SLIDING WINDOW TECHNIQUE FOR FOREST FIRE PREDICTION

This thesis is submitted to the Division of Applied Sciences, College of Arts and Sciences in partial fulfilment of the requirements for the degree of Master of Science (Information Technology)
Universiti Utara Malaysia

By

Khor Jia Yun

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

KHOR JIA YUN
(89407)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/ her project paper of the following title)

SLIDING WINDOW TECHNIQUE FOR FOREST FIRE PREDICTION

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **PROF. DR. KU RUHANA KU MAHAMUD**

Tandatangan
(Signature) : _____

Tarikh
(Date) : 11 November 2008

PERMISION TO USE

This thesis presents a partial fulfilment of the requirement for a postgraduate degree from Universiti Utara Malaysia. I agree that the university library may make it freely available for inspection. I further agree that the permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or, in their absence by the Assistant Vice Chancellor of the College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

Assistant Vice Chancellor
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman.

ABSTRAK

Kebakaran hutan di Portugal setiap tahun mengakibatkan kemusnahan kawasan tanah yang luas serta kematian. Dalam penyelidikan ini, kaedah penemuan corak digunakan untuk mencari corak kebakaran yang mengaitkan di antara keadaan meteorologi dan saiz kebakaran hutan. Faktor meteorologi yang disiasat merangkumi suhu, kelembapan udara, kelajuan angin dan hujan. Teknik *sliding window* digunakan untuk mencari corak kebakaran untuk keadaan cuaca yang mempunyai kaitan dengan kejadian kebakaran hutan. Set data awal diolah dengan menukar nilai selanjar ke nilai kategori. Teknik *sliding window'* kemudiannya digunakan untuk mencari corak kebakaran. Corak yang ditemui dikumpulkan mengikut saiz kawasan kebakaran. Dapatan kajian menunjukkan lapan *rules* yang boleh digunakan untuk menjangkakan saiz kebakaran hutan. Di samping itu, penyelidikan ini menunjukkan bahawa teknik *sliding window* juga boleh digunakan bagi set data yang tidak berupa *temporal*.

ABSTRACT

Every year, forest fire in Portugal causes large areas of land being destroyed and there are cases of death. In this research pattern discovery is being used to generate patterns of meteorological conditions in relation to area burnt of forest fire. The meteorological conditions that are being investigated are temperature, relative humidity, wind speed and rainfall. The combination of these four conditions forms the patterns that are of interest in this research. The sliding window technique is being used to generate patterns for meteorological conditions that are significant to forest fire. The initial dataset is being transformed by changing the continuous values of the attributes into categorical values of the attributes. The patterns are then being generated through the sliding window methodology. Patterns that could not be validated are being regarded as invalid and thus are discarded while the patterns that could be validated are taken for further analysis. Patterns that are valid are then being grouped based on the burnt area associated with a pattern. The rules are then generated by transforming the categorical values into intervals and the merging of different records into the same rules. The rule generation stage produces eight distinct patterns of meteorological conditions that could predict the size of forest fire. In addition, this study showed that the sliding window technique could be used in non-temporal data.

ACKNOWLEDGEMENTS

Now, my most profound thankfulness goes to my final project supervisor Professor Dr. Ku Ruhana Ku Mahamud for her help, guidance and encouragement. I would also like to thank her continuous faith and support in me.

Secondly, I wish to thank all my dearest family members, especially to my parents, brother and sister their support for me to complete this study. Next I would like to thank all the lecturers who have taught me before throughout the Masters Degree course because the knowledge they have imparted have allowed me to be more knowledgeable and thus in a better position to complete this research.

In addition, I would like to thank all my friends who had given me emotional support and taken care of me at times of difficulties.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBA	Automated Biomass Burning Algorithm
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BIRD	Bi-Spectral InfraRed Detection
BUI	Buildup Index
CFFDRS	Canadian Forest Fire Danger Rating System
DC	Drought Code
DM	Data Mining
DMC	Duff Moisture Code
DNA	Deoxyribonucleic Acid
ERS	European Earth Resource Satellite
FARSITE	Fire Area Simulator
FFI	Finnish Forest Fire Risk Index
FFMC	Fine Fuel Moisture Code
FWI	Fire Weather Index
GIS	Geographic Information Systems
GOES	Geostationary Operational Environmental Satellite
HMM	Hidden Markov Model
ICRIF	Indice Combinado de Risco de Incendio Florestal (Forest Fire Risk Index)
ISI	Initial Spread Index
MODIS	The Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation

NDVI	Normalized Difference Vegetation Index
RADARSAT	Radar Satellite
SAR	Synthetic Aperture Radar
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

Forest fire plays a crucial role in shaping forest ecosystems all over the world. In Mountain Ash forests in Australia, for example, the regeneration of the ecosystems depends on fire. However, other forest ecosystems such as tropical lowland and peat forest are most likely to be seriously damaged by fire. There is increasing evidence to show that the link between climate change and El Nino phenomenon is causing an escalation in number and size of forest fire. According to new evidence from Amazon, tropical forests that have experienced burning before would be more susceptible to future burning. Thus, there is an increased possibility that wildfire episodes will occur more frequently, and in the magnitude not endurable by the tropical forest ecosystem. Scientists believed that the entire Amazon would be threatened, and the consequence affects the biodiversity and climate change globally (Rowell & Moore, 2000).

During the El Niño phenomenon in the late 1997 and early 1998, the major forests throughout the world were experiencing burning, causing severe damages to the environment as well as economy. It was a time where South East Asia, South and Central America, Europe, Russia and China were raged with fire. In South East Asia, hundreds of thousands of hectares of forest and other lands were damaged, costing US$5-10 billion economically and 70 million of the population's health were adversely affected. Moreover, forests all over the world burnt in 1999 have yet to recover, thus indicating that the impact of forest fire on ecology, culture, social and economy is immense (Rowell & Moore, 2000). In addition, a severe drought

The contents of the thesis is for internal user only

REFERENCES

Aicep. (2008). *Portugal Global: Portugal Country Profile*. Retrieved September 4, 2008, from http://www.portugalglobal.pt/CmsAPI/AICEP/portugalcountryprofilebookmarks_ing.pdf

Bugalho, L. & Pessanha, L. (2007). The Forest Fire Risk Index (ICRIF) Operational Processing and Validation. In *4th International Wildland Fire Conference*, Seville, SPAIN, 13-17 May 2007. www.fire.uni-freiburg.de/sevilla-2007/contributions/doc/SESIONES_TEMATICAS/ST4/

Carvalho, A., Flannigan, M. D., Logan, K., Miranda, A. I., & Borrego, C. (2008). Fire activity in Portugal and its relationship to weather and the Canadian Fire Weather Index System. *International Journal of Wildland Fire, 17*, 328-338.

Chen, K., Kurgan, L. & Ruan, J. (2006). Optimization of the Sliding Window Size for Protein Structure Prediction. *IEEE symposium on Computational Intelligence and Bioinformatics and Computational Biology,* 1-7.

Cheng, H., Yan, X., Han, J. & Hsu, C. W. (2007). Discriminative Frequent Pattern Analysis for Effective Classification. *IEEE 23rd International Conference on Data Engineering* (pp. 716-725). Istanbul.

Cheng, H., Yan, X., Han, J. & Yu, P. S. (2008). Direct Discriminative Pattern Mining for Effective Classification. *IEEE 24th International Conference on Data Engineering* (pp. 169-178). Cancun.

Chi, Y., Wang, H., Yu, P. S., & Muntz, R. R. (2004). Moment: Maintaining Closed Frequent Itemsets over a Stream Sliding Window. *Fourth IEEE International Conference on Data Mining* (pp. 59-66). Brighton, UK.

Cortez, P. & Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. *13th Portugese Conference on Artificial Intelligence, 3-7 Dec 2007.*

Darmawan, M., Aniya, M. & Tsuyuki, S. (2001). Forest Fire Hazard Model Using Remote Sensing and Geographic Information Systems: Toward understanding of Land and Forest Degradation in Lowland areas of East Kalimantan, *Indonesia, 22nd Asian Conference on Remote sensing, 5-9 November, 2001, Singapore.*

De Saint Vincent, A. R., Grazzini, F., Cruz, J. M., Valcance, F., & Martinez, S. (2005). A Comprehensive Service Portfolio for Forest Fire Risk Management Within the Context Of ESA/GMES and EC/Eurorisk: Present and Future Of Risk

Management Services in Europe. In *Proceedings of 2nd International Conference of Recent Advances in Space Technologies, RAST* ,113-117.

Elfeky, M. G., Aref, W. G. & Elmagarmid, A. K. (2006). STAGGER: Periodicity Mining of Data Streams using Expanding Sliding Windows. *Proceedings of the Sixth International Conference on Data Mining*, 188-199.

