Call for Proposals

BigData & Artificial Intelligence

HIRP OPEN 2017

Application Deadline: 09:00 A.M., 16th June, 2017 (Beijing Standard Time, GMT+8).

If you have any questions or suggestions about HIRP OPEN 2017, please send Email (innovation@huawei.com). We will reply as soon as possible.

Catalog

HIRPO2017060301: Deep learning models for recommender system and Context-Aware recommendation

1 Theme: Big Data & Artificial Intelligence

2 Subject: Technology in Recommender System

List of Abbreviations

DNN: Deep Neural Network

FNN: Factorization-Machine supported Neural Network

PNN: Product based Neural Network

3 Background

Learning feature interactions automatically via non-linear models have been an important research and applicable direction in recommender system. Researchers have developed some deep learning models for recommendation system, such as FNN, PNN, which perform well on public datasets. While in industry, several companies applied deep learning models successfully in real applications, e.g. Google proposed Wide & Deep model for Google Play and developed its DNN for YouTube. Researchers in Noah's Ark lab also devised several advanced deep learning models, which performs better than the state-of-the-art models on both public datasets and commercial datasets.

In Huawei Application Store, there are several challenging scenarios for recommendation, like "hot apps in local", "guess you like", "fancy new apps", etc. Besides main lists in Huawei Application Store, the above scenarios will

bring significant improvement about user's experience, if the personalized recommendation performs well.

"Deep learning in recommendation" and "Context-Aware recommendation" are not only interesting directions for academia researchers, but also important in industry applications.

4 Scope

1) Research on deep learning models for recommender system: investigate the state-of-the-art deep learning models, design new deep learning models for recommender system and evaluate the performance of the models on datasets we provide;

2) Research on Context-Aware recommendation in recommender system: survey on the state-of-art algorithms, design recommendation algorithms for Context-Aware recommendation.

5 Expected Outcome and Deliverables

1) Technical report of deep learning models for recommender system and context-aware recommendation;
2) New models, algorithm design and implementation on aforementioned research areas.

6 Acceptance Criteria

All the deliverables are accepted by technical team, Huawei.

7 Phased Project Plan

Phase1 (~6 months): survey the state-of-the-art algorithms in deep learning for recommender system and context-aware recommendation, and design the models and algorithms in related areas;

Phase2 (~6 months): Implement and evaluate of the designed models and algorithms.

Click here to back to the Top Page

HIRPO2017060302: Learning-based anomaly detection and prediction

1 Theme: Big Data & Artificial Intelligence

2 Subject: Network Intelligent Management Technology

List of Abbreviations

AI: Artificial Intelligent

MBB: Mobile BroadBand

3 Background

With the advent of MBB network technologies, mobile Internet access becomes the norm in everyday life. In the meantime, network becomes more and more complicated and guarantee subscribers' quality of experience (QoE) becomes very challenging. From a network operator's perspective, achieving accurate and timely network fault localization and anomaly detection are critical for reduce OPEX and improve QoE.

So, it is a valuable research direction of data driven, learning based network fault localization and network anomaly prediction / detection.

4 Scope

1) Research on network fault localization and root cause analysis: investigate the state-of-art methods, design new solution for network fault localization and root cause analysis based on datasets we provide;

2) Research on network KPI, traffic, anomaly prediction / detection: survey on the state-of-art algorithms, compare their performance and accuracy, design optimized algorithm for related network problems.

5 Expected Outcome and Deliverables

Technical reports of State-of-art algorithms in data driven, learn-based network fault localization, and root cause analysis, anomaly detection / prediction.

New solution design and implement on aforementioned research areas;

6 Acceptance Criteria

All the deliverables are accepted by technical team, Huawei.

7 Phased Project Plan

Phase1 (~3 months): survey the state of the art of data driven, learn-based network fault localization, and root cause analysis, anomaly detection / prediction.

Phase2 (~5 months): Research on optimized algorithms and solutions in related areas.

Phase3 (~4 months): Implement and validate of optimized algorithms and solutions.

