OUTCOME PREDICTION IN TUMOUR THERAPY BASED ON DEMPSTER-SHAFER THEORY

Chunfeng Lian[1,2] *Su Ruan*[2] *Thierry Denœux*[1] *Pierre Vera*[2,3]

[1] Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, 60205 Compiègne, France
[2] Université de Rouen, QuantIF - EA 4108 LITIS, 76000 Rouen, France
[3] Centre Henri-Becquerel, Department of Nuclear Medicine, 76038 Rouen, France

ABSTRACT

Outcome prediction plays a vital role in cancer treatment. It can help to update and optimize the treatment planning. In this paper, we aim to find discriminant features from both PET images and clinical characteristics, so as to predict the outcome of a treatment to adapt the therapy. As both information sources are imprecise, we propose a novel feature selection method based on Dempster-Shafer theory to tackle this problem. Then, a specific objective function with sparsity constraint is developed to search for a feature subset that leads to increasing prediction performance and decreasing data imprecision simultaneously. Our approach was applied to two real data sets concerning to lung tumour et esophageal tumour, showing good performance.

Index Terms— Outcome Prediction, Feature Selection, Dempster-Shafer Theory, PET imaging

1. INTRODUCTION

Accurate outcome prediction prior to or even during the cancer treatment is of great clinical value, upon which more effective treatment planning can be updated. Medical imaging plays a fundamental role in this task, since it realizes noninvasive monitoring of tumour lesions. Some research has proven that functional information provided by fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) is predictable for response of therapy [1]. In FDG-PET, well-explored imaging features include, but are not limited to, metabolic tumour volume (MTV), total lesion glycolysis (TLG), as functional indices describing metabolic tumor burden, and standardized uptake values (SUVs) describing FDG uptake within a region of interest (ROI), e.g. SUVmean, SUVpeak, or single pixel (SUVmax). Additionally, texture analysis through PET images may also provide potential complementary predictive value for outcome assessment [1]. However, there is no consensus to determine the most predictive variables.

This work was partly supported by China Scholarship Council.

Feature selection is a good way to tackle this challenge. It aims to select a subset of features that facilitates data understanding and improves prediction performance [2]. Filter, wrapper and embedded methods are the three main categories of feature selection algorithms. Utilizing variable ranking as the principal selection mechanism, filter methods are simple and scalable; however, they can produce a sub-optimal subset as they ignore correlation between features [2]. Wrapper and embedded methods, such as sequential selection algorithms [3] and direct objective optimization methods [4, 5], use the prediction accuracy of classifiers as the criteria for selecting feature subsets. They are more likely to yield optimal feature subsets than filter methods. However, available wrapper or embedded methods are not designed to work for imperfect data with uncertainty, as often encountered in the medical domain. This motivates us to design a new feature selection approach to select the most discriminant PET-based and clinical features from imperfect patient data.

In this paper, a novel wrapper selection method based on Dempster-Shafer theory is proposed. A modified evidential K-nearest neighbour classifier (mEK-NN) is designed at first, in which the previous evidence construction protocol of classical EK-NN [6] is improved to achieve a more robust representation of the uncertainty about given data. Then, based on mEK-NN, a special objective function is developed to search for predictive feature subsets that lead to high classification accuracy and low overlapping between classes (low imprecision and uncertainty).

The rest of this paper is organized as follows. The fundamental background on Dempster-Shafer theory is reviewed in Section 2. Then the proposed method is presented in Section 3, followed by some experimental results presented in Section 4. Finally, Section 5 concludes this paper.

2. DEMPSTER-SHAFER THEORY

As a formal framework to model and fuse imperfect information (partial knowledge) for reasoning and decision making, Dempster-Shafer theory (DST) [7] is the extension of both

978-1-4799-2374-8/15/$31.00 ©2015 European Union

probability theory and set-membership approach.

Let ω be a variable taking values in $\Omega = \{\omega_1, \cdots, \omega_c\}$, called the *frame of discernment*. An item of evidence regarding the actual value of ω is represented by a *mass function* m from 2^Ω to [0,1], such that $\sum_{A\subseteq\Omega} m(A) = 1$. Each $A \subseteq \Omega$ corresponds to a hypothesis that "$\omega \in A$". Function m is said to be *normalized* if $m(\emptyset) = 0$, which is assumed in this paper.