Europe [Map]. (2008). *Wikipedia: The Free Encyclopedia*. Retrieved September 4, 2008, from http://en.wikipedia.org/wiki/Europe

Fiorucci, P., Gaetani, F., Minciardi, R., Scipioni, A. (2007). Nation-wide forest fire risk assessment: Performance analysis and validation of two different dynamic fire weather systems in Italy. In *4th International Wildland Fire Conference*, Seville, SPAIN, 13-17 May 2007. www.**fire**.uni-freiburg.de/sevilla-2007/contributions/doc/SESIONES_TEMATICAS/ST4/

Gillespie, T., Chu, J., Frankenberg, E. & Thomas, D. (2007). Assessment and Prediction of Natural Hazards from Satellite Imagery. *Progress in Physical Geography, 31*(5), 459-470.

Ho, C. C., Li, H. F., Kuo, F. F. & Lee, S. Y. (2006). Incremental Mining of Sequential Patterns over a Stream Sliding Window. *Sixth IEEE International Conference on Data Mining Workshops*, 677-681.

Huang, Y. P., Kao, L. J. & Sandnes F. E. (2008). Identifying Insightful Salinity and Temperature Variations in Ocean Data. *8th IEEE International Conference on Computer and Information Technology*, 191-196.

Jin, L., Chai, D. J., Lee, Y. K. & Ryu, K. H. (2007). Mining Frequent Itemsets over Data Streams with Multiple Time-Sensitive Sliding Windows. *Sixth International Conference on Advanced Language Processing and Web Information Technology*, 486-491.

Jin, X., Zuo, X., Lam, K. Y., Wang, J. & Sun, J. (2006). Efficient Discovery of Emerging Frequent Patterns in Arbitrary Windows on Data Streams. *Proceedings of the 22nd International Conference on Data Engineering*, 113-113.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2001). An Online Algorithm for Segmenting Time Series. In *Proceedings of IEEE International Conference on Data Mining*, 289-296.

Kontaki, M., Papadopoulos, A. N. & Manolopoulos, Y. (2006). Efficient Incremental Subspace Clustering in Data Streams. *10th International Database Engineering and Applications Symposium*, 53-60.

Kottlowski, D. (2006). *Humidity*. Retrieved August 30, 2008, from AllExperts, Accuweather Web site: http://en.allexperts.com/q/Accuweather-2546/Humidity-2.htm

Laneve, G. & Cadau, E. G. (2007). Quality assessment of the fire hazard forecast based on a fire potential index for the Mediterranean area by using a MSG/SEVIRI based fire detection system. *IEEE International Geoscience and remote sensing symposium*, 2447-2450.

LaRosa, C., Xiong, L. & Mandelberg, K. (2008). Frequent pattern mining for kernel trace data. *Proceedings of the 2008 ACM symposium on Applied computing,* (pp. 880-885), Fortaleza, Ceara, Brazil.

Li, H. & Chen, H. (2008). DSOSW: A Deleting Strategy in Mining Frequent Itemsets over Sliding Window of Stream. *2008 International Symposiums on Information Processing*, 135-138.

Li, H. F., Ho, C. C., Kuo, F. F. & Lee S. Y. (2006). A New Algorithm for Maintaining Closed Frequent Itemsets in Data Streams by Incremental Updates. *Sixth IEEE International Conference on Data Mining Workshops*, 672-676.

Li, H. F., Ho, C. C., Shan, M. K. & Lee, S. Y. (2006). Efficient Maintenance and Mining of Frequent Itemsets over Online Data Streams with a Sliding Window. *IEEE International Conference on Systems, Man, and Cybernetics, Taiwan, 3*, 2671-2677.

Liang, Q. A., Chung, J.Y., Miller, S. & Ouyang, Y. (2006). Service Pattern Discovery of Web Service Mining in Web Service Registry-Repository. *IEEE International Conference on e-Business Engineering*, 286-293.

Lloyd, N. (n.d.). *Forest Fire in Spain*, Retrieved September 1, 2008, from IberiaNature Web Site: http://www.iberianature.com/material/fire.html

METSERVICE. (n.d.). *FAQ: Weather Terminology*, Retrieved August 26, 2008, from http://www.metservice.co.nz/default/help.php?subject=faq-terminology

Orozco, S., J. (2008). *Forest Fire Risk Model for Michoacan, Mexico*. Unpublished M.Sc. thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Park, A. S. & Glass, J. R. (2008). Unsupervised Pattern Discovery in Speech. *IEEE Transactions on Audio, Speech, and language processing, 16(1)*, 186-197.