Click here to back to the Top Page

HIRPO2017060303: Multivariate Time Series Prediction Techniques with Fixed and Non-Fixed Relationship

1 Theme: Big Data & Artificial Intelligence

2 Subject: Supply Chain Planning Technology

List of Abbreviations

SCM (Supply Chain Management)

PTO(Pick to Order)

PTO(Pick to Order)

AI (Assembly Item)

P(Purchased Item)

3 Background

Huawei is one of the leading global communication device providers, of which products are launched in more than 170 countries/areas. In Huawei, an efficient supply chain management (SCM) is essential to guarantee a good service quality as well as a profitable business. Demand forecasting, which creates reliable forecasts of demand to align inventory levels with peaks and troughs in demand, is one of the core tasks in SCM. One percent improvement of forecasting accuracy can lead to million dollars of cost savings. Huawei's supply chain needs to fulfill requirements imposed by tens of thousands of business cooperators all over the world. In addition, there are more than 100K different items (PTO, AI, P) with fixed and non-fixed hierarchical structures and substitutable relationships in our productions. For example, the non-fixed structure of PTO and AI is determined by the customer order and the structure

of AI and P which is fixed depends on the manufacture BOM. This high production complexity as well as the high customer diversity bring new challenges to demand forecasting. Therefore, it is a valuable research direction to analyze and investigate demand forecasting techniques.

4 Scope

1) Research on demand forecasting: investigate the relationship of different items and apply the fixed and non-fixed structure to design demand forecasting algorithm'

Input: historical custom order, manufacture BOM and historical outbound quantity of each item.

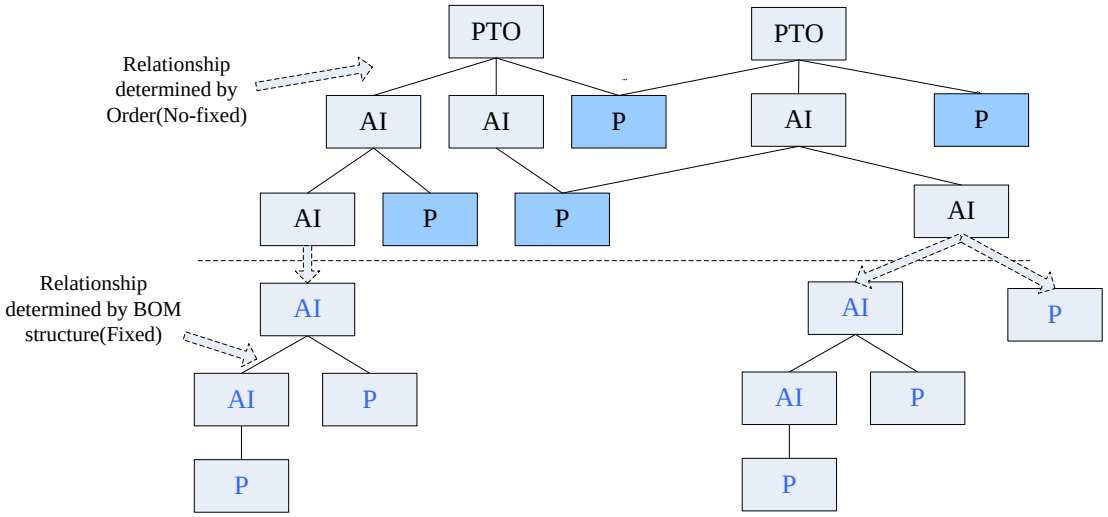
Output: demand of each item in future $(X^{i}_{t+1}, X^{i}_{t+2}, \dots, X^{i}_{t+k})$

Challenge： the fixed and non-fixed structure of Items

State of the Art:

1. **Vector Auto-regression:** estimate the relationship of different items directly from the outbound quantity.
2. **Hierarchy time series models：** apply the fixed hierarchy structure to adjust the forecast.
3. **Network vector regression:** apply the fixed network structure to improve the forecast accuracy

Picture 1 Product structure



5 Expected Outcome and Deliverables

Technical reports of demand forecasting model, including data/parameter measurement study, technical survey, algorithm design and verification result.

