**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KỸ THUẬT MÁY TÍNH**

**NGUYỄN MẠNH THẢO**

**QUÁCH THẾ HÀO**

**KHÓA LUẬN TỐT NGHIỆP**

**THIẾT BỊ QUAN TRẮC, DỰ BÁO THỜI TIẾT PHẠM VI NHỎ SỬ DỤNG CÔNG NGHỆ MÁY HỌC**

**LOCAL WEATHER FORECASTING SYSTEM USING NEURAL NETWORK**

**KỸ SƯ/ CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH**

**TP. HỒ CHÍ MINH, 2018**

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KỸ THUẬT MÁY TÍNH**

**NGUYỄN MẠNH THẢO – 14520853**

**QUÁCH THẾ HÀO – 14521124**

**KHÓA LUẬN TỐT NGHIỆP**

**THIẾT BỊ QUAN TRẮC, DỰ BÁO THỜI TIẾT PHẠM VI NHỎ SỬ DỤNG CÔNG NGHỆ MÁY HỌC**

**LOCAL WEATHER FORECASTING SYSTEM USING NEURAL NETWORK**

**KỸ SƯ/ CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH**

**GIẢNG VIÊN HƯỚNG DẪN**

**T.S TRỊNH LÊ HUY**

**TP. HỒ CHÍ MINH, 2018**

DANH SÁCH HỘI ĐỒNG BẢO VỆ KHÓA LUẬN

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số ………….…. ngày …………………….. của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

* 1. – Chủ tịch.
  2. – Thư ký.
  3. – Ủy viên.

**MENU**

**PICTURE MENU**

**TABLE MENU**

**LIST OF ABBREVIATIONS**

**ACKNOWLEDGEMENTS**

Foremost, we would like to express our sincere thanks to teachers of University of Information Technology in general and the Faculty of Computer Engineering for imparting knowledge and valuable experience during the past 5 years to us.

We wish to send our sincere thanks to Mr. Nguyen Minh Son, Dean of the Faculty of Computer Engineering, for providing us with all the necessary facilities for the research.

We are grateful to Mr. Trinh Le Huy, lecturer, in the Faculty of Computer Engineering. We are extremely thankful and indebted to him for sharing expertise, and sincere and valuable guidance and encouragement extended to us.

We deeply thanks to the help of our friends, seniors who helped us to find information during the course.

We are also grateful to our parents for the unceasing encouragement, support, and attention.

Once again, we would like to thank all those who are interested in our thesis. However, in the process of working there may be some unavoidable mistakes, we are looking forward to receiving comments on this thesis.

Nguyen Manh Thao - Quach The Hao

**ABSTRACT**

According to the statistics of 2017, the structure of Viet Nam’s economy is as follows: agriculture, forestry, and aquaculture account for 15.34%; industrial and construction area account for 33.34%; service area accounts for 41.32%. In particular, agriculture, forestry, and aquaculture, construction play essential roles, occupying a significant share and they are trendy. Most of the models in these industries are affected by the weather, and local weather forecasts are of great benefit. Along with the development of the Internet of Things models in the industry, the demand for weather forecasts is increasing. However, there is currently no simple system to meet this need.

The current solution to this problem is using National weather forecasting channels. An advantage of this solution is that it is no need for additional hardware. It is entirely free because forecast information is published on media channels. However, there is a disadvantage that the forecast information from National weather forecasting channels is not completely correct for all local areas.

The thesis "Local weather forecasting system using neural network" focus on forecasting the weather for a small localized area using a neural network, personalizing for users, serving for family and small businesses.

The system includes a central device and other environmental monitoring devices that collect data and transmit to the central device using LoRa technology - the new wireless, long-range transmission technology. The thesis using Neural Network as the main algorithm for forecasting weather with input data collected from the environment such as temperature, humidity, atmospheric pressure, the hour in the day, month in the year, rainfall, wind direction and wind speed. The system also has a web server so that users can connect to when they have internet.

