



دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد

در رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان

استفاده از تکنیک RVM برای کلاس بندی و رگرسیون داده های موهومی

استاد راهنما:

دکتر بهزاد مظفری تازه کند

استاد مشاور:

مهندس کریم صمدزمینی

پژوهشگر:

معصومه پیرفلک سرشکه

تاریخ ارائه پایان نامه

شهریور ۱۳۹۳

چکیده

در این پایان‌نامه که به روش کاربردی نظری انجام شده است و اطلاعات آن به صورت کانال مخابراتی موهومی تعیین شده است به بررسی روش طبقه‌بندی ماشین بردار ارتباطات^۱ (RVM) بر روی داده‌های موهومی پرداخته شده است. با افزایش اهمیت طبقه‌بندی داده‌ها در کاربردهای متفاوت، طبقه‌بندی کننده‌های مختلفی تعریف شده است که از آن جمله می‌توان به روش K نزدیک‌ترین همسایه، روش بیز، SVM و RVM اشاره داشت.

اخیرا RVM به دلیل تامین کردن تعداد زیادی از مزایا علاقه زیادی را در جامعه تحقیقاتی به خود جلب کرده است. RVM می‌تواند با قدرت نسبتا خوبی پیچیدگی رگرسیون^۲ و کلاس‌بندی^۳ را بر اساس مفهوم چارچوب بیزی احتمالاتی مدیریت کند و به طور گسترده‌ای در برخورد با مشکلات تشخیص‌های مختلف هم از آن استفاده می‌شود. این روش قابل تعمیم است و نتیجه را با یک هزینه محاسباتی کم تولید می‌کند. از طرفی بایازین^۴ روشی است که با استفاده از متغیرهای تصادفی پارامترها را آموزش می‌دهد تا محدودیت‌ها را مشخص کند تا در نهایت بتواند وزن توابع را به درستی تعیین کند. این الگوریتم در محیط متشکل از اعداد حقیقی پیاده‌سازی شده است. در حقیقت RVM بر مبنای حقیقی بودن اعداد کلاس‌بندی را انجام می‌دهد. در مقایسه روش ماشین بردار ارتباطات و ماشین بردار پشتیبان نتایج نشان می‌دهد که RVM برای ارزیابی داده‌ها در کاربردهای مختلف از نظر دقت مدل‌سازی منطقی و سرعت مدل‌سازی سریع‌تر و مناسب‌تر است. در این پایان‌نامه ابتدا نظریه‌های پایه از روش‌های کلاس‌بندی از جمله RVM و SVM برای رگرسیون و کلاس‌بندی بیان شده است و در نهایت الگوریتم پیشنهادی برای ارزیابی بر روی کانال مخابراتی اعمال شده است. در ادامه این روند تاثیر هسته‌های متفاوت بر روی دقت نتایج سنجیده می‌شود. در انتهای این پایان‌نامه نتایج حاصل در قسمت نتیجه‌گیری ارائه شده است.

کلمات کلیدی: RVM، SVM، کلاس‌بندی، رگرسیون، بایازین

¹ Relevance vector machine

² Regression

³ Classification

⁴ Bayesian

فهرست مطلب

صفحه

فصل اول: کلیات تحقیق	۱
مقدمه	۲
۱-۱ بیان مساله	۲
۲-۱ ضرورت و اهمیت تحقیق	۳
۳-۱ پرسش‌های تحقیق	۴
۴-۱ فرضیات تحقیق	۴
۵-۱ اهداف اصلی تحقیق	۴
۶-۱ روش تحقیق	۵
۷-۱ روش جمع‌آوری اطلاعات	۵
۸-۱ کاربرد تحقیق	۵
فصل دوم: پیشینه تحقیق	۷
مقدمه	۸
۱-۲ تشخیص توده‌های سرطانی در سینه	۸
۱-۱-۲ تشخیص خودکار خوشه‌های سرطانی MC	۹
۲-۱-۲ طبقه‌بندی خوشه‌های سرطانی MC	۱۳
۲-۲ استفاده از رگرسیون لجستیک و RVM برای تشخیص سرطان	۱۴
۱-۲-۲ نتایج تجربی تشخیص سرطان	۱۶
۳-۲ به کار بردن RVM برای GCM های نامرغوب هیدرولوژی	۱۸
۱-۳-۲ پیش‌پردازش داده‌ها	۲۰

۲۱	۲-۳-۲ نتایج تشخیص GCM
۲۳	۴-۲ کاربرد RVM در طبقه‌بندی متن XML
۲۸	۵-۲ کاربرد RVM در سیستم‌های مبتنی به DMT
۲۹	۲-۵-۱ مدل سیستم
۳۰	۲-۵-۲ مدولاتور DMT
۳۱	۲-۵-۳ نتایج شبیه‌سازی DMT
۳۴	۶-۲ جمع بندی
۳۵	فصل سوم: مفاهیم پایه برای کلاس‌بندی داده
۳۶	مقدمه
۳۶	۱-۳ ماشین‌های بردار پشتیبان
۳۷	۳-۱-۱ نحوه یادگیری در ماشین بردار پشتیبان
۴۰	۳-۱-۲ دسته‌بندی داده‌های جدید
۴۱	۳-۱-۳ معماری ماشین بردار پشتیبان
۴۱	۳-۲ دسته‌بندی داده‌ها به روش بایازین
۴۱	۳-۲-۱ قانون تصمیم‌گیری بیز برای کمترین خطا
۴۳	۳-۳ ماشین بردار ارتباطات
۴۵	۳-۳-۱ بهینه‌سازی هایپرپارامترها
۴۶	۴-۳ تخمین کانال
۴۸	۵-۳ تخمین کانال بر مبنای سمبل آموزشی
۴۹	فصل چهارم: روش پیشنهادی و ارزیابی آن
۵۰	مقدمه
۵۰	۱-۴ الگوریتم پیشنهادی