Improve 10% of the prediction accuracy compared to methods of state of the art (Arima)

Algorithm source codes and documents.

One high quality industrial patent/paper of high impact factor (accepted by the first tier data mining, machine learning or Artificial Intelligence conferences/journals).

6 Acceptance Criteria

All the deliverables are accepted by technical team, Huawei.

7 Phased Project Plan

Phase1 (~3 months): survey the state-of-the-art techniques of demand forecasting, perform data/parameter measurements and provide relevant technical reports.

Phase2 (~5 months): research on forecasting models, verify the prototype solutions and provide relevant technical reports.

Phase3 (~4 months): optimize the solutions, verify the effectiveness of methods by comparing with related techniques coping with Huawei's data, and summarize patent/paper.

Click here to back to the Top Page

HIRPO2017060304: Exploration of new methodology of human brain-inspired computing

1 Theme: Big Data & Artificial Intelligence

2 Subject: Human Brain-Inspired New Computing Paradigm

List of Abbreviations

DNN: Deep Neural Network

A.I.: Artificial Intelligence

3 Background

We are on one hand enjoying the outcomes brought by DL in the tasks of image recognition, speech recognition. On the other hand, the development of DL is facing a bottle-neck due to its limitation. Breakthrough of A.I., in both theoretical and practical perspectives, will not be possible, if we do not come up with new ideas other than DL.

We are currently seeking for a potentially viable alternative to DL which can enhance any of the A.I. sub-fields, including machine learning, natural language processing, computer vision, recommendation and searching, knowledge inference, human-robot interaction, etc.

Our attention has been drawn to computing paradigm inspired by the very mechanism of how a brain works. For a thorough understanding of our brain, one has to consult neuroscience. Therefore, we look for a multi-disciplinary collaboration in the fields of neuroscience and computer science, to achieve a

novel design of brain-inspired algorithms and its realization. We expect such collaboration would produce an application-driven outcome.

4 Scope

1) **Research on neuroscience:** investigate how our brains process with cognition issues on deeper levels.
2) **Research on brain-inspired algorithm:** based on the aforementioned neuroscience research outcome, design and devise algorithms which mimic the cognition mechanism of human brains. The algorithms must be different compared to the existing DL models.

5 Expected Outcome and Deliverables

At least one novel brain-inspired algorithm in any of the A.I. sub-field is proposed and tested.

1 – 2 published papers in the top-tier journals and conferences.

No special requirement on patents.

6 Phased Project Plan

Phase1 (~6 months): Research on how our brains represent information and process the information.

Phase2 (~6 months): Research on an algorithm which addresses the representation method and processing method.

Click here to back to the Top Page

HIRPO2017060305: AI Tool for Patent Novelty Search

1 Theme: Big Data & Artificial Intelligence

2 Subject: AI Tool for Patent Novelty Search

List of Abbreviations

AI: Artificial Intelligence

3 Background

Currently the artificial intelligence technology is increasingly being used in different areas, and achieved very good results. Patent novelty search is a necessary basic work in the patent application and examination. A large number of patent applications and massive prior art require a lot of examiners and time, and the novelty search efficiency is low and the quality is difficult to be guaranteed. Research and development an AI tool can be helpful for improving the quality and efficiency of patent novelty search.

4 Scope

Research and develop the AI tool for patent novelty search.

5 Expected Outcome and Deliverables

Technical reports of the AI tool for Patent Novelty Search;

Solution design of the AI tool;

The software package of the AI tool with source codes and description;

6 Acceptance Criteria

Project proposal is accepted by the evaluation team, Huawei.

Project deliverables are accepted by the evaluation team, Huawei.

The search quality of AI tool is better than the search quality of the high level patent examiner from Patent Office.

7 Phased Project Plan

Phase1 (~2 months): Research and design the solution of AI tool for patent novelty search.

Phase2 (~3 months): Developing the AI tool for patent novelty search.