Pearce, V.(2008). *Windows to the Universe: Classification of Wind Speeds*. Retrieved August 25, 2008, from the University Corporation for Atmospheric Research (UCAR) Website:

http://www.windows.ucar.edu/tour/link=/earth/Atmosphere/wind_speeds.html&edu=high

Pedrosa, F. T. & Herrman, M. L. (2007). Forest Fires in Protected Areas in Portugal and in the State of Santa Catarina (Brazil): Comparative Analysis of Policies of Prevention and Recovery of Burned Areas. In *4th International Wildland Fire Conference*, SPAIN.

Peng, W., Li, T. & Narasimhan, G. (2006). Mining the Database of Transcription Binding Sites. *Sixth IEEE Symposium on Bioinformatics and BioEngineering* (pp. 61-64).

Pereira, M. G., Trigo, R. M., da Camara, C. C., Pereira, J. M. C., Leite, S. M. (2005). Synoptic patterns associated with large summer forest fires in Portugal. *Agricultural and Forest Meteorology, 129*, 11-25.

Portugal [Map]. (2008). *Google Maps*. Retrieved September 4, 2008, from http://maps.google.com/maps?sourceid=navclient&ie=UTF-8&rlz=1T4RNWN_enMY278MY279&q=portugal&oe=UTF-8&um=1&sa=N&tab=wl&oi=property_suggestions&resnum=0&ct=property-revision&cd=1

Rainha, M. & Fernandes, P. M. (2002). Using the Canadian Fire Weather Index (FWI) in the Natural Park of Montesinho, NE Portugal: calibration and Application to fire management. In *Forest Fire Research & Wildland Fire Safety, Viegas*.

Ren, J. & Li, K. (2007). Find Recent Frequent Items with Sliding Windows in Data Streams. *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing IIHMSP*, 2, 625-628.

Rowell, A. & Moore, D. P. F. (2000). Global Review of Forest Fires. *International Union for Conservation of Nature and Natural Resources*.

Soja, A., Al-Saadi, J., Pierce, B., Kittaka, C., Szykman, J., Williams *et al.* (2006). How well does satellite data quantify fire and enhance biomass burning emissions estimates? *15th International Emission Inventory Conference: Reinventing Inventories – New Ideas in New Orleans*, May 15-18, New Orleans.

Su, T. H., Zhang, T. W. & Qiu, Z. W. (2007). HMM-Based System for Transcribing Chinese Handwriting. *International Conference on Machine Learning and Cybernetics*, 6, 3412-3417.

Tanskanen, H. & Venalainen, A. (2008). The relationship between fire activity and fire weather indices at different stages of the growing season in Finland. *Boreal Environment Research, 13*, 285-302.

Taylor, S. W. & Alexander, M. E. (2006). Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire, 15*, 121-135.

Umamaheshwaran, R., Bijker, W., & Stein, A. (2007). Image Mining for Modeling of Forest Fires From Meteosat Images, *IEEE Transactions on Geoscience and Remote Sensing, 45(1)*, 256-253

Voight, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite Image Analysis for Disaster and Crisis-Management Support. *IEEE Transactions on Geoscience and Remote Sensing, 45*(6), 1520-1528.

Wang, W., Zhao, H., Chen, J., Chen, J. & Xi, G. (2008). Bridge the gap between syndrome in Traditional Chinese Medicine and proteome in western medicine by unsupervised pattern discovery algorithm. *IEEE International Conference on networking, Sensing and Control, ICNSC,* 745-750.

Wells, G., Rohling, K., Kapus, J., & Swedberg, T. (2007). The Fire-Climate Connection. *Fire Science Digest: Research Supporting Sound Decisions.*

Wikipedia. (2008). *Rain*. Retrieved August 25, 2008, from http://en.wikipedia.org/wiki/Rain

Wong, A. K. C. & Li, G. C. L. (2008). Simultaneous Pattern and Data Clustering for Pattern Cluster Analysis. *IEEE Transactions on Knowledge and Data Engineering. 20*(7), 911-923.

Yan, R., Jurisica, I., Boutros, P. C. & Penn. L. Z. (2007). Comparison of Machine Learning and Pattern Discovery Algorithms for the Prediction of Human Single Nucleotide Polymorphisms. *IEEE International Conference on Granular Computing*, 452-452.