

REFERENCES

- Affymetrix. Affymetrix genome-wide human SNP Nsp/Sry Assay5.0 [online]. 2007 [cited 2011 Oct 5]. Available from: URL: http://media.affymetrix.com/support/downloads/manuals/genome_wide_snp_5_0_manual.pdf.
- Anyanwu MN, Shiva, SG. Comparative analysis of Serial decision tree classification algorithms. *International Journal of Computer Science and Security*. 2009; 3(3): 230–40.
- Bakir B, Batmaz Í, Güntürkün FA, Ipekci ÍA, Köksal G, Özdemirel NE. Defect cause modeling with decision tree and regression analysis. *World Acad Sci Eng Technol*. 2006; 24: 1–4.
- Bekkerman R, El-Yaniv R, Tishby N, Winter Y. Distributional word clusters vs. words for text categorization. *JMLR* 2003; 3: 1183–208.
- Besaggio D, Fuselli S, Srikumool M, Kampuansai J, Castri L, Tyler-Smith C, et al. Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evol Biol*. 2007 Aug 16; 7(Suppl 2): 1-10.
- Boontang, N. Geographical region separation. The Royal Institution. 2003 [cited 2011 June 2]. Available from: URL: <http://www.royin.go.th/th/knowledge/detail.php?ID=1378> [5].
- Caruana R and de Sa V. Benefitting from the variables that variable selection discards. *JMLR* 2003; 3: 1245–64.

Cavalli-Sforza LL, Edwards AW. Phylogenetic analysis models and estimation procedures. *Am. J. Hum. Genet.* 1967 May; 19(3 Pt 1): 233–57.

Christopher DM, Prabhakar R, Hinrich S. Introduction to information retrieval [online]. 2008 [cited 2011, January 10]. Available from: URL: <http://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>.

Condominas G. From Lawa to Mon, from Saa' to Thai. The Australian national university, Canberra. 1990.

Department of provincial administration, Ministry of Interior. Announcement of the central bureau of department of provincial administration in the number of Thai population divided by Bangkok and the other provinces from civil registry evidence on December 31st, 2011 [online]. 2012 [cited 2012, Jan 20]. Available from: URL: http://stat.bora.dopa.go.th/stat/pk/pk54/pk_54.pdf.

Dhillon I, Mallela S, and Kumar R. A divisive information-theoretic feature clustering algorithm for text classification. *JMLR* 2003; 3: 1265–87

Dınc H. ALU Insertion polymorphisms in Anatolian Turks [M.S. Thesis]. Çankaya Ankara :The Middle East Technical University; 2003.

Dorak T. Basic population genetics [online]. 2011 [cited 2011, July 5]. Available from: URL: <http://www.dorak.info/genetics/popgen.html>

Duerinck KF. SNP-single nucleotide polymorphism [online]. 2001 Feb 12 [cited 2011 Jun 20]. Available from: URL: <http://www.duerinck.com/>

Efron B, Halloran E, Holmes S. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. U.S.A.* 1996 Nov 12; 93(23): 13429–34.

- Ender P. Applied categorical and nonnormal data analysis [online]. 2005 [cited 2011 Jun 28]. Available from: URL: <http://www.philender.com/courses/categorical/intro.html>
- Eriksen TH. Ethnicity and nationalism: anthropological perspectives. Pluto Press 1993.
- Everitt BS, Hothorn T. A Handbook of Statistical Analyses Using R [online]. 2009 Jul 20 [cited 2012 March 27]. Available from: URL: http://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_multidimensional_scaling.pdf.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; 39: 783–91.
- Forman G. An extensive empirical study of feature selection metrics for text classification. *JMLR* 2003; 3: 1289–306.
- Fuerle R.D. The evolution of modern humans [Online]. 2008 [cited 2011 Jul 23]. Available from: URL: <http://erectuswalksamongst.us/Chap7.html>.
- Greenacre MJ, Degos L. Correspondence analysis of HLA gene frequency data from 124 population samples. *Am J Hum Genet* 1977; 29: 60-75.
- Greenacre MJ. Theory and applications of correspondence analysis. London. Academic Press. 1984.
- Guyon I, Elisseeff A. An Introduction to variable and feature selection. *JMLR* 2003; 3: 1157–82.
- Illumina. Genome-wide DNA analysis bead chips data sheet. California: San Diego; 2009.
- Jaroenwong K. Lua in Lan Na. 1991 March 7-8th; Chiang Mai, Thailand. Department of Humanity and Sociology, Chiang Mai Teacher collage; 1991.

Kalinowski ST. Evolutionary and statistical properties of three genetic distances.

Mol. Ecol. 2002 Aug; 11(8): 1263–73.