This thesis report contains five chapters:

Chapter 1: Overview

Chapter 2: Theory and Experimental Study

Chapter 3: System Analysis and Design

Chapter 4: Result and Summary

Chapter 5: Future work

Chapter 1. **OVERVIEW**

**1.1 Background**

**1.2 Related work**

**1.3 Target, object and scope of the thesis**

**1.3.1 Target and object of this thesis**

Our thesis focuses on researching and upgrading the previous thesis named "THIẾT BỊ QUAN TRẮC, DỰ BÁO THỜI TIẾT PHẠM VI NHỎ SỬ DỤNG CÔNG NGHỆ LORA" of Mr. NGUYEN TRAN TIEN ĐAT and Mr. HO QUI ĐAY. Build a system that can collect environmental data and forecast without using third-party data. Environmental monitoring devices can operate for long periods of time outside of the actual environment. Package the product so that the device can normally operate under the rain and wind.

The thesis uses Machine Learning, specifically the Back-propagation Neural Network model to solve the problem of small-scale rain forecasting. Utilizing the advantages of the Back-propagation Neural Network model is to improve predictive accuracy over time.

Applying data transfer technology - Lora in transmit and receive operations can increase the device installation distance. Besides, Lora technology can save more power than other transmit technologies. Also, thanks to solar cells, the battery life of the monitoring devices is extended, so that the system can be installed in many different areas without other power supplies.

Using Wi-Fi module in Gateway can send weather data to Linux VPS in other to predict with Machine Learning. Also, Linux VPS run a Web Server application, provides users with an easy-to-use managing website.

Overall, the biggest problem we focus on solving is that accuracy and reliability of the system must be optimized.

Detailed target:

* Build a hardware system that collects weather data independently. The hardware must be packaged in other to able to work in windy weather environment. The hardware system includes:
* Environmental monitoring devices: collect weather data such as temperature, humidity, atmosphere, wind direction, wind speed, and rainfall. Using a solar panel to maintain the battery life.
* Gateway: Receive data from Environmental monitoring devices via Lora and push them to Linux VPS via Wi-Fi. Receive weather forecasting results from the server on VPS and display them on the LCD.
* Research weather forecasting theory and study papers about weather forecasting. Analyze data collected from the environment in a long time to obtain a theoretical basis for weather forecasting.
* Research and develop a local weather forecasting application using Back-propagation Neural Network model. Compare it to the local weather forecasting application using Fuzzy Logic. Increase forecasting accuracy to 60%.
* Develop a system managing sensor data, forecasting results and user’s data with MongoDB
* Develop a Web Server on Linux VPS with an easy-to-use interface and stable operation.

Object:

* Weather Station with Anemometer/Wind vane/Rain bucket SKU:SEN0186
* Solar panel, Li-ion battery and battery charger circuit.
* Mesh-protected Weather-proof Temperature/Humidity Sensor
* Barometric Pressure/Temperature/Altitude Sensor BMP180
* Theoretical basis in weather forecasting field
* Machine Learning Back-propagation Neural Network model
* LoRa Transceiver RFM96
* Wifi Module ESP8266
* MQTT Protocol
* Web Server with NodeJS programming language
* MongoDB Database

**1.3.2 Scope**

Build an environmental monitoring and weather forecasting system with high accuracy in local areas. The system must be easy-to-install, the accuracy of forecasting results increases along with system use time.

Detail:

* Collect environmental data in 3 months at lease. Process data and use it to train the Back-propagation Neural Network model in the beginning.
* Develop a local weather forecasting application using Back-propagation Neural Network model. The application has 7 inputs for forecasting include temperature, humidity, atmosphere, the hour in the day, month in the year, wind direction and wind speed. For training operation, the application will use rainfall.
* Communicate and get data from sensors
* Build a power supply system with Li-ion batteries and use a solar panel to maintain battery life
* Communicate environmental monitoring devices and a gateway with LoRa technology
* Develop a website with HTML, CSS, and NodeJS. The website must have an easy-to-use interface and stable operation.
* Develop a system managing sensor data, forecasting results and users data with MongoDB

