازه معین (معمولاً بین 0 و 1)، با هدف داشتن مقیاس مشابه برای همه ویژگی‌ها تبدیل می‌شود.

نرمال سازی باعث افزایش سرعت و تعمیم پذیری مدل می‌شود. همچنین هنگامی که مقادیر ویژگی‌ها دارای مقیاس‌های متفاوت هستند، الگوریتم‌های یادگیری گروهی ممکن است به ویژگی‌هایی با مقیاس بزرگتر بیش‌برازش کنند. با نرمال‌سازی داده‌ها، این احتمال کاهش می‌یابد و عملکرد مدل در داده‌های جدید بهبود می‌یابد.

* 1. آموزش موقعیت‌یاب با الگوریتم‌های یادگیری گروهی

مدل های یادگیری گروهی معرفی شده در فصل 3 با مجموعه دادگان معرفی شده در بخش 4-2 آموزش داده می شود. برای مقایسه عملکرد مدل ها در شرایط یکسان، می بایست تمام پردازش ها در یک سیستم سخت‌افزاری واحد انجام شود. مشخصات این سیستم سخت افزاری در ‏جدول (4˗1) ذکر شده است. مطابق آن، سیستم سخت افزاری مورد استفاده، یک سیستم پایین‌رده برای محاسبات موقعیت محسوب می‌شود.

همچنین در پیاده سازی الگوریتم های یادگیری گروهی، از به کارگیری الگوریتم‌های جاسازی شده برای توقف

مشخصات سیستم سخت افزاری موقعیت‌یاب برای اجرای شبیه‌سازی

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CPU** | **Memory** | **Disk** | **OS** | **GPU** |
| Intel Core i5-4670 @3.4GHz | 16.0 GB DDR3 | HDD | Windows 11 | × |

زودهنگام الگوریتم در روش های تکرار شونده ممانعت به عمل آمده است تا مدل‌های خروجی قابل مقایسه باشد. برای هر الگوریتم یادگیرنده از درخت با حداکثر عمق به عنوان یادگیرنده پایه استفاده شده است.

برای ارزیابی بهتر، از روش اعتبارسنجی متقابل طبقه‌بندی شده 5 گانه[[2]](#footnote-3) استفاده می‌شود. در این روش 20 درصد از دادگان به عنوان دادگان اعتبارسنجی، با درنظرگیری نسبت برچسب ها انتخاب می‌گردد و مابقی به عنوان آموزش به مدل داده می‌شود. در نهایت پس از 5 تکرار آموزش مدل، از دقت تخمین دادگان اعتبارسنجی میانگین گیری می‌شود.

در ابتدا، یک درخت تصمیم به عنوان یادگیرنده پایه با چهار مقدار حداکثر عمق متفاوت آموزش داده شده است و نتایج ارزیابی آن برای 20 درصد از مجموعه دادگان در ‏جدول (4˗2) عنوان شده است. اشاره شد که ابعاد هر بلوک 1 متر در 1 متر است لذا در محاسبات میانگین خطا مطابق رابطه می توان به سادگی و مستقیم مختصات و را جای‌گذاری کرد. همچنین زمان آزمایش برای 5072 داده محاسبه شده است.

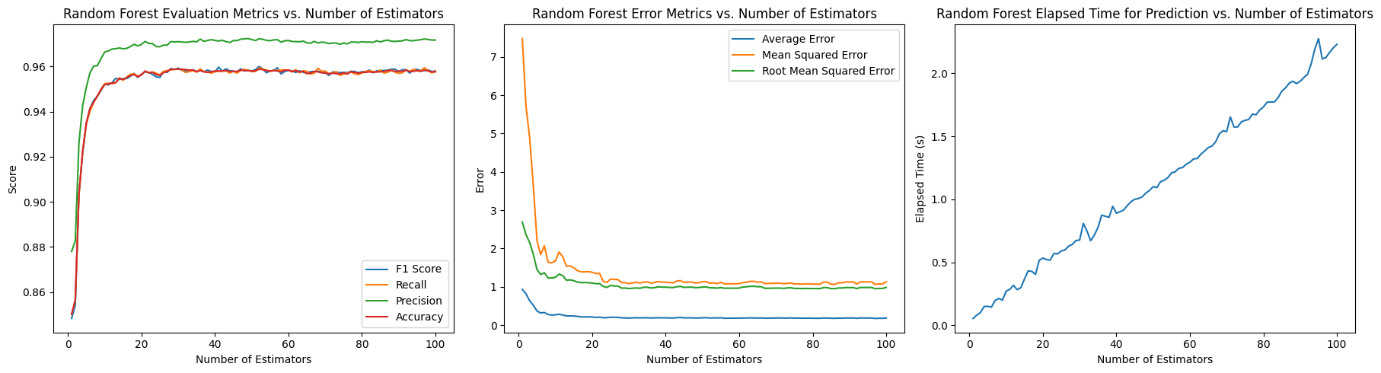
نتایج شبیه‌سازی درخت تصمیم با دادگان ارزیابی و حداکثر عمق‌های متفاوت

