

فهرست مطالب

۱-۱- مقدمه	۱۱
۲-۱- بیان مسئله	۱۵
۳-۱- ضرورت و اهمیت مسئله	۱۷
۴-۱- فرضیات و محدودیت‌های تحقیق	۱۷
۵-۱- نوآوری‌های راه‌حل پیشنهادی	۱۸
۶-۱- مقالات استخراج‌شده از پژوهش	۲۲
۷-۱- ساختار رساله	۲۲
۲- فصل دوم: مرور ادبیات	۲۳
۱-۲- مقدمه	۲۴
۲-۲- تلفیق اطلاعات	۲۴
۳-۲- بررسی مدل‌های تلفیق داده	۲۵
۴-۲- پیش‌پردازش داده‌ها	۳۵
۵-۲- چالش‌های موجود	۴۴
۶-۲- معرفی ابزارهای داده‌های حجیم	۴۵
۱-۶-۲- ابزارهای جریان‌ی داده‌های حجیم	۴۵
۲-۶-۲- ابزارهای غیرجریان‌ی داده‌های حجیم	۴۸
۳-۶-۲- ابزارهای پیش‌پردازش داده‌های حجیم	۵۳
۴-۶-۲- طبقه‌بندی ابزارهای داده‌های حجیم	۵۴

۵۴	۷-۲-خلاصه فصل
۵۶	۳- فصل سوم: روش پیشنهادی
۵۷	۳-۱- مقدمه
۵۷	۳-۲- ساختار کلی روش پیشنهادی
۵۹	۳-۳- فاز اول: راه‌اندازی
۵۹	3-3-1- روال پیش‌پردازش تعریف‌شده توسط کاربر
۶۷	۳-۳-۲- فاز دوم: اجرایی
۸۰	۳-۳-۳- ارائه یک مثال توصیفی
۹۰	۳-۳-۱- فاز سوم: رصد کردن
۱۰۰	۳-۴- خلاصه فصل
۱۰۱	۴- فصل چهارم: ارزیابی روش پیشنهادی
۱۰۲	۴-۱- مقدمه
۱۰۲	۴-۲- سناریوی اول: الگوریتم خوشه‌بندی بهینه‌شده با استفاده از الگوریتم کلونی زنبور عسل
۱۰۳	۴-۲-۱- مجموعه داده
۱۰۴	۴-۲-۲- ارزیابی مدت زمان اجرا
۱۰۶	۴-۲-۳- ارزیابی کیفی
۱۰۷	۴-۲-۴- تحلیل نتایج
۱۰۹	۴-۳- سناریوی دوم: پیش‌بینی بار ترافیکی در مناطق شهری
۱۱۲	۴-۳-۲- پیاده‌سازی با استفاده از چارچوب پیشنهادی
۱۱۴	۴-۳-۳- ارزیابی مدت زمان اجرا

۱۱۵	۴-۳-۴- تحلیل نتایج
۱۱۶	۴-۴- خلاصه فصل
۱۱۷	۱- فصل پنجم: نتیجه‌گیری و کارهای آتی
۱۱۸	۱-۱- مروری بر روش پیشنهادی
۱۱۸	۲-۱- چالش‌های پژوهش
۱۱۸	۱-۲-۱- محدودیت منابع پردازشی
۱۱۹	۲-۲-۱- شبیه‌سازی داده‌های جریان‌ی
۱۱۹	۳-۱- کارهای آتی
۱۱۹	۱-۳-۱- پشتیبانی از منابع داده‌ای و ابزارهای بیشتر حوزه داده‌های حجیم
۱۲۰	۲-۳-۱- استفاده از الگوریتم‌های تقریبی
۱۲۰	۳-۳-۱- استفاده از تئوری بازی‌ها
۱۲۱	۲- فصل ششم: منابع

فهرست جداول

جدول ۱-۱	تقسیم‌بندی داده‌ها برحسب اندازه	۱۱
جدول ۲-۱	مقایسه داده‌های سنتی و داده‌های حجیم	۱۲
جدول ۳-۱	جزئیات مسئله اصلی تحقیق	۱۶
جدول ۴-۱	فهرست نوآوریهای پژوهش	۱۹
جدول ۱-۲	مزایا و معایب چارچوب‌های پیش‌پردازش ارائه‌شده برای داده‌ها	۳۸
جدول ۲-۲	مقایسه روش‌های موجود	۴۰
جدول ۳-۲	مقایسه نوع عمل پیش‌پردازش، مجموعه‌داده و چارچوب در روش‌های پیش‌پردازش داده‌های حجیم	۴۲
جدول ۴-۲	طبقه‌بندی ابزارهای داده‌های حجیم	۵۵
جدول ۱-۳	دسته‌بندی عملگرها از نظر نوع وابستگی به اجرای کارهای قبل و بعد	۷۵
جدول ۲-۳	گراف معادل عملگرهای زبان ارائه شده برای تعریف روال پیش‌پردازش	۷۷
جدول ۳-۳	فهرست عدم تطبیق‌ها و سیاست مقابله با آن‌ها	۹۹
جدول ۱-۴	سخت‌افزار محیط شبیه‌سازی	۱۰۲
جدول ۲-۴	زمان اجرای الگوریتم خوشه‌بندی بر روی مجموعه داده Vehicle_sensIT برحسب ساعت بدون استفاده از چارچوب پیشنهادی	۱۰۵
جدول ۳-۴	زمان اجرای الگوریتم خوشه‌بندی بر روی مجموعه داده Vehicle_sensIT برحسب ساعت با استفاده از چارچوب پیشنهادی	۱۰۵
جدول ۴-۴	اطلاعات مجموعه داده citypulse	۱۱۰

