33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Supervised Regularized Canonical Correlation Analysis: Integrating Histologic and Proteomic Data for Predicting Biochemical Failures

Abhishek Golugula[1], George Lee[2], Stephen R. Master[3], Michael D. Feldman[3], John E. Tomaszewski[3] and Anant Madabhushi[2]

[1]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854
[1]Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey 08854
[1]Department of Pathology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104

Abstract— **Multimodal data, especially imaging and non-imaging data, is being routinely acquired in the context of disease diagnostics; however computational challenges have limited the ability to quantitatively integrate imaging and non-imaging data channels with different dimensionalities for making diagnostic and prognostic predictions. The objective of this work is to create a common subspace to simultaneously accommodate both the imaging and non-imaging data, called a metaspace. This metaspace can be used to build a meta-classifier that produces better classification results than a classifier that is based on a single modality alone. In this paper, we present a novel Supervised Regularized Canonical Correlation Analysis (SRCCA) algorithm that (1) enables the quantitative integration of data from multiple modalities using a feature selection scheme, (2) is regularized, and (3) is computationally cheap. We leverage this SRCCA framework towards the fusion of proteomic and histologic image signatures for identifying prostate cancer patients at risk for biochemical recurrence following radical prostatectomy. For a cohort of 19 prostate cancer patients, SRCCA was able to yield a lower fused dimensional metaspace comprising both the histological and proteomic attributes. In conjunction with SRCCA, a random forest classifier was able to identify patients at risk for biochemical failure with a maximum accuracy of 93%. The classifier performance in the SRCCA space was statistically significantly higher compared to the fused data representations obtained either with Canonical Correlation Analysis (CCA) or Regularized CCA.**

I. Introduction

With the plentitude of multi-scale, multi-modal, disease pertinent data being routinely acquired for diseases such as breast and prostate cancer, there is an emerging need for powerful data fusion (DF) methods to integrate the multiple orthogonal data streams for the purpose of building diagnostic and prognostic meta-classifiers [1]. A major limitation in constructing integrated meta-classifiers that can leverage imaging (histology, MRI) and non-imaging (proteomics, genomics) data streams is having to deal with different data representations spread across different scales and dimensionalities [1]. This creates a need to represent the different modalities in a common subspace called a metaspace.

Several researchers have previously attempted to fuse such heterogeneous data [2] but all of these DF techniques have their own weaknesses in creating an appropriate metaspace that can simultaneously accomodate multiple imaging and non-imaging modalities. Generalized Embedding Concatenation [3] relies on dimensionality reduction methods that face the risk of extracting noisy features which degrade the metaspace [4]. Other DF techniques, including consensus embedding, multi-kernel graph embedding, and boosted embedding [2] have yielded promising results, but come at a high computational cost.

Canonical Correlation Analysis (CCA) and its regularized version, (RCCA), are DF techniques for fusing two modalities. They capitalize on the knowledge that the different modalities represent different sets of descriptors for characterizing the same object. In recent years, CCA has been used to find linear relationships between the pixel values of images and the text attached between these images [5]. RCCA has been used to study expressions of genes measured in liver cells and compare them with concentrations of hepatic fatty acids in mice [6].

CCA is a simple technique but it suffers from over fitting when the modalities have large dimensions. RCCA is a modification to CCA that prevents over fitting but this procedure is computationally very expensive. Both these algorithms also fail to take complete advantage of class label information, when available. In this paper, we present an efficient Supervised RCCA (SRCCA) algorithm that performs DF without being plagued by issues of over fitting while also being computationally cheap. Moreover, it makes better use of labeled information that can significantly help stratify the data in the metaspace.

In this work, we apply SRCCA to the problem of predicting biochemical recurrence in prostate cancer (CaP) patients, following radical prostatectomy, by fusing histologic imaging and proteomic signatures. Biochemical recurrence is commonly defined as a doubling of Prostate Specific Antigen (PSA), a key biomarker for CaP. However, the nonspecificity of PSA leads to over-treatment of CaP, resulting in many unnecessary treatments, which are both stressful and costly [3]. Thus, the overarching goal of this study is to leverage SRCCA to construct a fused histologic, proteomic marker for predicting biochemical recurrence in CaP patients following surgery.

