



دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پروپوزال پایان نامه کارشناسی

گرایش معماری سیستم‌های کامپیوتری

پیاده‌سازی موازی تشخیص ناهنجاری در فعالیت‌های خانه‌های هوشمند

نگارش

فاطمه هاشمی چالشتی

استاد راهنما

دکتر مریم امیرحائری

بهمن ۱۳۹۷

## ۱ مقدمه

تجزیه و تحلیل داده‌ها موضوع قابل توجهی است که استفاده گسترده آن در شبکه‌های اینترنت اشیا (IoT<sup>۱</sup>) اهمیت آن را دوچندان کرده است. یکی از جنبه‌های تجزیه و تحلیل داده‌ها تشخیص ناهنجاری<sup>۲</sup> در آن‌هاست که در برخی موارد لازم است بی‌درنگ<sup>۳</sup> باشد و برای مثال در نگهداری پیشگویانه، نظارت بر آب و هوا، تأمین امنیت و یا پزشکی قانونی سایبری کمک شایانی می‌کند. با وجود این که برنامه‌های تحت وب<sup>۴</sup> بسیاری وجود دارند که کاربران بوسیله آن‌ها می‌توانند بر اساس محور زمان داده‌ها را تصویرسازی کنند و روی آن تجزیه و تحلیل آماری انجام دهند ولی در این بین معمولاً تشخیص ناهنجاری صورت نمی‌گیرد و یا بسیار ساده، غیرقابل اعتماد و غیردقیق است [۱].

### ۱.۱ تشخیص ناهنجاری

گسل، شکست، خطا و ... کلمات مورد استفاده برای ارجاع به یک سیستم در یک پیکربندی ناخواسته هستند. به جای دسته‌بندی این کلمات در یک سلسله مراتب، مفهوم سطح بالاتر ناهنجاری مورد استفاده قرار می‌گیرد. یک ناهنجاری وضعیت غیرطبیعی یا غیر منتظره و یا رفتار غیرعادی بخشی از سیستم است.

پیدا کردن الگوهای پرت یا ناهنجار در مسائل بسیاری مثل تشخیص تقلب، تشخیص نفوذ، تشخیص تخلف کاربرد دارد. زمانی که داده‌ها زیاد هستند و نمی‌توان به راحتی داده‌های پرت را در میان آن‌ها شناسایی نمود، تکنیک‌های داده‌کاوی که برای تشخیص الگوهای پرت به کار می‌روند، می‌توانند در این راه مفید واقع شوند. ناهنجاری‌ها نمونه‌هایی از یک کلاس فعالیت هستند که از توالی معمول یا مورد انتظار رویدادها عملکرد متفاوتی دارند. در این پروژه قصد داریم ناهنجاری‌های فعالیت‌های انجام شده در یک خانه هوشمند را شناسایی کنیم. به منظور تشخیص ناهنجاری، اطلاعات تعدادی از فعالیت‌ها و مدت زمان درگیر در انجام یک نمونه فعالیت مورد بهره‌برداری قرار می‌گیرد.

---

<sup>۱</sup> Internet of Things

<sup>۲</sup> Anomaly Detection

<sup>۳</sup> Real-time

<sup>۴</sup> Web Application

بسیاری از الگوریتم‌های تشخیص ناهنجاری زمان اجرای به نسبت بالایی دارند. به طور مثال الگوریتم ضریب دورافتادگی محلی<sup>۱</sup> یا LOF یک الگوریتم بدون نظارت برای تشخیص ناهنجاری و داده‌های پرت است که با اندازه‌گیری انحراف محلی یک نقطه داده با توجه به همسایه‌های آن داده‌های پرت را شناسایی می‌کند. ایده اساسی این الگوریتم مقایسه چگالی محلی یک نقطه با چگالی همسایه‌های آن است. محلی بودن بر اساس  $k$  نزدیک‌ترین همسایگی تعیین می‌شود که فاصله آن‌ها برای تخمین چگالی مورد استفاده قرار می‌گیرد. با مقایسه چگالی محلی یک شیء با چگالی‌های همسایه‌های آن می‌توان نواحی دارای چگالی مشابه و نقاطی که اساساً چگالی کمتری نسبت به همسایه‌های خود دارند را تعیین کرد. این موارد به عنوان دورافتادگی یا داده پرت در نظر گرفته می‌شوند. چگالی محلی به وسیله فاصله معمولی که یک نقطه داده توسط همسایه‌های خود «دسترسی‌پذیر» است تخمین زده می‌شود [۲].