**1.4 Advantage and disadvantage**

**Advantage**

During the period of the thesis, we have received the enthusiastic support of our relatives, family, friends and especially thesis instructors. Besides, we also have received a lot of support from the Faculty of Computer Engineering, such as the equipment, research tools, working rooms and teachers in the department helped us in finding a lot of useful materials as well as solutions to complete the thesis. The University of Information Technology Facilities Management Department also helped a lot by permitting us to put environmental monitoring devices at the university campus. The collaboration between members of our team makes a significant contribution to the thesis success. We have a good plan for this thesis and a sense of responsibility. Also, the sharing of knowledge and experience of other research teams is also a resource for our team to complete this thesis.

**Disadvantage**

Besides the advantages mentioned above, the group has also encountered difficulties during the period of the thesis. Typically, the lack of expertise in meteorology or knowledge we found is not exactly accurate. This affects the accuracy of forecasting results. Moreover, we need to collect environmental data and build environmental monitoring devices at the same time. For this reason, environmental data collecting was interrupted at some specific time, which lead to a lack of environmental data. Some sensors we use in this thesis are costly, rare and hard to find in the domestic market. Building forecasting application along with Web Servers on University's Linux VPS also faces many difficulties due to resource constraints and the need for network access permissions.

Chapter 2. **THEORY AND EXPERIMENTAL STUDY**

**2.1 Theory**

**2.1.1 Weather forecasting**

**2.1.2 Machine Learning and Neural Network**

**2.2 Experimental Study**

**2.2.1 Components in the system**

**2.2.1.1 LoRa and LoRaWan**

LoRa, stands for Long Range Radio, is a wireless communication technology developed by Cycleo, and bought by Semtech in 2012. LoRa uses a modulation technique called Chirp Spread Spectrum. According to Semtech publication, this technique reduces the complication and accuracy requirement of receivers that need to achieve to decode data. In addition, LoRa does not require high transmit power to transmit to a receiver at a long distance cause the receiver can receive data even if received signal strength lower than ambient noise.

Working bandwidths of LoRa for regions of the world:

* 868 MHz for Europe
* 915 MHz for North America
* 433 MHz band for Asia

The technology is presented in two parts — Lora, the physical layer and LoRaWan, the upper layers. LoRaWan is a protocol based on LoRa developed by LoRa Alliance. It's used in the industrial, scientific and medical (ISM) radio bands. A LoRaWan network structure usually is a star-of-stars model. Gateways are bridges forwarding messages between end nodes and server at the backend. Gateways connect to the internet via normal IP connections. Meanwhile, end nodes wirelessly single-hop connect to one or many gateways. A single LoRaWan Gateway can cover up to 10Km. With the benefits of LoRaWan's distance, The Things Network covers Amsterdam with just 10 gateways and costs $1200.

**2.2.1.2 MQTT**

MQTT was developed by IBM and Eurotech, the latest version is MQTT 3.1.1. MQTT (Message Queuing Telemetry Transport) is a publish / subscribe protocol commonly used for Internet of Things devices with low bandwidth, high reliability and the ability to be used in unstable networks. It is based on a Broker and is designed to be open and non-specific to any application, very simple and easy to integrate. MQTT is suitable for M2M (Mobile to Mobile) applications, WSN (Wireless Sensor Networks) or IoT (Internet of Things).

There are 3 levels of QoS (Qualities of service) in MQTT:

* QoS 0 – Almost once: The broker/client sends the data once, only TCP/IP protocol confirm the successful receiving
* QoS 1 – At least once: The broker/client sends at least one successful receiving packet to the sender
* QoS 2 - Exactly once: The broker/client sends only one packet to the sender to confirm that the receiving process is successful

**2.2.1.3 Unix Socket**

Sockets enable communication between two processes in the same computer or different computer. It's a method to communicate with other processes via standard Unix file descriptors. In the Unix environment, every I/O operation is done by reading or writing to a file descriptor. A file descriptor is an integer associated with an open file such as text files, network connections, etc.