Our main contributions in this paper are:

- A novel data fusion algorithm, SRCCA, that builds an accurate metaspace representation that can simultaneously represent and accommodate two heterogeneous

This work was made possible by grants by the Walter H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program, The Cancer Institute of New Jersey and the Society for Imaging Informatics in Medicine.

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

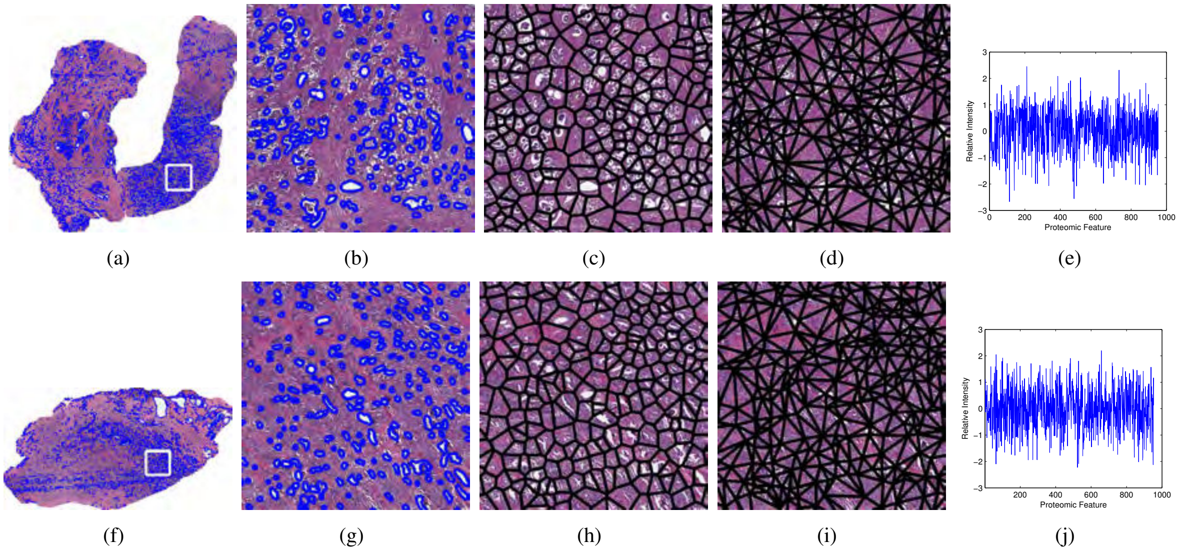

Fig. 1. Multi-modal patient data (top row: relapsed case, bottom row: non-relapsed case). (a), (f) Original prostate histology section showing region of interest, (b), (g) Magnified ROI showing gland segmentation boundaries, (c), (h) Voronoi Diagram (d), (i) Delaunay Triangulation depicting gland architecture, (e), (j) Plot of the proteomic profile obtained from the dominant tumor nodule regions (white box in (a), (f) respectively) via mass spectrometry.

imaging and non-imaging modalities.

- Leveraging SRCCA to build a meta-classifier to predict risk of 5 year biochemical failure in prostate cancer patients following radical prostatectomy by integrating histological image and proteomic features.

II. Supervised Regularized Canonical Correlation Analysis (SRCCA)

A. Canonical Correlation Analysis (CCA)

CCA [5] is a way of using cross-covariance matrices to obtain a linear relationship between two multidimensional variables, $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$, where p and q are the number of features in X and Y and n is the number of overall samples. CCA obtains two directional vectors $w_x \in \mathbb{R}^{p \times 1}$ and $w_y \in \mathbb{R}^{q \times 1}$ such that $Xw_x \in \mathbb{R}^{n \times 1}$ and $Yw_y \in \mathbb{R}^{n \times 1}$ will be maximally correlated. It is defined as the optimization problem [5]:

$$\rho = \max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \tag{1}$$

where $C_{xy} \in \mathbb{R}^{p \times q}$ is the covariance matrix of the matrices X and Y, $C_{xx} \in \mathbb{R}^{p \times p}$ is the covariance matrix of the matrix X with itself and $C_{yy} \in \mathbb{R}^{q \times q}$ is the covariance matrix of the matrix Y with itself. The solution to CCA reduces to the solution of the following two generalized eigenvalue problems [7]:

$$C_{xy} C_{yy}^{-1} C_{yx} = \lambda C_{xx} w_x \tag{2}$$

$$C_{yx} C_{xx}^{-1} C_{xy} = \lambda C_{yy} w_y \tag{3}$$

where λ is the generalized eigenvalue representing the canonical correlation, and w_x and w_y are the corresponding generalized eigenvectors. CCA can further produce exactly $\min\{p, q\}$ orthogonal embedding components (sets of Xw_x and Yw_y) which can be sorted in order of decreasing correlation, λ.

DF is performed as described in Foster et al. [8]. When the Xw_x and Yw_y are maximally correlated, each modality represents similar information. In order of decreasing λ, the top d embedding components can be chosen to represent the two modalities in a metaspace.

B. Regularized Canonical Correlation Analysis (RCCA)

When $n << p$ or $n << q$, the features in X and Y tend to be highly collinear. This leads to ill-conditioned covariance matrices C_{xx} and C_{yy} such that their inverses are no longer reliable. The greatest λ's tend to be nearly 1 and the remaining $d-1$ dimensions do not provide any new meaningful information.

RCCA [6] corrects for noise in X and Y by assuming first that X and Y are contaminated with $N_X \in \mathbb{R}^{n \times p}$ and $N_Y \in \mathbb{R}^{n \times q}$. Since the p and q columns of N_X and N_Y, respectively, are gaussian, independent and identically distributed, all combinations of the covariances of the p columns of N_X and q columns of N_Y will be 0 except the covariance of a particular column vector with itself. This variance of each column of N_X and N_Y is labeled λ_x and λ_y. The matrix C_{xy} will not be affected but the matrices C_{xx} and C_{yy} become $C_{xx} + \lambda_x I_x$ and $C_{yy} + \lambda_y I_x$. The solution to RCCA becomes the solution to these generalized eigenvalue problems [7]:

$$C_{xy}(C_{yy} + \lambda_y I_y)^{-1} C_{yx} = \lambda (C_{xx} + \lambda_x I_x) w_x \tag{4}$$

$$C_{yx}(C_{xx} + \lambda_x I_x)^{-1} C_{xy} = \lambda (C_{yy} + \lambda_y I_y) w_y \tag{5}$$

The noise parameters next have to be chosen. For $i \in \{1, 2, ..., n\}$, let w_x^i and w_y^i denote the weights calculated from RCCA when samples X_i and Y_i are removed. λ_x and λ_y are varied in a certain range $\theta_1 \leq \lambda_x, \lambda_y \leq \theta_2$ and chosen via the optimization [6]:

$$\max_{\lambda_x, \lambda_y} \left[corr(\{X_i w_x^i\}_{i=1}^n, \{Y_i w_y^i\}_{i=1}^n) \right] \tag{6}$$

C. Extending RCCA to SRCCA

SRCCA chooses λ_x and λ_y using a supervised feature selection method (Wilks Lambda Test [9]). The data in the metaspace, $\gamma = Xw_x$ or Yw_y, can be split using its labels into α and β, where α contains the n_1 samples that belong to Class 1 and β contains the n_2 samples that belong to Class 2. These three vectors are then used to calculate Wilks Lambda (Λ), which is defined as the ratio of within group variance to total variance, and minimized as:

$$\min_{\lambda_x, \lambda_y} \frac{(\alpha - 1\bar{\alpha})^T(\alpha - 1\bar{\alpha}) + (\beta - 1\bar{\beta})^T(\beta - 1\bar{\beta})}{(\gamma - 1\bar{\gamma})^T(\gamma - 1\bar{\gamma})} \qquad (7)$$

where $\alpha \in \mathbb{R}^{n_1 \times 1}$, $\beta \in \mathbb{R}^{n_2 \times 1}$, $\gamma \in \mathbb{R}^{n \times 1} = [\alpha \ \beta]$, and $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ are denoted as the means of vectors α, β and γ respectively. A lower Λ value will indicate that the data will be more discriminatory in the lower dimensional metaspace.