Phase3 (~1 months): Testing and acceptance the AI tool ;

Click here to back to the Top Page

HIRPO2017060501: Self-learning microservice governance autopilot

1 Theme: Big Data & Artificial Intelligence

2 Subject: Microservice Technology

3 Background

In a microservice system, there are many governance approaches are provided to make sure the whole system is workable. The APM information is the input of these governance actions. The user need to check those data, find the abnormal services and take actions which are provided by the system. Also, the ways to do the governance is growing and those events are hiding in the data and very difficult to find out in a large system. In this situation, a self-learning robot is need to do this replacing the human.

4 Scope

1) Research on self-learning system upon the APM data

2) Research on the modeling to the microservice governance actions which leads different results and may affect each other.

3) Research on robot which take actions based on the knowledge

5 Expected Outcome and Deliverables

- An autopilot system which is able to forecast problems and automatically manage microservices system, such as take measures before a microservice failure and recover failure automatically

- Integrate with APM system, and provide a UX for operator to monitor microservice status and operate manually instead of autopilot
- Daily Report for Autopilot System
- 1~2 Invention/patents;

6 Acceptance Criteria

Autopilot operation false rate <0.1%

24*7 High-available system

7 Phased Project Plan

Phase1 (~3 months): Research about AI framework, decide which is best for autopilot. learn on APM data, choose what is the valuable data for autopilot to manage microservices and what autopilot is able to do with those data. How to integrate with APM and microservice system. Provide a technical report

Phase1 (~6 months): Implement Autopilot system, and run it in production environment

Phase1 (~3 months): Write patents

Click here to back to the Top Page

HIRPO2017060502: Serverless function scheduling prediction

1 Theme: Big Data & Artificial Intelligence

2 Subject: Design: Serverless Technology

3 Background

In Serverless system, there are a varieties of approaches to reduce the function call latency. Data mining historical function call data might be able to find function reference patterns. It helps scheduling related functions closely and intelligently shorten the latency in between. In addition, periodical function reference peaks can also be predicted, which allows function runtime to be initialized in advance.

4 Scope

1) Research on self-learning system from Serverless function call data
2) Research on modeling Serverless function calling patterns and how to affect scheduling
3) Research on real-time data mining and strategy adoption.

5 Expected Outcome and Deliverables

- An prototype system which is able to find related functions and predict function peaks.
- Integrate with Serverless system to schedule function in proper resource and prepare resource pools for function scheduling.

- 1~2 Invention/patents;

6 Acceptance Criteria

Find 90% related function call

Give function call scheduling priority for all the combination (resource requirement, language, network requirement, etc).

7 Phased Project Plan

Phase1 (~3 months): Research about self-learning framework, decide which is best for data mining. learn on Serverless historical function, choose what is the valuable data for prediction. what prediction is able to do with those data. How to integrate with Serverless system. Provide a technical report

Phase2 (~6 months): Implement prototype system, and integrating with Serverless with

Phase3 (~3 months): Write patents

Click here to back to the Top Page

HIRPO2017060503: A Double-Edged Sword: The Vulnerable Side of Mobile Application Recommendations

1 Theme: Big Data & Artificial Intelligence

2 Subject: Search and Recommendation

List of Abbreviations

App: mobile application

CF: collaborative filtering

3 Background

The number of mobile applications (or apps) is huge and is still growing rapidly and continuously. It has become increasingly tedious for users to locate the right apps to meet their needs. As a consequence, there is an immense demand for effective app recommendations. App recommendation is a double-edged sword. On the one hand, it can help users find the right apps that can remarkably improve user experiences in mobile devices and can establish the trust between users and the system. On the other hand, it also has attracted malicious users (or attackers) to manipulate the system and to make the system produce recommendations as they desire (i.e., promoting apps from their own companies or demoting apps from their competitors), which can damage the reputation of the system and can degrade the trust from users tremendously. Currently the majority of existing works focus on the side of building better recommender systems; while ignoring the vulnerable side. In this project, we will investigate the dark side of app recommendations to bridge this gap and aim to build robust app recommender systems.