فهرست شکل‌ها

- شکل ۱-۱ فرآیند استخراج دانش از داده ۱۴
- شکل ۱-۲ چرخه هوش ۲۶
- شکل ۲-۲ حلقه OODA ۲۷
- شکل ۳-۲ مدل تلفیق داده JDL ۲۷
- شکل ۴-۲ مدل آبخاری ۳۰
- شکل ۵-۲ مدل Dasarathy ۳۱
- شکل ۶-۲ مدل تلفیق داده بصری ۳۲
- شکل ۷-۲ مدل Omnibus ۳۳
- شکل ۸-۲ مدل دیدن Endsley ۳۴
- شکل ۹-۲ ساختار spark ۴۷
- شکل ۱۰-۲ معماری آپاچی کافکا ۴۸
- شکل ۱۱-۲ معماری پایگاه داده hbase ۵۰
- شکل ۱۲-۲ معماری hive ۵۲
- شکل ۱-۳ روال کلی چارچوب پیشنهادی ۵۸
- شکل ۲-۳ مثالی از روال پیش‌پردازش تعریف‌شده توسط کاربر ۷۰
- شکل ۳-۳ روال افراز گراف کارها ۷۳
- شکل ۴-۳ افراز گراف جهت‌دار کارها ۷۸
- شکل ۵-۳ افراز گراف کارها ۷۸
- شکل ۶-۳ فاز نگاشت الگوریتم زمانبند ۷۹
- شکل ۷-۳ فاز کاهش الگوریتم زمانبند ۸۰
- شکل ۸-۳ روال کلی الگوریتم کلونی زنبور عسل ۸۳
- شکل ۹-۳ فاز نگاشت الگوریتم خوشه‌بندی FCM بهینه شده با الگوریتم کلونی زنبور عسل ۸۶
- شکل ۱۰-۳ فاز کاهش الگوریتم خوشه‌بندی FCM بهینه شده با الگوریتم کلونی زنبور عسل ۸۷
- شکل ۱۱-۳ روال پیش‌پردازش تعریف شده برای الگوریتم خوشه‌بندی ارائه شده ۸۸
- شکل ۱۲-۳ گراف کار معادل روال پیش‌پردازش تعریف شده ۸۹
- شکل ۱۳-۳ افراز گراف کارها ۹۱
- شکل ۱۴-۳ نمونه‌ای از انتشار خطا در فرآیند پیش‌پردازش ۹۳
- شکل ۱۵-۳ خروجی مؤلفه نگاره‌سازی تا انتهای افراز اول در شکل ۳-۱۳ ۹۷
- شکل ۱۶-۳ گراف ایجاد شده از خروجی مؤلفه نگاره‌سازی ۹۷
- شکل ۱۷-۳ خروجی مؤلفه نگاره‌سازی که عدم تطابق در اجرای روال کارها را نشان می‌دهد ۹۸

شکل ۴-۱ مقایسه زمان اجرای الگوریتم خوشه‌بندی بر روی مجموعه داده Vehicle_sensIT در دو حالت با استفاده از چارچوب پیشنهادی و بدون استفاده از چارچوب پیشنهادی ۱۰۶

شکل ۴-۲ مقایسه زمان اجرای الگوریتم خوشه‌بندی بر روی مجموعه داده Caltech101 در دو حالت با استفاده از چارچوب پیشنهادی و بدون استفاده از چارچوب پیشنهادی ۱۰۷

شکل ۴-۳ مقایسه زمان اجرای الگوریتم خوشه‌بندی بر روی مجموعه داده MSRC-v1 در دو حالت با استفاده از چارچوب پیشنهادی و بدون استفاده از چارچوب پیشنهادی ۱۰۸

شکل ۴-۴ مقایسه کیفیت خوشه‌بندی روش ارائه شده در مقایسه با روش‌های دیگر به تفکیک کلاس‌های مختلف در مجموعه داده Vehicle_sensIT ۱۰۸

شکل ۴-۵ ساختار پیشنهادی برای مسئله پیش‌بینی ترافیک شهری ۱۱۲

شکل ۴-۶ روال چارچوب پیشنهادی برای پیش‌بینی ترافیک شهری ۱۱۴

شکل ۴-۷ مقایسه زمان اجرای الگوریتم پیش‌بینی ترافیک شهری با استفاده و بدون استفاده از چارچوب پیشنهادی ۱۱۵