The *credibility* and *plausibility* of subset A are, respectively, quantified by the *belief* and *plausibility function*:

$$Bel(A) = \sum_{B\subseteq A} m(B); \quad Pl(A) = \sum_{B\cap A\neq\emptyset} m(B). \tag{1}$$

Two mass functions m_1 and m_2 derived from independent items of evidence can be combined by the *Dempster's rule* to generate a new mass function:

$$(m_1 \oplus m_2)(A) = \frac{1}{1-Q} \sum_{B\cap C=A} m_1(B)m_2(C) \tag{2}$$

for all $A \in 2^\Omega/\emptyset$, where $Q = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$ is the *degree of conflict* between these two pieces of evidence.

If Q is large, which reveals a strong conflict between the items of evidence, *Yager's rule* [8] can also be used to combine two mass functions:

$$m(A) = \begin{cases} 0 & , \quad \text{if } A = \emptyset; \\ \sum_{B\cap C=A} m_1(B)m_2(C), & \text{if } A \subset \Omega, A \neq \emptyset; \\ m_1(\Omega)m_2(\Omega) + Q, & \text{if } A = \Omega. \end{cases} \tag{3}$$

Using Yager's rule, the conflicting mass of belief is conservatively assigned to the whole frame Ω.

3. APPROACH

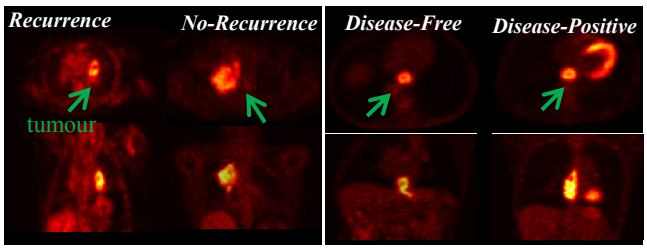
In this section, the construction protocol of mass function in our modified EK-NN (mEK-NN) is described at first, upon which the feature selection procedure is then presented.

3.1. Construction of Mass Functions

Let $\{(X_i, Y_i)|i = 1, \cdots, N\}$ be a collection of N training pairs, $X_i = [x_1, \cdots, x_v]$ the ith training instance with v features, and Y_i be its class label. Assume all class labels form a frame of discernment $\Omega = \{\omega_1, \cdots, \omega_c\}$. A mass function regarding each query instance X_t's class can be constructed as follows.

Step 1: Let X_j be the jth ($j \in \{1, \ldots, K\}$) nearest neighbour of X_t. According to [6], the knowledge that $Y_j = \omega_q$ ($q = 1, \cdots, c$) supports the hypothesis that the query instance X_t also belongs to this class. This piece of evidence can be quantified as

$$\begin{cases} m_{t,j}(\omega_q) = \alpha \cdot e^{-\gamma_q d_{t,j}^2} \\ m_{t,j}(\Omega) = 1 - \alpha \cdot e^{-\gamma_q d_{t,j}^2} \end{cases}, \tag{4}$$



(a) Lung Tumours (b) Esophageal Tumours

Fig. 1. Example tumour uptakes on FDG-PET imaging from different views; (a) recurrence and no-recurrence instances before treatment of lung tumour; (b) disease and disease-free instances before treatment of esophageal tumour.

where $d_{t,j}$ is the distance between X_j and X_t. In our application, this distance represents the dissimilarity between two patients (X_j and X_t). In addition, α and γ_q are two constant parameters. We can heuristically set $\alpha = 0.95$, and let γ_q be the inverse of mean distance between training instances with label ω_q.

Step 2: The nearest neighbours with the same class label ω_q are grouped as a set Γ_q. Since the corresponding mass functions have relatively low conflicts, Dempster's rule [7] is used to fuse them to obtain Γ_q's group mass functions $m_t^{\Gamma_q}$.

Step 3: We can assume that, when more neighbours belong to the same class, X_t will also be more likely to be in this class. So the mass functions obtained in last step are further discounted according to each set's size:

$$\begin{cases} dm_t^{\Gamma_q}(\omega_q) = (\frac{|\Gamma_q|}{|\Gamma_{max}|})^\eta \cdot m_t^{\Gamma_q}(\omega_q) \\ dm_t^{\Gamma_q}(\Omega) = 1 - (\frac{|\Gamma_q|}{|\Gamma_{max}|})^\eta \cdot m_t^{\Gamma_q}(\omega_q) \end{cases}, \tag{5}$$

where $|\Gamma_{max}|$ is the maximum cardinality in $\{|\Gamma_1|, \cdots, |\Gamma_c|\}$, and η (≥ 0) is a coefficient that controls the discounting level. A larger η brings stronger discounting. Heuristically, average good results with $\eta = 0.5$ have been found.