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| حداکثر عمق  درخت تصمیم |  |  |  |  | (متر) | (متر مربع) | (متر) | زمان آزمایش |
| 10 | 8.21% | 12.21% | 8.22% | 7.97% | 15.58 | 362.46 | 17.98 | 0.054 ثانیه |
| 50 | 82.50% | 85.73% | 82.49% | 83.10% | 1.06 | 9.85 | 3.11 | 0.056 ثانیه |
| 100 | 85.02% | 87.94% | 85.27% | 84.85% | 0.93 | 7.48 | 2.69 | 0.057 ثانیه |
| 176 | 85.02% | 88.80% | 85.01% | 84.84% | 0.093 | 7.47 | 2.69 | 0.058 ثانیه |

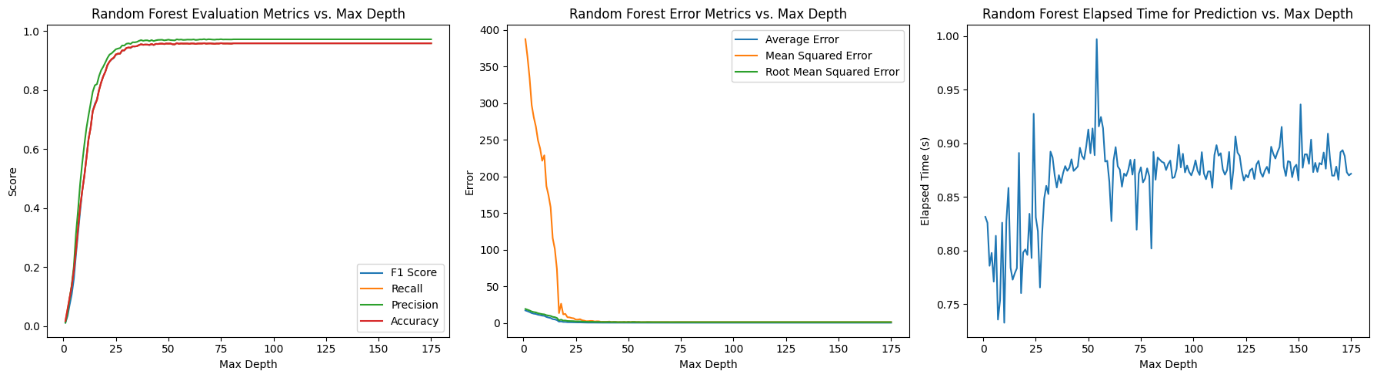
با توجه به نتایج حاصل از یادگیرنده پایه، مدل‌های یادگیری گروهی مورد ارزیابی قرار می‌گیرد.

* + 1. مدل موقعیت‌یاب الگوریتم جنگل تصادفی

همانطور که در مطرح شد، الگوریتم جنگل تصادفی به سه پارامتر )تعداد ویژگی‌های هر گره(، (تعداد درختان در جنگل) و اندازه درخت نیاز دارد. برابر با نیمی از تعداد ویژگی‌های کل (176 ویژگی) درنظر گرفته شده است. اندازه درخت برابر با حداکثر عمق درخت تعیین شده است که در این بخش مقدار آن محدود نشده است. نتایج شبیه‌سازی با تعداد متفاوت درختان () در ‏شکل (4˗4) نمایش داده شده است. با توجه به آن، با افزایش تعداد درختان در جنگل، زمان آزمایش افزایش می‌یابد. همچنین مشاهده می‌شود که خطا و معیارهای سنجش، پس از تعداد مشخصی از درختان، با افزایش آن تغییر چندانی نمی‌کند. در ‏جدول (4˗3) نیز مقدار دقیق متغیرها برحسب سه مقدار از تعداد درختان ذکر شده است.