Kampuansai J, Bertorelle G, Castri L, Nakbunlung S, Seielstad M, Kangwanpong D.

Mitochondrial DNA variation of Tai speaking peoples in northern Thailand.

ScienceAsia. 2007; 33(4): 443-8.

Kampuansai J. X-,Y-Chromosomal and mitochondrial DNA variations among Tai

speaking peoples of northern Thailand [Ph.D Thesis]. Chiang Mai: Chiang

Mai University; 2008.

Kim D-H, Uhm S, Ko Y-W, Cho SW, Cheong JY, Kim J. Chronic hepatitis and

cirrhosis classification using SNP data, Decision Tree and Decision Rule.

ICCSA [online]. 2007 [cited 2011 Oct 12]; 4707: 585–96. Available from:

URL: <http://www.springerlink.com/content/wp3703043007655r/>.

Kohavi R, Quinlan R. Decision tree discovery [online]. 1999 [cited 2011 Jun 15].

Available from: URL: <http://ai.stanford.edu/~ronnyk/treesHB.pdf>.

Kutanan W, Kampuansai J, Nakbunlung S, Lertvicha P, Seielstad M, Bertorelle G,

et al. Genetic structure of Khon Mueang populations along a historical

Yuan migration route in northern Thailand. Chiang Mai International Journal

of Science and Technology 2010; 38(2): 295 – 305.

Kutanan W, Kampuansai J, Fuselli S, Nakbunlung S, Seielstad M, Bertorelle G et al.

Genetic structure of the Mon-Khmer speaking groups and their affinity to the

neighbouring Tai populations in Northern Thailand. BMC Genetics. 2011 Jun

15;12(1):56.

Kwak N, Chong-Ho Choi. Input feature selection by mutual information based on Parzen window. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002 Dec; 24: 1667–71.

Kwok PY. Single nucleotide polymorphisms methods and protocols. New Jersey: Human Press; 2003.

Laakso M. Computational identification of recessive mutations in cancers using high throughput SNP-arrays [online]. 2007 [cited 2011 Oct 5]. Available from: URL: <http://www.doria.fi/bitstream/handle/10024/4008/computat.pdf?sequence=1>.

LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res.* 2009 Jul; 37(13): 4181–93.

Laoongpliw W. The border people and area building in social of immigrant: in case study of Dara Aung in Chiang Dao district [Master degree of Art]. Chiang Mai: Chiang Mai University; 2003.

Latter BD. Selection in finite populations with multiple alleles. 3. genetic divergence with centripetal selection and mutation. *Genetics.* 1972 Mar; 70(3): 475–90. (31)

Lee BL. Chapter 11 correspondence analysis [online]. 2007 [cited 2011 Jun 12]. Available from: URL: <http://www.uv.es/visualstats/vista-frames/pdf/chap11.pdf>.

Leepreecha, P. The Lao Hmong in Thailand: state policies and operations (1975-2009) [online]. 2011[cited 2011 Oct 25]. Available from: URL: http://arcmtailand.com/publications_detail_view.php?id=80.

Leepreecha P. Hmong various life from mountain into city. Chiang Mai: Ming Muang publishing; 2005.

Malasam S. The migration and settlement of Yong people in Lamphun. Bangkok: Thammasart university publisher; 1997.

Malasam S. Population mobility across states in Lan Na history. Proceedings of the National Demography Conference; 2010 Nov 25-26; Bangkok, Thailand.

McDonald D, Pauli J. Population genetics glossary population ecology, ZOO4400/5400 [online]. 2009 [cited 2011 July 7]. Available from: URL: <http://www.uwyo.edu/dbmcd/popecol/Maylects/PopGenGloss.html> .

Ministry of Social Development and Human Security. Directory of high land community 20 provinces in Thailand 2545 A.D. Bangkok: 2003.

Munwalee. Documentary of Phi Tong Luang. Bangkok: Ruamsarn publisher; 1984.

Nei M. Genetic Distance between Populations. The American Naturalist. 1972; 106(949): 283–92

Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;76: 379-390.

Nei M. Molecular evolutionary genetics. New York: Columbia university press. 1987.

Nei M. Phylogenetic analysis in molecular evolutionary genetics. Annu. Rev. Genet. 1996; 30: 371–403.

Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. gene frequency data. J. Mol. Evol. 1983; 19(2): 153–70.

Nei M, Roychoudhury AK. Evolutionary relationships of human populations on a global scale. Mol. Biol. Evol. 1993 Sep; 10(5): 927–43.

- Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford: Oxford university press; 2000.
- Nenadić O, Greenacre M. Correspondence analysis in R, with two- and three-dimensional graphic: The ca package. Journal of Statistical software 2007 May; 20(3): 1–13.
- Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG., Stengard J, et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat. Genet 1998; 19: 233–240.
- Ongsakul S. The era of domains and city-states before the beginning of the Lan Na Kingdom. In: history of Lan Na. 1st ed. Chiang Mai: Silkworm Books; 2005.
- Park J, Hwang S, Lee YS, Kim S-C, Lee D. SNP@Ethnos: a database of ethnically variant single-nucleotide polymorphisms. Nucleic Acids Res. 2007 Jan; 35: D711–5.
- Penth H. A brief history of Lanna : civilization of north Thailand. Chiang Mai: Silkworm; 2000.
- Pookajarn S. The Phi Tong Luang (Mlabri): A hunter-gatherer group in Thailand. Bangkok: Odeon store; 1992.
- Quinlan JR. C4.5: programs for machine learning. Morgan Kaufmann; 1993.
- Quinlan JR, Rivest RL. Infering decision tree using the minimum description length principle. Information and Comutation. 1989; 80: 227–48.
- Rieder MJ, Taylor S L, Clark AG, Nickerson DA. Sequence variation in the human angiotensin converting enzyme. Nat.Genet 1999; 22: 59–62

Rokach L, Maimon O. Data mining with decision trees. Singapore: World Scientific Publishing; 2008.

Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern. 1991 Jun; 21(3): 660–74.

Sailer S. Pedigree collapse' due to inbreeding: iSteve Blog,[online]. 2006 Mar [cited 2011 Jul 25]. Available from: URL: <http://isteve.blogspot.com/2006/03/pedigree-collapse-due-to-inbreeding.html>.

Sampson JN, Kidd KK, Kidd JR, Zhao H. Selecting SNPs to identify ancestry. Annals of Human Genetics. 2011 Jul 1; 75(4): 539–53.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 1987 Jul; 4(4): 406–25.

Schliesinger J. Tai group of Thailand, Volume 1: Introduction and overview. Bangkok: White Lotus Press; 2000.

Sneath PHA, Sokal RR. Numerical Taxonomy. San Francisco: W.H. freeman and company;1973.

Srikummool M. X-, Y-chromosome and mitochondrial DNA variations of the Karen, Hmong and Iu Mien in the upper northern part of Thailand [Ph. D Thesis]. Chiang Mai University, Chiang Mai, Thailand. 2005.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 2011; 28: 2731-9.

Tan P, Steinbach M, Kumar V. Introduction to data mining: chapter 4 classification: basic concepts, decision trees, and model evaluation [online]. 2006 [cited 2011 Oct 10]. Available from: URL: <http://www.users.cs.umn.edu/~kumar/dmbook/index.php>.

Thearling K. An introduction to data mining [online]. 2010 [cited 2011 Sep 23] Available from: URL: <http://www.thearling.com/text/dmwhite/dmwhite.htm>.

The Association of Hilltribe Cultural Studies of Thailand et al. Knowledge and folk wisdom of the Mien (Yao). 2002.

The HUGO Pan-Asian SNP Consortium et al. Mapping Human Genetic Diversity in Asia. Science. 2009; 326: 1541–5.

Thianpanya P, Luangtip P, Sripangnen S, Damsaard P. The right of Mon community in case of gas building effect on the native and Sangklaburi Mon community district . Bangkok: Nititam Publisher; 2004.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, et al. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 1996; 271: 1380–1387.

Torkkola K. Feature extraction by non-parametric mutual information maximization. JMLR 2003; 3: 1415–38.

Vignal A, Milan D, Sancristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 2002; 34: 275–305.

Weir BS, Hill WG. Estimating F-statistics. Annu. Rev. Genet. 2002; 36: 721–50.

Weston J, Elisseeff A, Schoelkopf B, and M. Tipping. Use of the zero norm with linear models and kernel methods. JMLR 2003; 3:1439–61.

Wichaenkaew A, editor. Article collection of Tai Lue: Chiang Kam . Chiang Mai: 1986.

Wright S. Evolution and the genetics of populations. Chicago: Chicago Press; 1978.

Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2008; 14: 1–37.

Xu S, Jin L. PEAS: Package for elementary analysis of SNP data [online]. 2010 [cited 2011 Aug 4]. Available from: URL: <http://www.picb.ac.cn/~xushua/index.files/PEAS.htm>.

Zhou N, Wang L. Effective selection of informative SNPs and classification on the HapMap genotype data. BMC. 2007; 8(484): 1–9.

Zhou N, Wang L. Perfect population classification on Hapmap data with a small number of SNPs [online]. In: Ishikawa M, Doya K, Miyamoto H, Yamakawa T, editors. Neural Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2011 Sep 25]. 789–97. Available from: URL:<http://www.springerlink.com/content/pp6mw26114525834/>.