D. Computational Complexity

Assume v potential λ_x and λ_y sampled evenly between θ_1 and θ_2 . Given $\varphi = \min\{p, q\}$, RCCA has a computational complexity of $vn\varphi!$ because RCCA requires CCA, which has a computational complexity of $\varphi!$ (based on the source code in [10]) to be computed n times, where n is the sample size, over v intervals. SRCCA only requires CCA to performed once each interval, leading to a much cheaper computational complexity of $v\varphi!$.

III. Experimental Results and Discussion

A. Data Description

19 CaP patients at the Hospital at the University of Pennsylvania were identified, all of whom had gland resection. 10 of these patients had biochemical recurrence within 5 years following surgery (Non-Failure) and the other 9 did not (Failure). For each patient, a representative histology section was chosen and the tumor nodule identified. Mass Spectrometry (MS) was performed at this site to yield a protein expression vector. The aim of this experiment is to combine quantitative image descriptors on histology of the tumor with the proteomic vector to build a meta-classifier to distinguish the patients at risk of recurrence from those who are not.

B. Proteomic Feature Selection (denoted $\phi^P \in \mathbb{R}^{19 \times 953}$)

Active genes encode proteins that are present in a tissue sample, and these proteins can be measured and serve as quantitative markers of cancer activity. For this study, MS was used to measure the relative abundance of peptides in cancerous regions of the tissue. A high dimensional feature vector was obtained, characterizing each patient's protein expression profile at the time of treatment. This data underwent quantile normalization, log(2) transformation, and mean and variance normalization on a per-protein basis.

C. Quantitative Histologic (denoted $\phi^H \in \mathbb{R}^{19 \times 151}$) Feature Extraction

Following an automated gland segmentation process used to define the gland centroids and boundaries (see [11] for details), morphological (denoted $\phi^M \in \mathbb{R}^{19 \times 100}$) and architectural (denoted $\phi^A \in \mathbb{R}^{19 \times 51}$) image features (quantifying glandular arrangement) were extracted from the dominant tumor region on histology [3].

D. Fusing Proteomic, Histologic Features for Predicting Biochemical Recurrence in CaP Patients Post-Surgery

We perform CCA, RCCA and SRCCA on the non-imaging modality, ϕ^P, and the selected imaging modality, ϕ^J, where $J \in \{H, M, A\}$. ϕ^P was reduced to 25 features as ranked by the t-test, with a p-value cutoff of p = .05, using a leave one out validation strategy. For CCA, ϕ^P and ϕ^J were used as the two multidimensional variables, X and Y, as mentioned above in Sec II. For RCCA and SRCCA, ϕ^P and ϕ^J were used in a manner similar to CCA except they are tested with regularization parameters λ_x and λ_y evenly spaced from $\theta_1 = .001$ to $\theta_2 = .2$ with $v = 200$.

Experiment 1 - Obtaining a Fused Proteomic, Histologic meta-classifier

After using the top $d = 2$ embedding components, the classification accuracies of K-Nearest Neighbor (ϕ^{KNN}), with k = 1, and Random Forrest (ϕ^{RF}), with 50 Trees, were determined using leave-one-out cross-validation.

Experiment 2 - Comparing classifier accuracy for SRCCA, CCA, and RCCA based metaspace representations

Using these 10 different values for $d \in \{1, 2, ..10\}$, and the 3 fusion schemes considered (ϕ^P, ϕ^M), (ϕ^P, ϕ^A) and (ϕ^P, ϕ^H), 30 different embeddings were obtained for CCA, RCCA and SRCCA. The maximum and median of these 30 different measurements for each classifier were calculated. In addition, two paired Student *t*-tests were employed to identify whether there were statistically significant improvements for the ϕ^{KNN} and ϕ^{RF} when: (1) CCA and SRCCA and (2) RCCA and SRCCA.