Sockets were first introduced in 2.1BSD and edited to current form in 4.2BSD. Socket feature is available to most popular UNIX system.

A Unix Socket is usually used in a server-client framework. Most protocols like FTP, SMTP and POP3 use socket to establish a connection between server and client to transmit data.

There are four types of sockets available to the users:

* Stream Sockets
* Datagram Sockets
* Raw Sockets
* Sequenced Packet Sockets

**2.2.1.4 MongoDB**

MongoDB is a cross-platform document-oriented database. It is higher performance and easier to scale than RDBMS. Documents of MongoDB have the same structure as JSON. MongoDB is first introduced by MongoDB Inc in February 2009. After that, MongoDB quickly developed into one of the most popular NoSQL databases.

The current version of MongoDB is 4.1.1 published at Git Repository: [github.com/mongodb/mongo](https://www.github.com/mongodb/mongo).

Relational concepts in RDBMS

|  |  |
| --- | --- |
| **RDBMS** | **MongoDB** |
| Database | Database |
| Table | Collection |
| Tuple/Row | Document |
| Column | Field |
| Table Join | Embedded Documents |

Main features of MongoDB

* Schema less
* High availability by cluster
* Structure of a single object is clear
* No complex joins
* Deep query-ability
* Tuning
* Ease of scale-out
* Uses internal memory for storing the (windowed) working set, enabling faster access of data

Use case

* Big Data
* Content Management

**2.2.1.5 Sensors and modules**

**Weather Station with Anemometer/Wind vane/Rain bucket SKU:SEN0186**

Equipped with a rain bucket and wind accelerometer, the SKU:SEN0186 from DFRobot is a popular solution to many weather monitoring applications at a reasonable price.

Specification

* Operating voltage: 5V
* Temperature range: -40~80℃
* Humidity range: 0~99%
* Package Dimension: 20\*18\*30 CM
* Weight: 4480g

Application

* Weather station
* Weather monitor

Data interface

Serial: 9600bps with 1s interval

Format of Data Output

35 bytes per second, including the end CR/LF

For example: c000s000g000t086r000p000h53b10020\r\n

c000： Air direction, 0 degree

s000： Air speed (1 minute), 0 miles per hour

g000： Air speed (5 minutes), 0 miles per hour

t086： Temperature, 86 Fahrenheit Degree

r000： Rainfall (1 hour), 0 inches

p000： Rainfall (24 hours), 0 inches

h53： Humidity, 53%

b10020： Atmosphere, 1002 hpa

**Temperature/Humidity Sensor SHT10**

SHT10 is designed by Sensirion The Sensor Company with a 14-bit-analog-to-digital converter and a serial interface circuit. It is a low-cost series sensor comparing to other SHT1x series.

Specification

* Series: SHT10
* Type Humidity Accuracy %: ±3
* Type Temperature Accuracy °C: ±0.4
* Supply Voltage Range V: 2.4 to 5.5
* Interface: Digital SBus
* Package Size: 7.5mm x 4.9mm x 2.6mm

The version used in this thesis is modified with protective case protecting the sensor from physical environmental factors like water, dust, etc.

**Barometric Pressure Sensor BMP180**

BMP180 is allow-cost sensing solution for measuring barometric pressure and temperature produced by Bosch. Because of pressure changes with altitude, this sensor can also be used as an altimeter.

Specifications

* Operating voltage: 1.8 ~ 3.6V
* Operating Current: 0.5uA at 1Hz
* Interface: I2C
* Pull-up resistors integrated on the I2C pins
* Max interface speed: 3.5MHz
* Low error: 0.03hPa (25cm)
* Integrated calibrate module available
* Pressure Measuring range: 300hPa ~ 1100hPa (+9000m to -500m)
* Weight: 1.18g
* Package size: 21mm x 18mm

**Ultra-Long-Range Transceiver RFM95**

The LoRa transceiver module RFM95 produced by HopeRF provides ultra-long-range communication with minimizing current consumption. Using the Chirp Spread Spectrum modulation technique, the RFM95 can reduce the sensitivity to ~ -148dBm, roughly equal to the ambient noise level. Combined with the 20dBm Power Amplifier, the communication distance can reach up to 15km.

