Strategic Plan

RII Track-2 FEC: Precise Regional Forecasting via Intelligent and Rapid Harnessing of National Scale Hydrometeorological Big Data

Vision

Better regional weather forecasting for wide deployment to facilitate applications with societal and economic benefits.

Mission

To develop portable machine learning models and effective model acceleration methodologies for intelligent and rapid data harness toward precise regional forecasting with inputs from both WRF-HRRR computed weather parameter values and near-surface weather observation results.

Research Theme 1: Accelerating and Enhancing Big Data Harness

Goal 1.1: *Design and evaluate system architectural approaches for accelerating bigdata analytics based on machine learning over massive weather datasets.*

- Objective 1.1.1: In-Memory Processing (aka Project Task 3 led by **N. Tzeng)**

  To accelerate data harness via utilizing massive physical memory available on an entire system through on-demand memory requests over low-latency connecting mechanisms. This research objective involves the following two activities, with their specific milestones provided.

  + GPU-Assisted memory expansion (GAME)

    Year 1: GAME architectural design and platform establishment

    Year 2: Performance evaluation and fine-tuning based on various benchmark applications

    Year 3: GAME-oriented environment development for Modelet training

  + Adaptive memory expansion

    Year 1: Develop adaptive control on memory expansion onto remote machines or GPU cards

    Year 2: Model adaptive control performance based on various system parameters

- Objective 1.1.2: Stochastic Computing (aka Project Task 5 led by **H. Najafi**)

  To address data harness acceleration resulting from the desirable bit-stream representations that enjoy the property of progressive-precision in computation, to permit dynamic run-time control of the quality of results. This research objective exploits stochastic computing (SC) support for neural network (NN) computation improvement, with subsequent project activities.

  + Fast converging low-discrepancy (LD) deterministic bit-stream exploitation

    Year 1: Exploring Low-cost and energy-efficient methods for generating LD bit-streams Finite-State-Machine-based LD SC for extremely parallel data processing

    Year 2: Integrating LD bit-streams and residue number systems for exponentially lower latency and energy

    Year 3: Exploiting LD bit-streams for accurate and high-performance SC-based CNN computation

  + SC to improve Modelet computations

    Year 2: developing ultra-low-power SC-based accelerators for hardware-based implementation of convolutional and deep neural networks

Year 3: Evaluating performance and hardware efficiency of SC-based Meteo and Hydro Modelets compared to the state-of-the art hardware-based counterparts

+ Exploiting skew tolerance of SC to improve performance and energy consumption

Year 2: Removing and relaxing clock distribution network (CDN) of the developed SC-based designs to minimize the overhead cost of their CDNs

Year 3: Exploring aggressive voltage and frequency scaling for lower energy but comparable performance

+ Near-sensor processing of massive weather data

Year 1: Employing time-based SC for near sensor design of multiplication operations and convolution functions

Year 2: Near-sensor processing of weather data using SC-based Modelet computations in the time domain

Goal 1.2: *Devise and assess algorithmic enhancement for machine learning (ML)-based bigdata analytics.*

- Objective 1.2.1: Execution Resilience (aka Project Task 4 led by **L. Peng**)

  It is critical to enhance execution resilience for lengthy big data harness executions, like modelet training. Given such an execution on a multiple-core server typically spawns multiple threads for its speedup via parallelism, this research objective involves the project activities below.

  + Comprehensive software-based error detection to enhance resilience

    Year 1: Synchronization mechanism protection in parallel big data programs

    Year 2: Resilience analysis, detection, and protection for developed Meteo and Hydro modelets

  + Fault-tolerant architecture design for bigdata execution

    Year 2: Fault-tolerant ReRAM accelerator for neural networks

    Year 3: Dependability analysis and improvement for neural network accelerators

- Objective 1.2.2: GPGPU Support (aka Project Task 6 led by **L. Chen**)

  This research objective focuses on data parallelism since our proposed weather forecasting involves a large number of Modelets with simple NN structures for each region, involving subsequent project activities.