4 Scope

Aim 1: identify new types of attacks (or vulnerabilities) for app recommendations

App recommendations need to consider some unique factors (e.g., privacy preference and retention rate) in addition to the primary one as traditional recommender systems. Hence, we will first investigate whether we can apply traditional attacks, then identify unique attacks for app recommendations and finally reveal the typical strategies of implementing these attacks.

Aim 2: detect attacks with multiple sources

One app can be related to multiple sources (e.g., app descriptions, comments, and download histories) that could enable us to develop more accurate attack detectors than using one single source. We will study important cues from each source that can discriminate attacks and then build machine learning algorithms with multiple sources to detect attacks automatically.

Aim 3: build app recommender systems robust to attacks

Different techniques can be used to build an app recommender system such as content- and CF- based methods. Thus, we will investigate the robustness of these techniques to attacks and study solutions to reduce their vulnerabilities in app recommendations.

5 Expected Outcome and Deliverables

- Technical reports of a list of possible attacks or vulnerabilities to app recommendations and their frequently adopted strategies to conduct such attacks;

- Technical reports of attack detection algorithms including important cues to distinguish attacks, methods to combine multiple sources and the performance evaluations;
- Technical reports of robust recommender systems including various solutions to reduce the vulnerabilities in different types of recommendations.
- 1~2 Invention/patents;

6 Acceptance Criteria

The proposed research will identify and detect attacks (or vulnerabilities) in app recommendations and provide solutions to defend such attacks, which can be integrated to boost the performance of current app recommender systems and greatly improve user experiences.

7 Phased Project Plan

Phase 1 (~3 months): survey the state-of-the-art of attacks in traditional recommender systems, analyze potential attacks in app recommendations and prepare the related technical report.

Phase 2 (~4 months): study attack detection, including important attack patterns, and methods to combine multiple sources, empirical evaluations, provide the related technical report and prepare a patent.

Phase 3 (~5 months): investigate the defense strategies for detected attacks, combine them to build novel and robust recommendation algorithms, conduct evaluations and prepare related technical report and a patent.

Click here to back to the Top Page

HIRPO2017060504: Markov Logic Networks for Switch Fault Analysis and Diagnosis

1 Theme: Big Data & Artificial Intelligence

2 Subject: Casual Analysis and Association Analysis

List of Abbreviations

MLNs: Markov Logic Networks

SFAD: Switch Fault Analysis and Diagnosis

3 Background

Switch fault analysis and diagnosis play a fundamental role in the operation and maintenance of communication equipment manufacturer. Due to the complexities of the composition and structure, the large network switches usually have various types of faults. Based on the association analysis of the big data from the switch fault maintenance, timely checking the reasons of the failure is absolutely necessary. In most cases, people manually identify the sources of the faults by tracing and replaying the execution of the switches, which can be very time consuming and labor-intensive. This naturally raises the question of whether it is possible to apply the machine learning and statistical inference techniques to the whole process.

Actually, this problem can be formalized as interactions or influence between various types of objects such as behaviors, events, status, reasons and diagnosis types. However, the inherent complexity and uncertainty are both quite difficult to be expressed by traditional features. By unifying the strengths of probabilistic graphical models and first-order logic, Markov logic networks (MLNs) offer a general framework for representing well-defined probability

distributions over uncertain, relational data. Its powerfulness and flexibility thus make it highly suited for switch fault analysis and diagnosis (SFAD) in an efficient way.

Therefore, it is a valuable research direction to develop efficient MLN approaches to automatically analyze the associative characteristics of the large-scale switch fault data which has been collected by Huawei, and to ensure high accuracy in detecting the principle sources of the switch faults. This undoubtedly has great theoretical and practical values.

4 Scope

1) Learning MLN Structure for Switch Fault Analysis:

Investigate MLN-based formulation for the analysis of switch faults signals; design efficient algorithms for learning the latent MLN structure that capture complex dependencies among pairwise interactions; explore the scoring strategies to identify a compact representation from the learned new rules and those supplied by the domain experts.