## ۱.۲ خانه هوشمند

خانه هوشمند به عنوان یک برنامه مدرن به جهت مدیریت هوشمند آسایش، مراقبت‌های پزشکی، ایمنی، امنیت، و مصرف بهینه انرژی شناخته می‌شود. در طول دهه گذشته، پیشرفت‌های وسیع دستگاه‌های هوشمند جدیدی که می‌توانند به اینترنت متصل شوند و با استفاده از برنامه‌های کاربردی از راه دور کنترل شوند، افزایش یافته است. این شبکه از دستگاه‌ها و موارد دیگر تعبیه شده با حسگر، الکترونیک، نرم افزار و ... اینترنت اشیاء نامیده می‌شود و در مجموع می‌تواند به پدید آمدن مفهومی همچون خانه هوشمند منجر شود [۳]. بر اساس یک پایگاه داده موضوعی، این سیستم فعالیت‌های زندگی روزمره ساکنین را بررسی می‌کند. از این رو، رفتار غیرطبیعی یا مشکلات مزمن می‌تواند توسط سیستم شناسایی شود. با در نظر گرفتن پیچیدگی این سیستم‌ها، ممکن است انواع مختلف ناهنجاری‌ها رخ دهد و عملکرد برنامه‌های ضروری را مختل سازد. با توجه به گستره وسیع برنامه‌ها، ناهنجاری‌ها در خانه‌های هوشمند می‌توانند زندگی ساکنین را در معرض خطر قرار دهند. به عنوان مثال، اگر حسگرهایی که توسط ساکنین پوشیده شده‌اند، اطلاعات غلطی ارسال کنند، مراقبین یا پرستاران نمی‌توانند بیماری‌های مزمن را شناسایی کنند. به همین ترتیب، مثلاً در صورتی که حسگر تشخیص سقوط انسان، پس از مدت زمان طولانی زمین خوردن یکی از ساکنین خانه را گزارش کند مأموران اورژانس ممکن است به موقع در جریان مشکل قرار نگیرند و این امر باعث ایجاد استرس بیش از حد یا افزایش رنج

---

<sup>۱</sup> Local Outlier Factor

ساکنین شود[۵]. از دیگر موارد ناهنجاری می‌توان به دزدی یا میزان دما، رطوبت و یا دود نامتعارف اشاره کرد. از این رو، سیستم مدیریت ناهنجاری یک بخش ضروری از معماری کلی یک خانه هوشمند است.

## ۲ پیاده‌سازی موازی الگوریتم

با اینکه یافتن داده پرت براساس داده‌های همسایه نتیجه خوبی به همراه دارد، محاسبه مقدار LOF برای تک تک داده‌ها نیازمند بدست آوردن  $k$  نزدیک‌ترین همسایه است که این سربار محاسباتی می‌تواند موارد استفاده از الگوریتم LOF را محدود کند و عملاً امکان استفاده از آن را برای داده‌های با حجم بالا غیر ممکن می‌کند[۶]. در این پروژه قصد داریم با پیاده‌سازی موازی و توزیع شده الگوریتم روی خوشه‌ای از بردهای رزبری پای<sup>۱</sup> زمان اجرا را کاهش دهیم. رزبری پای یک رایانه تک-برد<sup>۲</sup> در اندازه یک کارت اعتباری است که بنیاد رزبری پای آن را ساخته است تا آموزش علوم رایانه را در مدرسه‌ها تشویق کند. این رایانه دارای دیسک سخت نیست ولی به کمک حافظه SD، خود را راه‌اندازی می‌کند و قادر است با سیستم‌عامل‌های مختلفی راه‌اندازی شود. رزبری پای تعداد زیادی سوکت (GPIO<sup>۳</sup>) دارد که امکان وصل شدن آن به سخت‌افزارها و قطعات دیگر را فراهم می‌آورد. همچنین ارزان بودن برد رزبری پای از دیگر مزایای مهم آن است.