  + Scheduling Modelets on a GPGPU for rapid inference

    Year 1: Scheduling policy design for rapid processing and maximal throughput of inference workloads

    Year 2: Performance evaluation and fine-tuning with various Modelet benchmark workloads

  + Framework development for optimal concurrent Modelet training over multiple GPGPUs

    Year 1: Mechanism design for inter-Modelet and intra-Modelet scheduler of the training framework, with performance evaluation and fine-tuning

    Year 2: Mechanism design for communication optimizer of the training framework, with performance evaluation and improvement

Research Theme 2: *Improving Meteorological Prediction and Its Applications*

Goal 2.1: *Meteorological ML model (Meteo Modelet) development and evaluation.*

- Objective 2.1.1: Surface Weather Data Preprocessing and 1-D Meteo modelet improvement (aka Project Task 1 led by **X. Yuan**)

This research objective aims to deal with spatial and temporal resolution mismatches of datasets from multiple sources, to develop algorithms and software for integrating those diverse datasets as inputs to Meteo Modelets, and to rectify and fine-tune Meteo Modelets under development for better prediction accuracy. It addresses following project activities.

+ Mesonet data regridding

  Year 1: Effective regridding to refine spatial and temporal resolutions of Mesonet data, with an aid of geographic and topographic information; completing the spatial interpolation model under development based on geographically weighted regression (GWR)

  Year 2: 1-D Meteo Modelet evaluation and comparison under regridded Mesonet data versus under GWR modeling results for ungauged location prediction

+ Personal weather station (PWS) data rectification and regridding

  Year 1: Effective regridding to refine spatial and temporal resolutions of PWS data

  Year 2: 1-D Meteo Modelet evaluation and comparison under regridded PWS data

+ 1-D Meteo modelet improvement

  Year 2: Meteo Modelet prediction on ungauged locations using regridded Mesonet data

  Year 3: Meteo Modelet improvement with fused Mesonet data and PWS data as inputs

Goal 2.2: *2-D Meteo Modelet development and assessment.*

- Objective 2.2.1: Developing and evaluating improved Meteo Modelets with 2-D inputs (aka Project Task 7 co-led by **S. Foster and X. Yuan**)

  This research objective is to let Meteo Modelets admit regridded Mesonet and PWS datasets as 2-D inputs for better prediction accuracy, as a result of utilizing the measurements and computed results of nearby grids. It addresses following project activities.

  + 2-D Meteo Modelet development and evaluation

    Year 2: New NN-based model design, development, and evaluation with regridded Mesonet datasets; evaluating GWR-based spatial interpolation model with both Mesonet data fused with PWS data available to the public

    Year 3: 2-D Meteo Modelet evaluation and fine-tuning the inputted 2-D Mesonet and PWS datasets, with proper data fusing schemes; elevating prediction accuracy by possibly taking GWR results into consideration

Goal 2.3: *Meteo Modelet execution acceleration.*

- Objective 2.3.1

  This execution acceleration objective evaluates applying acceleration techniques devised under Research Theme 1 to both 1-D Meteo Modelets and 2-D Meteo Modelets which are inputted with regridded Mesonet and PWS datasets under proper fusion. It is to be conducted after those acceleration techniques have been fully developed and evaluated, mostly during Year 4.

  + In-memory processing
  + Stochastic computing
  + Execution resilience
  + GPGPU execution support

Goal 2.4: *Meteo Modelet result applications.*

- Objective 2.4.1: Local meteorological research and operational forecasting (aka Project Task 8 co-led by **S. Kimball and X. Yuan**)

It is critical for predicting severe thunderstorm activities, landfalling tropical systems, and local convection initiation (CI) process studies based on detailed information of local atmospheric profiles and stability. This research objective includes subsequent project activities.