2) Probabilistic Inference for Identifying Switch Fault Reasons:

Investigate a scalable probabilistic inference model that is suitable for fast switch fault diagnosis; develop self-supervised and minimally supervised techniques, helping the system autonomously infer additional relevant facts for higher diagnosis performance; design robust algorithms to combine the uncertain facts and infer the underlying possible fault reasons with good precision and recall rate.

5 Expected Outcome and Deliverables

- Academic papers on MLN-based learning approach for switch fault detection;

- Academic papers on probabilistic inference for identifying switch fault reasons.
- Related simulation platform with source codes and description.
- 1 Invention/patents on intelligent analysis and diagnosis of the switch faults;

6 Acceptance Criteria

- Project proposal is accepted by the evaluation team, Huawei.
- Project deliverables are accepted by the evaluation team, Huawei.
- No fewer than 2 academic papers indexed by SCI.

7 Phased Project Plan

- Phase1 (~3 months): Survey the state of the art of switch fault diagnosis in enterprise networks. Identify the problems, metrics and requirements in this topic. Form the project research report.
- Phase2 (~6 months): Build the MLN-based formulation for the analysis of switch faults signals, and implement the MLN learning and inference algorithms. Form the solution design documents and the brief evaluation of the core idea. Prepare the academic papers.
- Phase3 (~3 months): Deliver the concrete simulation results of all the solutions proposed in the project. Verify the simulation platform based on the big data provided by Huawei.

Click here to back to the Top Page

HIRPO2017060601: Research on Graph Analytics over Stream Processing

1 Theme: Big Data & Artificial Intelligence

2 Subject: Stream Graph Analytics

List of Abbreviations

Gelly: Graph library over Flink

Flink: Is unified batch and stream processing engine

3 Background

Nowadays an increasing number of Big Data applications are dealing with heterogeneous and diverse data sources which carry implicit and explicit associative relations among the objects (or entities) represented in them. The analysis of such networks of objects can be mapped to graph processing scenarios. Hence, supporting graph analysis at large scale becomes essential also in various vertical scenarios in the context of IoT, finance or telco, particularly for applications that deal with risk analytics, customer profiling, targeting marketing or market evolution.

Processing graphs at scale requires on the one hand a robust execution engine able to cope with the complexity of large graphs algorithms executions and on the other hand to a rich library of graph algorithms to support a wide range of applications and analysis. This brings forward many challenges: 1) scaling the underlying processing engines, in terms of nodes, relations or nodes used for the computation; 2) enhancing the algorithms themselves for running in a distributed fashion; 3) reducing or eliminating the disk IO by moving the computation in memory; 4) decreasing the overall computation

time to meet stringent SLA. Furthermore, the constant demand for timely analytics based on the latest data pushes all these challenges to be consider the execution of graph processing over continuous incoming streams of events.

Unfortunately, the modern graph processing engines are not yet ported on the stream paradigm, limiting their application to batch executions. Real-time graph processing, unlike typical stream processing application, add a further factor of complexity given by the fact that a large number of items with large number of relations needs to be constantly updated in time. This requires dedicated algorithms capable of providing online incremental approaches and coping with the complexity at large scale computation. This shows the necessity for an extensive research study to explore architectural options to enhance modern stream engines, such as Apache Flink, to work in stream graph processing scenarios.

4 Scope

The goal of the project is to investigate solutions for supporting in-memory graph analytics at large scale for stream analytics. The ideal solution would propose an extension to Gelly, the graph processing library of Flink, for supporting stream analytics for graph processing. The online set of graph algorithms should be able to support millions of graph nodes (i.e., of various types), and several times (e.g., 2x – 20x times) more relations between nodes, with relations of different types. Moreover, the approaches developed should consider various techniques for enhancing the processing performance such as incremental computation, iterative computation, partition of data and computation across nodes, caching, etc. The goal is that by using such

techniques and optimizations most of the graph algorithms ported to the stream engine are able to run with second-level performance.