۵۱	۲-۴ فلوجارت کار پیشنهادی
۵۲	۳-۴ محیط پیاده‌سازی
۵۲	۴-۴ اجرا با هسته گوسی
۵۴	۵-۴ اجرا با هسته Cauchy
۵۶	۶-۴ اجرا با هسته فاصله اقلیدسی
۵۸	۷-۴ اجرا با هسته شاخص همسایگی
۶۰	۸-۴ اجرا با هسته لاپلاسی
۶۴	۹-۴ نتیجه‌گیری
۶۵	۱۰-۴ پیشنهادهایی برای کارهای آینده
۶۶	مراجع

مراجع

- [1] Dimitris G. T. , Wei L, Likas A. , Yang Y. , and Galatsanos P. , A Tutorial On Relevance Vector Mashine For Regression And Classification Whith Applications , Department of Computer Science, University of Ioannina, Ioannina, Chicago, 2006.
- [2] Li Y. , Zhang X., Yang X., Wang J., Kang Y., XML Text Classification Model by RVM Classifier, College of Mathematics and Information Science & Technology, Hebei Normal University of Science & Technology, China, 2013
- [3] Krishnapuram.B, Hartemink. A, e Carin.L, Applying Logistic Regression And RVM To Achive Accurate Probailistic Cancer Dlagnosis From Gene Expression Profiles, Dept. of Electrical Engineering, Duke University, Dept. of Computer Science, Duke University
- [4] Fan-Zi Z. and Zheng-Ding Q. A survey of classification learning algorithm. International Conference on Signal Processing 1504, 2004.
- [5] Fletcher.T. Relevance Vector Machines Explained. October 19, 2010
- [6] Tipping M. E. “Sparse Bayesian Learning and the Relevance Vector Machine”, *Journal of Machine Learning Research*, pp. 211-244, 2001.
- [7] Berges M. Simplified Support Vector Decision Rules. InL.Saitta,editor,Proceeding of the Thirteenth International Conference on Machine Learning ,page 71 -77,.Morgan Kaufmann. Bari,Italy1996
- [8]chen, D.L,Donoho,and M.A.saunders.Atomic decomposition by basispursuit.Thchnical Report 479,Department of Statistic, Stanford University,1995.
- [9] Bishop C.M .Neural Network For Pattern Recognition.Oxford University Press,1995
- [۱۰] اسدی.ع. تخمین کانال بی‌سیم با استفاده از تبدیل ویولت در استاندارد IEEE802.16. کارشناسی ارشد. دانشکده مهندسی برق و کامپیوتر. دانشگاه تبریز. شهریور ۹۱
- [11] *Cancer Facts and Figures 1998*. Atlanta, GA: American Cancer Society, 1998.
- [12] El-Naqa I. , Yang Y., Wernick M. N., Galatsanos N. P., and Nishikawa R. M., “A support vector machine approach for detection of microcalcifications,” *IEEE Trans. on Medical Imaging*, vol. 21, 1552-1563, 2002
- [13] Wei L., Yang Y., Nishikawa R. M., Wernick M. N. and Edwards A., “Relevance Vector Machine for Automatic Detection of Clustered Microcalcifications,” *IEEE Trans. on Medical Imaging*, vol. 24, 1278-1285, 2005.

- [14] Knutzen A. M. and Gisvold J. J., "Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions," *Mayo Clin. Proc.*, vol. 68, pp. 454- 460, 1993.
- [15] Kopans D. B., "The positive predictive value of mammography," vol. 158, pp. 521-526, 1992.
- [16] Figueiredo M. and Jain A. , "Bayesian learning of sparse classifiers," in *Proc. Computer Vision and Pattern Recognition*, pp. 35–41, 2001.
- [17] Neal R.M..Bayesian Learning for Neural Network.spring 1996.
- [18] Tipping M. E., and Faul A. "Fast Marginal Likelihood Maximization for Sparse Bayesian Models" Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Jan 3-6, 2003
- [19] Tipping M. "Sparse Bayesian Learning and the Relevance Vector Machine", *Journal of Machine Learning Research*, Vol. 1, Jun. 2001, pp. 211-244.
- [20] Ramaswamy S., et al. "Multiclass Cancer Diagnosis Using Tumor Gene Expression Signatures", *PNAS*, Vol. 98, No. 26, Dec. 2001, pp. 15149-15154.
- [21]. Chen H., Guo S.L., Xu C-Y., Singh V.P.,. Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin. *Journal of Hydrology*, Elsevier Science, Netherlands, 2007, 334,pp.171-184.
- [22] Guo J., Xiong W., Chen H., Application Of Relevance Vector Mashine To Downscale GCMs to Runoff Hydrology,fifth international conference om fuzzy systems and knowledge discovery
- [23] Shi J., Malik J., Normalized cuts and image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2010) 888 – 905.
- [24] Denoyer L. Gallinari P. Bayesian network model for semi-structured document classification, *Information Processing & Management* 40 (2004) 807 – 827.
- [25] Golden, P., Dedieu, H., and K. Jacobsen , Implementation and Application of DSL Technology. Boca Raton, FL: CRC Press, USA, 2007.
- [26] Weinstein S. and Ebert P., "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Transactions on Communications Technology*, October 1971, pp. 628–634.
- [27] Li C., and Lin Y., "Receiver window designs for radio frequency interference suppression in DMT systems," *IET signal processing*, January 2009, pp. 33-39.