+ Weather prediction improvement with an aid of Meteo Modelets

  Year 1: Developing local climatology of low-level inversions and investigating sea breezes on CI

  Year 2: Refining regional severe weather prediction models under development; deploying UAVs and balloons for observing atmospheric boundary layer processes

  Year 3: Predicting regional severe weather precisely with the aid of 2-D Meteo Modelets to identify severe thunderstorm formation and landfalling tropical initiation

+ Targeted forecasting to determine critical regions for emergency management enhancement

  Year 3: Applying information on predicted local CI and landfalling tropical onsets to traffic alerts

  Year 4: Prototyping such alert application with visualization interface (under Objective 2.4.2 below), for possible technology transfer

- Objective 2.4.2: Regional weather forecasting plots/charts based on Meteo Modelet results, to be developed by **S. Foster and E. Rappin**.

  + Meteo Modelet result visualization

    Year 1: Regional forecast plots/charts using only WRF-HRRR data

    Year 2: Improved regional forecast plots/charts using 1-D meteo Modelet results

    Year 3: Precise regional forecast plots/charts using 2-D meteo Modelet results

Research Theme 3: *Hydrological Prediction Improvement and Result Applications*

Goal 3.1: *Hydrological ML model (Hydro Modelet) development and evaluation.*

- Objective 3.1.1: Hydro modelet design and development (aka Project Task 9 co-led by **L. Chen and N. Tzeng**)

  This research objective aims to complete NN-based ML modelets for flood monitoring and river inundation forecasting in real time at regions of interest, inputted with the readings of both deployed streamgages and water gauges to be deployed. It addresses the following project activities.

  + Improving and evaluating 2-D Hydro Modelets

    Year 2: Rectifying Hydro Modelets with their inputs from deployed streamgage and water gauge readings and also from the rainfall estimates via the Multi-Radar Multi-Sensor (MRMS) system implemented at National Centers for Environmental Prediction (NCEP) with high spatial (of 1 *km*) and temporal (of 2 minutes) resolutions

    Year 3: Hydro Modelet improvement and evaluation by taking readings of shallow water gauges (deployed over a nearby backwater wetland area) as additional data points for model training and inference use

- Objective 3.1.2: Water gauge deployment and gauge data applications. This research objective is to address the following project activities, achieved mainly by **R. Day**.

  + Deploying shallow water gauges over a backwater wetland area

    Year 1: Acquiring suitable water gauges and identifying proper locations for their deployment

    Year 2: More gauge deployment, with gauge reading results to help improve predictions on floods by the nearby river/stream via developed 2-D Hydro Modelets

+ Water gauge data applications

  Year 2: Investigating the wetland storage capacity for flood mitigation based on gauge reading results

  Year 3: Investigating wetland ecological restoration after floods via gauge data for possible flow gate control between the wetland and its nearby water bodies

  Year 4: Developing regional flood forecasting plots/charts based on Hydro Modelet results

Goal 3.2: *Hydro Modelet execution acceleration.*

- Objective 3.2.1

  This execution acceleration objective deals with applying acceleration techniques devised under Research Theme 1 to developed Hydro Modelets for assessment. It is to be conducted after those acceleration techniques have been fully developed and evaluated, mostly during Year 4.

  + In-memory processing
  + Stochastic computing
  + Execution resilience
  + GPGPU execution support

Research Theme 4: *Workforce Development and Research Sustainability*

Goal 4.1: *Undergraduate and graduate student research enrichment.*

- Objective 4.1.1: Research experience for undergraduate students (aka Project Task 10 co-led by **P. Darby and N. Tzeng**)
  + Summer REU research
  + REU summer camps
- Objective 4.1.2: Co-directing theses and dissertations of graduate students

Goal 4.2: *Mentoring and collaborating research with junior investigators.*

- Objective 4.2.1: Mentoring workshops
- Objective 4.2.2: Joint research work, publications, and proposals

Goal 4.3: *Research sustainability.*