Specifications

* LoRa™ Modem
* 168 dB maximum link budget
* +20 dBm - 100 mW constant RF output vs V supply
* +14 dBm high efficiency PA
* Programmable bit rates up to 300 kbps
* High sensitivity: down to -148 dBm
* Bullet-proof front end: IIP3 = -125 dBm
* Excellent blocking immunity
* Low RX current of 103 mA, 200 mA register retention
* Fully integrated synthesizer with a resolution of 61 Hz
* FSK, GFSK, MSK, GMSK, LoRa™ and OOK modulation
* Built-in bit synchronizer for clock recovery
* Preamble detection
* 127 dB Dynamic Range RSSI
* Automatic RF Sense and CAD with ultra-fast AFC
* Packet engine up to 256 bytes with CRC
* Built-in temperature sensor and low battery indicator
* Package Size： 16\*16mm

Applications

* Automated Meter Reading
* Home and Building Automation
* Wireless Alarm and Security Systems
* Industrial Monitoring and Control
* Long range Irrigation Systems

**Wi-Fi Module ESP8266-V12**

The ESP8266-V12 is a low-cost Wi-Fi module which is highly rated for Internet and Wi-Fi applications as well as transmission applications. The ESP8266 is an integrated chip designed for the needs of the Internet of Things (IoT). It provides a complete and closed Wi-Fi network solution that allows it to host applications or to reduce the load on all Wi-Fi network connections from the controller. The ESP8266 has powerful processing and storage capabilities that allow it to be integrated with other sensors, microcontrollers and application devices via GPIOs.

Specifications

* Producer: Espressif
* Certification: Wi-Fi Alliance
* Protocols: 802.11 b/g/n
* Frequency Range: 2.4G ~ 2.5G
* CPU: Tensilica L106 32-bit processor
* Peripheral Interface: UART/SDIO/SPI/I2C
* Operating Voltage 2.5V ~ 3.6V
* Operating Current: ~80 mA
* Operating Temperature Range: –40°C ~ 125°C

**2.2.2 IDEs, software, and tools**

**2.2.2.1 Git and GitHub**

Git is a version control system developed by Linus Torvalds in 2005, with the purpose of developing Linux kernel at the beginning. Nowadays, Git becomes the most popular version control system. Git is an open source application with the GPL2 license.

Designed by Linus Torvalds

Developed by Junio Hamano and others

First publish in April 7th, 2005

Repository <https://git.kernel.org/pub/scm/git/git.git/>

Platform Linux, Windows, OS X

Homepage [git-scm.com](https://git-scm.com/)

GitHub is a service that provides web-based Git source code repositories for software development projects. GitHub offers both paid and free versions for accounts. Open source projects will be offered free storage. As of June 2018, GitHub has more than 40.1 million users, making it the largest source code server in the world. GitHub has become a trend in the open source development community. Even many developers have begun to consider Git as a substitute for resumes and some employers require applicants to provide a link to the GitHub account to evaluate candidates. Formally acquired by Microsoft on June 4th, 2018 for $ 7.5 billion.

Founded in February 8th, 2008

Headquarters in San Francisco, California, U.S

Founders: Tom Preston-Werner, Chris Wanstrath, PJ Hyett

Our Git repository is at <https://github.com/nguyenmanhthao996tn/ktln>

The repository has 3 main folders:

* **Documents**: Contains documents like related papers that relate to this thesis, report, testing results, cost, etc.
* **PCBs**: Contains all hardware designs.
* **SourceCode**: Contains source code of server and firmware of other components

**2.2.2.2 Slack**

Slack is a chat platform or a set of proprietary team collaboration tools and services. First introduced in August 2013, Slack is developed by Slack Technologies and its original author is Stewart Butterfield. Slack's users can be able to join conversations on different channels. Moreover, Slack provides many features, integration tools, and plugins. Typically, Git notifications are integrated through GitHub's Hook feature. Besides, it's free for everyone.