با داشتن خوشه‌ای از پردازنده‌ها قدرت و سرعت پردازش بالا می‌رود و در این پروژه هدف این است که بار پردازشی روی خوشه‌ای از بردهای رزبری پای توزیع شود که به این منظور از زبان برنامه‌نویسی MPI<sup>۴</sup> استفاده می‌کنیم. MPI یک سیستم پیام‌رسانی استاندارد و قابل حمل<sup>۵</sup> است که نحو<sup>۶</sup> و معنا<sup>۷</sup> را در هسته روال‌های<sup>۸</sup> کتابخانه برای طیف وسیعی از کاربران که برنامه‌های ارسال پیام قابل حمل را در زبان‌های C، C++ و Fortran می‌نویسند، تعریف می‌کند[۴]. رزبری پای با پردازنده چهار هسته‌ای خود، به عنوان یک پلتفرم کامل برای پردازش حافظه مشترک<sup>۹</sup> عمل می‌کند. یکی از بردهای خوشه به عنوان گره اصلی<sup>۱</sup> انتخاب می‌شود و دیگر بردها

---

<sup>۱</sup> Raspberry Pi

<sup>۲</sup> Single-board computer

<sup>۳</sup> General Purpose Input Output

<sup>۴</sup> Message Passing Interface

<sup>۵</sup> Portable

<sup>۶</sup> Syntax

<sup>۷</sup> Semantics

<sup>۸</sup> Routine

<sup>۹</sup> Shared Memory Processing

بردها گره‌های فرعی<sup>۲</sup> خواهند بود. باتوجه به تعداد گره‌ها داده‌های مسئله به چند بخش تقسیم می‌شوند. گره اصلی وظیفه تقسیم داده‌ها و ارسال به سایر گره‌ها را دارد و گره‌های فرعی پس انجام محاسبات محلی نتیجه را به گره اصلی باز می‌گردانند.

### ۳ منابع

[1] Mohamudally, Nawaz & Peermamode-Mohaboob, Mahejabeen. (2018). Building An Anomaly Detection Engine (ADE) For IoT Smart Applications. *Procedia Computer Science*. 134. 10-17. 10.1016/j.procs.2018.07.138.

[2] M. Breunig, Markus & Kriegel, Hans-Peter & Ng, Raymond & Sander, Joerg. (2000). LOF: Identifying Density-Based Local Outliers. *ACM Sigmod Record*. 29. 93-104. 10.1145/342009.335388.

[3] Mocrii, Dragos & Chen, Yuxiang & Musilek, Petr. (2018). IoT-based smart homes: A review of system architecture, software, communications, privacy and security. *Internet of Things*. 1-2. 81-98. 10.1016/j.iot.2018.08.009.

[4] Doucet, Kevin & Zhang, Jian. (2017). Learning Cluster Computing by Creating a Raspberry Pi Cluster. *ACM SE '17 Proceedings of the SouthEast Conference Pages* 191-194. 10.1145/3077286.3077324.

[5] Pardo, Etienne & Espes, David & Le-Parc, Philippe. (2016). A Framework for Anomaly Diagnosis in Smart Homes Based on Ontology. *Procedia Computer Science*. 83. 545-552. 10.1016/j.procs.2016.04.255.

[۶] ابیضی هدی؛ سید امیر اصغری؛ قربان خردمندیان و محمد رضا بینش مروستی، ارایه یک روش جدید در شناسایی الگوهای پرت با استفاده از روش‌های بدون نظارت و الگوریتم ژنتیک، نهمین کنفرانس ملی فرماندهی و کنترل ایران، تهران، دانشگاه خوارزمی - انجمن علمی فرماندهی و کنترل ایران، ۱۳۹۵.

---

<sup>۱</sup> Master

<sup>۲</sup> Slave