The key research challenges, issues and requirements to be addressed are:

- Achieving high performance and low latency for in-memory processing or graphs
- Porting the stream graph algorithms for distributed processing
- Update the graph in (nearly) real time
- Supporting large number of graph vertexes (in the order of millions), relations (2x-20x the number of vertexes) and relation types (tens)
- Efficient propagation of changes and stimulus in the graph with low latency
- Time evolution of the whole graph with stability checking, by creating future graph estimations (12-24 future graph iterations) and support for iterative graph algorithms
- Porting the graph operations (partition, distribution, statistics, algorithms) for online processing (i.e., in-memory stream processing)

5 Expected Outcome and Deliverables

The expected outcomes for the project are:

1. Evaluation study about existing graph algorithms and approaches on how to port them to stream processing. The goal is to identify best practices and techniques to tackle the requirements introduced in Section 4.

2. Architecture options for large scale real-time stream graph analytics. The goal is to design solutions for porting the graph algorithms and the approaches identified in the evaluation study for stream processing (e.g., Flink engine extension)

3. A library for Flink stream engine that implements the architecture of point 2. The goal is to show that the algorithms and approaches developed can meet the performance requirements mentioned in Section 4 (in-memory stream processing, scale to million nodes and relations, achieve most of the times second-level performance, create estimates for graph evolution in time).

6 Acceptance Criteria

The acceptance criteria with respect to the outcomes are:

1. Evaluation study meets the academic norms of algorithm/paper/technical writing and research analysis

2. Graph algorithms can manage more than 1 million nodes and 2 million relations in-memory.

3. Algorithms are compatible and run based on DataStream API of Flink

4. Most graph algorithms can perform the processing in second-level and run online. For the algorithms or graph summarization technique that cannot meet second-level performance, minute-level needs to be targeted

5. Library can support evolution of graphs for 24 iteration steps, can support graph updates and iterative algorithms and the intermediate results of the iterations and of algorithms can be accessed

6. The incremental calculation model is clearly defined, sufficiently common and can be applied to most graph algorithms
7. Code and demo system for the implementation of the architecture described in point 2 which demonstrate the performance requirements
8. Code and prototype should target open source standards (e.g. from Flink/Gelly project)
9. For the algorithms targeted for incremental processing the solution should include but not be limited to the following algorithm: PageRank, SSSP, SimRank, FastUnfolding, Community analysis, Similarity analysis. Additionally, summarization related computation such as limited working memory and million edges, graph summary, including eccentricity, diameter, cluster coefficient, community, structure analysis etc., should be computed and updated exactly and efficiently

7 Phased Project Plan

Expected project Duration (year): 1 year

Project Phase	Start Time	End Time	Content	Objective	Output
Phase 1	T	T+3	Evaluation of existing solutions and algorithms in the context of streaming	1. Identify limitations of existing solutions 2. Identify best architectural options for stream portability	1. Report 2. Algorithms approaches 3. Architecture design guidelines

				3. Identify options for scaling the algorithms (in time, based on vertices and relations) and across iterations	
Phase 2	T+3	T+6	Architecture design of stream graph analytics	1. Identify architecture options for integrating online graph algorithms into Flink and for supporting (near) real time graph updates 2. Identify approaches relations in the graph 3. Identify best approaches	1. Report 2. Algorithms 3. System architecture

				to scale the processing of the graph evolution in time	
Phase 3	T+6	T+12	The Stream Graph Processing for Flink	1. Implement the approaches into Flink 2. Scale the online graph algorithms to million nodes/relations 3. Optimize the performance of most graph algorithms to run in second-level	1. Demonstrator 2. Code 3. Final Report

Click here to back to the Top Page

HIRPO2017060602: Equipment Status and Data Intelligence Analysis Engine

1 Theme: Big Data & Artificial Intelligence

2 Subject: Intelligent Operations

3 Background

The operations trends of IT cloud hardware, especially heterogeneous hardware, will move forwards the intelligent in the future, the alarm-oriented service operations of traditional CT is difficult to adapt to the continuous increasing scale of IT, especially hardware isomerization, which makes this scene worse. Through the analysis of big data, it will transform the operations from the alarm-oriented manual to the data-oriented automation.

Lightweight engines help reduce deployment costs and explore the feasibility of lightweight data analytics engines to support private cloud applications.