Step 4: Mass functions in different sets possibly contain a certain degree of conflicts, especially when there are similar amounts of nearest neighbours. So, Yager's rule is chosen to fuse the corresponding mass functions of set Γ_1 to Γ_c. The global mass function regarding the class label of X_t is:

$$\begin{cases} m_t(\omega_q) = dm_t^{\Gamma_q}(\omega_q) \cdot \prod_{h\in\{1,\ldots c\}\backslash q} dm_t^{\Gamma_h}(\Omega) \\ m_t(\Omega) = 1 - \sum_{t=1}^{c} m_t(\omega_q) \end{cases}. \tag{6}$$

Final decision regarding the class label X_t is based on the credibility or plausibility criteria (i.e., Bel_t and Pl_t). They lead to the same prediction in our application.

3.2. Evidential Feature Selection

A good feature subset should meet three requirements:

1. High classification accuracy;
2. Low imprecision and uncertainty (small overlaps between different classes);
3. Sparsity to reduce the risk of over-fitting.

According to these requirements, we propose an evidential feature selection (EFS) method.

Let X_i and X_j be two training instances (patients have labeled outcome) with v features, a simple weighted Euclidian distance metric is used to measure their dissimilarity:

$$d_{i,j} = \left(\sum_{p=1}^{v} \lambda_p \cdot {d_{i,j}^p}^2 \right)^{1/2} \quad (7)$$

where $d_{i,j}^p$ is the difference between the pth dimension of the two feature vectors, and $\lambda_p \in \{0, 1\}$ is the corresponding *binary coefficient*. Features will be selected by determining $\{\lambda_1, \ldots, \lambda_v\}$. The pth dimension of the feature vectors will be selected if $\lambda_p = 1$, and vice versa.

Applying this weighted euclidian distance in Equation (4) to represent evidence from the nearest neighbours, then using the procedure described in Section 3.1 to construct mass functions for all training samples, an objective function is proposed to be minimized:

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{q=1}^{c} (Pl_i(\omega_q) - t_{i,q})^2 + \rho \times \frac{1}{n} \sum_{i=1}^{n} m_i(\Omega) + \xi \times Spa. \quad (8)$$

In this objective function, the first term is a square error measure corresponds to the first requirement, Pl_i is the plausibility (Equation (1)) of singletons for training sample X_i, $t_{i,q}$ is the qth element of a target output vector $t_i = \{t_{i,1}, \ldots, t_{i,c}\}$, with $t_{i,q} = \delta_{i,q}$ if $Y_i = \omega_q$. The second term corresponds to the second requirement, which penalizes subsets that result in high imprecision; while the last term forces the feature subset to be sparse. An approximation of the l_0-norm is utilized with $Spa = \sum_{p=1}^{v} (1 - e^{-5\lambda_p})$. As two hyper-parameters, ρ and ξ should be tuned for given data to compromise between prediction accuracy, imprecision ratio and feature subset size. Generally, appropriate values could be found by rough grid searching with $\rho \in [0, 0.5]$ and $\xi \in (0, 0.1]$.

As a global optimization method, the integer genetic algorithm [9] is used to solve this integer optimization problem.

4. EXPERIMENTAL RESULTS

In this section, the proposed feature selection approach is compared with several state-of-the-art feature selection methods. Its applicability is then assessed for different predictors. Two real patient data sets were used:

1) Lung Tumour Data: Twenty-five patients with stage II-III non small cell lung cancer were studied. 52 SUV-based (SUV_{max}, SUV_{mean}, SUV_{peak}, MTV and TLG) and texture-based (gray level size zone matrices (GLSZM) [1]) features were extracted. The definition of recurrence for patients at one year after the treatment is primarily clinical with biopsy and PET/CT. Local or distant *recurrence* is diagnosed on 19 patients, while *no recurrence* is reported on the remaining 6 patients (example images can be seen in Figure 1(a)).

2) Esophageal Tumor Data: Thirty-six patients with esophageal squamous cell carcinomas were studied. We have 29 SUV-based (SUV_{max}, SUV_{mean}, SUV_{peak}, MTV and TLG), GLSZM-based and patients' clinical features (gender, tumour stage and location, WHO performance status, dysphagia grade and weight loss from baseline). The disease-free evaluations include a clinical examination with PET/CT and biopsies. 13 patients were labeled *disease-free* when neither loco regional nor distant tumor recurrence is detected, while the remaining 23 patients were diagnosed as *disease-positive* (example images can be seen in Figure 1(b)).