**2.2.2.3 VirtualBox**

VirtualBox is a free, open-source powerful x86 and AMD64/Intel64 virtualization software. It supports on many platforms such as Windows, Linux, MacOS. VirtualBox is compatible with most 32-bit and 64-bit operating systems but not require additional virtualized hardware. In additional, VirtualBox provides a powerful Guest Additions mode that helps user can interact to the virtual machine much easier.

Original author: Innotek GmbH

Developer: Oracle Corporation

Initial release 17 January 2007

Repository: <https://www.virtualbox.org/browser/vbox/trunk>

Written in: C, C++, x86 Assembly

Operating system: Windows, macOS, Linux

License: Base Package: GNU General Public License version 2

**2.2.2.4 Visual Studio Code**

Visual Studio Code is a combination of a text editor and supporting tools that developers need such as an integrated terminal, Git client, plugins, code formatting feature, etc.

Visual Studio Code is a free, open source, cross-platform. Developed by Microsoft and first published in April 29th, 2015. It's repository: [github.com/Microsoft/vscode](https://github.com/Microsoft/vscode)

Chapter 3. **SYSTEM ANALYSIS AND DESIGN**

**3.1 System overview**

**3.2 Hardware design**

**3.3 Software design**

Chapter 4. **RESULT AND SUMMARY**

**4.1 Result**

**4.2 Summary**

Chapter 5. **FUTURE WORK**

**REFERENCES**

**English**

[1] Saktaya Suksri and Warangkhana Kimpan, "Neural Network training model for weather forecasting using Fireworks Algorithm", 2016 International Computer Science and Engineering Conference (ICSEC). [Online]. Available: <https://ieeexplore.ieee.org/document/7859952/>. [Accessed Jul. 14, 2018].

[2] Andrew Culclasure, "Using Neural Networks to Provide Local Weather Forecasts", Electronic Theses & Dissertations Georgia Southern University Spring 2013. [Online]. Available: <https://digitalcommons.georgiasouthern.edu/etd/32/>. [Accessed Jul. 14, 2018].

[3] Kumar Abhishek, "Weather Forecasting Model using Artificial Neural Network", Procedia Technology Volume 4, 2012, Pages 311-318. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S221201731200326X>. [Accessed Jul. 14, 2018].

**Vietnamese**

[1] Đoàn Văn Điếm, Giáo trình Khí tượng Nông nghiệp, Hà Nội: Trường đại học Nông nghiệp 1, 2008.

[2] Bộ văn hóa, thể thao và du lịch, Tổng cục du lịch, 2015. [http://www.vietnamtourism.com/index.php/about/items/2532, 2015](http://www.vietnamtourism.com/index.php/about/items/2532,%202015).

[3] Tổng cục Thống kê, <http://www.gso.gov.vn/SLTK/Chart.aspx?rxid=4227aeb8-dadb-4081-928c-0777afbce2b9&px_db=01.+%C4%90%C6%A1n+v%E1%BB%8B+h%C3%A0nh+ch%C3%ADnh%2c+%C4%91%E1%BA%A5t+%C4%91ai+v%C3%A0+kh%C3%AD+h%E1%BA%ADu&px_type=PX&px_language=vi&px_tableid=01.+%C4%90%C6%A1n+v%E1%BB%8B+h%C3%A0nh+ch%C3%ADnh%2c+%C4%91%E1%BA%A5t+%C4%91ai+v%C3%A0+kh%C3%AD+h%E1%BA%ADu%5cV01.09.px&layout=chartViewColumn>, 2015

[4] Trần Tân Tiến, Phương pháp số dự báo thời tiết, Hà Nội: Đại học Quốc gia Hà Nội, 2007.