Experiment 3 - Computational consideration for the 3 different CCA variants

We repeated Experiment 2 and measured the time for RCCA and SRCCA to distinguish between the failures and non-failures. These experiments were performed on a quadcore computer with a clock speed of 1.8GHz.

E. Experimental results

Experiment 1: Across both classifiers, SRCCA had a median classification accuracy of 71% compared to 42% for CCA and 42% for RCCA. SRCCA also performed better in 10 of 12 direct comparisons with CCA and RCCA, while underperforming only once (fusing ϕ^P and ϕ^H with the classifier KNN) (see Tables I and II). The higher classification accuracy results indicate that SRCCA produces a better metaspace compared to both CCA and RCCA.

These results, which strongly suggest that SRCCA outperforms CCA and RCCA, are observable in the embedding plots of Figure 2. More importantly, we see that because CCA lacks regularization, the corresponding covariance matrices have unreliable inverses. For this reason, in Figure 2 the embedding components are not orthogonal but are highly correlated to each other and yield the same information. RCCA overcomes this regularization problem but still does not produce the same level of discrimination between patient classes compared to SRCCA.

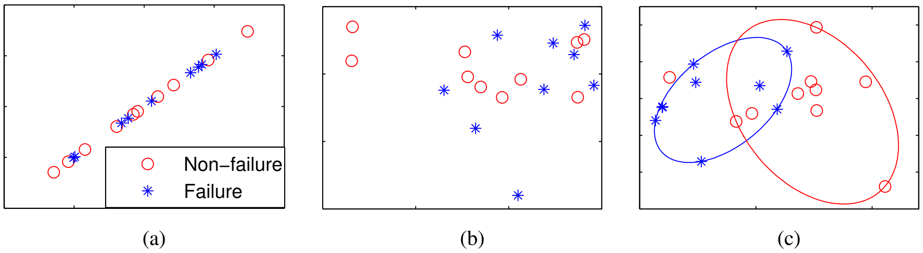

Fig. 2. 2-dimensional representation of (ϕ^P, ϕ^A) using (a) CCA, (b) RCCA, and (c) SRCCA where the X and Y axes are the two most significant embedding components produced by the 3 different algorithms. CCA (a) suffers from lack of regularization, RCCA (b) is regularized but does not produce the best metaspace while SRCCA (c) results in the best embedding components in terms of classification accuracy distinguished via best fit ellipses with one outlier.

TABLE I
AVERAGE K-NEAREST NEIGHBOR CLASSIFICATION ACCURACIES

Dataset (ϕ^I, ϕ^J)	CCA	RCCA	SRCCA
(ϕ^P, ϕ^M)	42%	37%	**68%**
(ϕ^P, ϕ^A)	37%	47%	**74%**
(ϕ^P, ϕ^H)	**74%**	31%	68%

TABLE II
AVERAGE RANDOM FOREST CLASSIFICATION ACCURACIES

Dataset (ϕ^I, ϕ^J)	CCA	RCCA	SRCCA
(ϕ^P, ϕ^M)	42%	48%	**70%**
(ϕ^P, ϕ^A)	36%	30%	**71%**
(ϕ^P, ϕ^H)	**79%**	46%	**79%**

Experiment 2: In Tables III and IV we see that the maximum and median ϕ^{KNN} and ϕ^{RF} of SRCCA for fusion of (ϕ^P, ϕ^J) were much higher than the corresponding values of CCA or RCCA. We also see that SRCCA attains a maximum classifier accuracy of 93.16% (see Table III). In Table V, we see that SRCCA yielded a statistically significant improvement over CCA and RCCA across both classifiers even at the $p = .01$ level.

Experiment 3 Figure 3 reveals that SRCCA is much faster and more efficient than RCCA. Even though the completion times are visibly different, a p-value of 1.9×10^{-3} even with just 3 samples, indicates that SRCCA is certainly statistically significantly faster than RCCA.