4.1. Feature Selection Performance

In the leave-one-out cross-validation protocol, the proposed evidential feature selection (EFS) was compared with two classical wrapper methods (SFS and SFFS [3]) and a widely used imbedded method, SVMRFE [5]. The classification accuracy of SVM (Gaussian kernel with $\sigma = 1$ was empirically chosen) serves as the selection criteria in SFS and SFFS. The optimal hyper-parameters used in EFS and cutoff thresholds for all the last three methods (feature subsets selection) were determined using a rough grid search strategy. In each iteration, the selected feature subsets were used to predict the test data. The same SVM classifier was still used after SFS and SFFS, while the proposed mEK-NN was executed after EFS. Finally, the average prediction accuracy with 95% confidence interval and the selected subset size were calculated. Based on feature frequency statistics, the robustness of selection methods was evaluated using the criteria introduced in [10]. All these results are summarized on Table 1, in which experiments of SVM with all features are presented too as baseline for comparison. As can be seen, the proposed EFS method leads to both higher accuracy and higher robustness.

The four features robustly selected by EFS in Lung Tumour are one SUV-based feature (SUVmax during radiotherapy) and three texture-based features; while the three features robustly selected in Esophageal Tumour are one SUV-based feature (TLG before the treatment) and two clinical features.

4.2. Classification Performance

We further tested whether feature subsets selected by EFS are applicable for other classifiers. To this end, Artificial Neural Networks (ANN), SVM, EK-NN, and the proposed mEK-NN were studied. The scaled conjugate gradient backpropagation network was used here in testing ANN. The number of neurons in the hidden-layer was empirically set as 10.

In the leave-one-out cross validation protocol, the selected feature subsets and all features were fed in these classifiers.

Table 1. Comparing feature selection methods using leave-one-out cross-validation. Average prediction accuracy (%) with 95% Confidence Interval (CI), selection robustness (%) and selected subset size are presented. EFS* denotes the proposed method.

Method	Lung Tumour Data			Esophageal Tumour Data		
	Accuracy [95% CI]	Robustness	Subset size	Accuracy [95% CI]	Robustness	Subset size
All features	76 [58-94]	n/a	52	64 [47-80]	n/a	29
SFS	84 [69-99]	60	3	53 [36-70]	63	3
SFFS	72 [53-91]	54	4	**81 [67-94]**	53	3
SVMRFE	92 [81-100]	57	5	75 [60-90]	80	5
EFS*	**100 [100-100]**	**94**	4	**81 [67-94]**	**92**	3

Table 2. Comparing the average prediction accuracy (with 95% confidence interval) of features selected by EFS with all features using different classifiers. mEK-NN* denotes the proposed classification method.

Classifier	Lung Tumour Data		Esophageal Tumour Data	
	without EFS	with EFS	without EFS	with EFS
ANN	68 [48-88]	92 [81-100]	67 [50-83]	83 [71-96]
SVM	76 [58-94]	**100 [100-100]**	64 [47-80]	81 [67-94]
EK-NN	68 [48-88]	96 [88-100]	64 [47-80]	83 [71-96]
mEK-NN*	56 [35-77]	**100 [100-100]**	53 [36-70]	**89 [78-100]**

The average classification accuracy is summarized in Table 2. As can be seen, the proposed EFS improves all classifiers' prediction accuracy. In addition, it has the best performance when working with our mEK-NN.

5. CONCLUSION

In this study, a novel approach based on Dempster-Shafer theory has been developed to find discriminant features from both PET images and clinical characteristics, so as to predict the outcome in cancer treatment. Experiments on two real data sets have been designed to evaluate the performance of the proposed method. The obtained results show that it can robustly select discriminant feature subsets to improve the prediction accuracy. The selected SUV-based features have also been confirmed important in other studies [1]. Our future work will evaluate the proposed method on more and larger data sets with different types of tumours.

6. REFERENCES

[1] F. Tixier, M. Hatt, C.C. Le Rest, A. Le Pogam, L. Corcos, and D. Visvikis, "Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18f-fdg pet," *Journal of Nuclear Medicine*, vol. 53, no. 5, pp. 693–700, 2012.

[2] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[3] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

[4] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, and Y. Zhu, "Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation," *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 256–269, 2011.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[6] T. Denœux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 25, no. 5, pp. 804–813, 1995.

[7] G. Shafer, *A mathematical theory of evidence*, vol. 1, Princeton university press Princeton, 1976.

[8] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[9] K. Deep, K.P. Singh, M.L. Kansal, and C. Mohan, "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518, 2009.

[10] P. Somol and J. Novovicova, "Evaluating stability and comparing output of feature selectors that optimize feature subset cardinality," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 1921–1939, 2010.