4 Scope

Problem to be resolved:

1. Offline learning: based on the hardware log, events, and status data for association learning, access to data and hardware status correlation judgment parameters, as an online reasoning system input.

2. Online learning: based on manual adjustment as input, the learning results are corrected or automatically added to the new reasoning judgment parameters.

3. Fault analysis: the ability to determine the status of the hardware device failure, and analyze the cause of the failure, the output fault measures recommended.
4. Fault prediction: can be on-line analysis of the data to make hardware what kind of failure prediction, and give the proposed measures.

Note: Deployment scale is less than 1000 servers, belonging to small-scale applications

5 Expected Outcome and Deliverables

Provide workable software code and documentation.

6 Acceptance Criteria

By entering the existing log, event or status data, the deliverables can analyze the correct status of the results to meet the lightweight needs:

1. Deployment size: minimum deployment requirements, 1 host, 4P, 8G memory
2. Deployment forms: support stand-alone deployment and distributed deployment mode, distributed deployments provide greater performance on multiple compute nodes
3. API interface: provide API interface for other software calls
4. Condition setting: support online settings, modify the fault judgment conditions, and support based on the new conditions for fault diagnosis.

7 Phased Project Plan

Expected project Duration: 18 months

Project Phase	Start time	End time	Work content	work goals	Output
Phase 1	T	T+4	Complete the feasibility of the program analysis	Identify alternatives	Submit the program feasibility study document
Phase 2	T+5	T+10	Build the basic framework	Complete the construction of the learning framework	Frame prototype
Phase 3	T+11	T+14	Data input and frame adjustment	Complete the data input learning and reasoning, and according to the results of the framework to adjust	Can work on the framework
Phase 4	T+15	T+18	Training and use of existing products	Product environment application, problem modification	Can be a product of the framework

Click here to back to the Top Page

HIRPO2017061101: Fault Management Algorithm under 5G Network Micro-service Architecture

1 Theme: Big Data & Artificial Intelligence

2 Subject: Smart operation and maintenance algorithm

List of Abbreviations

5G	5th Generation
OM	operation and maintenance

3 Background

With the cloud computing, virtualization and other internet technology become mature gradually in CT industry, cloudify become a trend in the entire CT industry. Thus, all cloud strategy is becoming a reality. The key features of cloudification are resource pooling, software fully distributed and operation automation. Therefore, a flexible architecture should be used to deal with the uncertainty of new business in order to achieve a smooth evolution to the 5G network. The micro-service architecture is the core enable technology that make a flexible 5G network environment. However, micro-service architecture not only bring flexibility to 5G network but also operational challenges, for example anomaly detection, fault diagnosis, root cause analysis, fault recovery, etc. The potential scenes of 5G micro-service architecture may include a single micro-service, interactive micro-service chain and cross-geographical distribution of micro-services. These scenes make operation and maintenance more challenging.

4 Scope

- Survey on characteristics of different micro-service, fault management algorithm solution in industry and academia, research trend of micro-service fault management as well as feasibility analysis of building simulation environment in cloud environment.
- Firstly, modeling the 5G micro-service, then design fault management algorithm for typical 5G micro-service use cases. Fault management algorithm may include anomaly detection, fault diagnosis, root cause analysis, fault recovery.
- According to previous analysis, building a simulation platform in the cloud to validate the proposed micro-service fault management algorithm.

5 Expected Outcome and Deliverables

- 1 survey report
- 1-2 research report
- 2-3 algorithms
- A demo
- 1-2 patents
- 1-2 papers

6 Acceptance Criteria

- A detailed deliverables in section 6

7 Phased Project Plan

Phase1 (~2 months): survey on characteristics of different micro-service, fault management algorithm solution in industry and academia, research trend of micro-service fault management as well as feasibility analysis of building simulation environment in cloud environment.

Phase2 (~5 months): design fault management algorithm for typical 5G micro-service use cases, which may include anomaly detection, fault diagnosis, root cause analysis, fault recovery.

Phase3 (~3 months): building a simulation platform in the cloud to validate the proposed micro-service fault management algorithm.

Click here to back to the Top Page