TABLE III
MAXIMUM ϕ^{KNN} AND ϕ^{RF} OF DF SCHEMES ACROSS $d \in \{1, 2, ..10\}$

Classifier	CCA	RCCA	SRCCA
ϕ^{KNN}	73.68%	68.42%	**84.21%**
ϕ^{RF}	80.20%	68.42%	**93.16%**

TABLE IV
MEDIAN ϕ^{KNN} AND ϕ^{RF} OF DF SCHEMES ACROSS $d \in \{1, 2, ..10\}$

Classifier	CCA	RCCA	SRCCA
ϕ^{KNN}	57.89%	47.37%	**68.42%**
ϕ^{RF}	58.42%	37.37%	**74.21%**

IV. CONCLUDING REMARKS

In this paper, we presented a novel supervised variation of CCA, Supervised Regularized Canonical Correlation Analysis (SRCCA). We applied this method to the problem of predicting 5 year biochemical failure in prostate cancer patients who have undergone radical prostatectomy. Overall, SRCCA allows for construction of a more accurate metaspace representation of imaging and non-imaging data compared to CCA and RCCA. Using the RF classifier, we are able to achieve a meta-classifier with classification results as high as 93%. Moreover, SRCCA is computationally much cheaper compared to RCCA. These results strongly indicate that SRCCA is a powerful tool in multimodal DF.

TABLE V
STATISTICAL SIGNIFICANCE (p-VALUE) OF SRCCA

Classifier	SRCCA vs CCA	SRCCA vs RCCA
ϕ^{KNN}	4.0×10^{-9}	7.1×10^{-11}
ϕ^{RF}	2.1×10^{-7}	3.6×10^{-16}

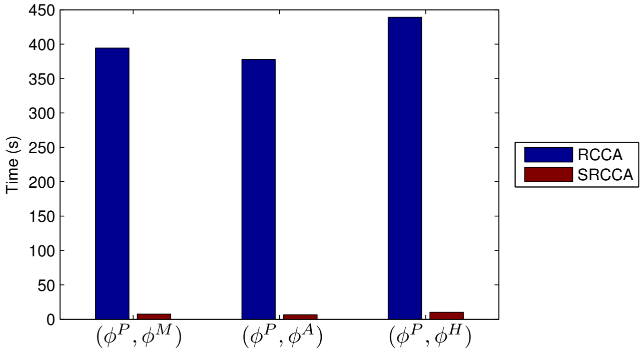

Fig. 3. *Experiment 3:* Computational run times for SRCCA and RCCA for the different fusion combinations. SRCCA significantly outperforms RCCA across all fusion experiments.

REFERENCES

[1] A. Madabhushi *et al.*, "Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multimodal data." *CMIG*, Feb 2011.

[2] P. Tiwari *et al.*, "Multi-modal data fusion schemes for integrated classification of imaging and non-imaging biomedical data," *ISBI*, pp. 165–168, 2011.

[3] G. Lee *et al.*, "A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology," *ISBI*, pp. 77–80, 2009.

[4] Y. Wu *et al.*, "Optimal multimodal fusion for multimedia data analysis," *ACM Conf. on Multimedia*, pp. 572 – 579, 2004.

[5] D. R. Hardoon *et al.*, "Canonical correlation analysis: an overview with application to learning methods." *Neural Comput*, vol. 16, no. 12, pp. 2639–2664, Dec 2004.

[6] I. Gonzalez *et al.*, "Cca: An r package to extend canonical correlation analysis," *Journal of Stat. Software*, vol. 23, no. 12, pp. 1–14, 1 2008.

[7] L. Sun *et al.*, "A least squares formulation for canonical correlation analysis," *ICML*, vol. 33, no. 1, pp. 1024–1031, 2008.

[8] D. Foster *et al.*, "Multi-view dimensionality reduction via canonical correlation analysis," *Technical Report TR-2008-4, TTI-Chicago*, 2008.

[9] D. Hwang *et al.*, "Determination of minimum sample size and discriminatory expression patterns in microarray data," *Bioinformatics*, vol. 18, pp. 1184–1193, 2002.

[10] M. Borga *et al.*, "Blind source separation of functional mri data," in *SSBA*, 2002.

[11] J. Monaco *et al.*, "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models," *Medical Image Analysis*, vol. 14(4), pp. 617–629, 2010.