نمودار نتیجه ارزیابی، خطا و زمان آزمایش جنگل تصادفی برای تعداد متفاوت درختان و حداکثر عمق ثابت



نمودار نتیجه ارزیابی، خطا و زمان آزمایش جنگل تصادفی برای عمق متفاوت درختان و تعداد 40 درخت

برای بررسی تأثیر عمق درختان بر معیارهای ارزیابی و زمان، تعداد درختان در مقدار 40 نگه داشته شده است و عمق درختان از صفر تا حداکثر مقدار تغییر داده شده است. نتایج ارزیابی در ‏شکل (4˗5) نمایش داده شده است. مطابق آن پس از یک عمق مشخص، دقت و خطا تغییر چندانی ندارد. دلیل این مسئله برآورده شدن شرایط توقف در ساخت درخت پس از تعداد معینی از گره ها است.این موضوع در نمودار زمان آزمایش برحسب حداکثر عمق و ‏جدول (4˗4) نیز تأیید می‌شود.

نتایج شبیه‌سازی جنگل تصادفی و تعداد متفاوت درختان و عمق ثابت درختان (حداکثر)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| تعداد درختان |  |  |  |  | (متر) | (متر مربع) | (متر) | زمان آزمایش |
| 10 | 85.38% | 88.26% | 85.66% | 89.17% | 0.91 | 8.43 | 2.38 | 0.082 ثانیه |
| 50 | 95.80% | 97.19% | 95.80% | 95.84% | 0.19 | 1.13 | 0.99 | 1.17 ثانیه |
| 100 | 95.76% | 97.17% | 95.76% | 95.81% | 0.61 | 15.26 | 3.90 | 3.00 ثانیه |

نتایج شبیه‌سازی جنگل تصادفی برای عمق متفاوت درختان و تعداد 40 درخت

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| حداکثر عمق درختان |  |  |  |  | (متر) | (متر مربع) | (متر) | زمان آزمایش |
| 10 | 50.99% | 61.46% | 60.00% | 50.70% | 9.58 | 228.90 | 11.59 | 0.80 ثانیه |
| 50 | 95.68% | 97.04% | 95.63% | 95.69% | 0.21 | 1.43 | 1.12 | 0.86 ثانیه |
| 100 | 95.68% | 97.04% | 95.63% | 95.69% | 0.21 | 1.43 | 1.12 | 0.87 ثانیه |

با درنظر گرفتن دو پارامتر آزاد از نظر محاسباتی، امکان ساخت جنگل تصادفی برای تمام حالات ممکن بسیار گران است. پیدا کردن مقدار بهینه این پارامتر در بخش با استفاده از الگوریتم ژنتیک توضیح داده می‌شود.

* + 1. مدل موقعیت‌یاب الگوریتم AdaBoost

در الگوریتم 2 که برای روش AdaBoost ارائه شده است، تنها از دو پارامتر (فرضیه یا یادگیرنده پایه) و (تعداد یادگیرندگان پایه) به عنوان ابرپارامتر تنظیم مدل یاد شده است. از آن جا که در این پژوهش از درخت تصمیم به عنوان یادگیرنده پایه استفاده شده است. پارامتر ، به (حداکثر عمق درخت) اشاره دارد. در این بخش همچنان نتایج ارزیابی در دو حالت تعداد یادگیرندگان متغیر/ثابت و حداکثر عمق متفاوت ثابت/متغیر

بحث می‌گردد. در حالت اول، حداکثر عمق درخت برابر با بیشینه مقدار ممکن آن؛ یعنی 176 گره، درنظر گرفته می شود و تعداد یادگیرندگان از صفر تا 100 تغییر می‌کند.

* 1. کاهش ابعاد دادگان موقعیت‌یابی

متن

* 1. تنظیم ابرپارامترها با استفاده از الگوریتم ژنتیک

متن

* 1. مقایسه و جمع‌بندی نتایج

متن

[1] J. Torres-Sospedra *et al.*, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *2014 international conference on indoor positioning and indoor navigation (IPIN)*, 2014, pp. 261-270: IEEE.

[2] J. Torres-Sospedra *et al.*, "The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work," *Sensors,* vol. 17, no. 3, p. 557, 2017.

[3] P. Roy, C. Chowdhury, D. Ghosh, and S. Bandyopadhyay, "JUIndoorLoc: A ubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity," *Wireless Personal Communications,* vol. 106, pp. 739-762, 2019.

1. Exploratory Data Analysis [↑](#footnote-ref-2)
2. Stratified 5-Fold Cross Validation [↑](#footnote-ref-3)