Particle Swarm Optimization with Hidden Markov Model Prediction Approach for Anti-Cancer Drug Sensitivity

Daniel NesaKumar.C[1], Jayasree.P[2]

[1]Department of MCA, Hindusthan College of Arts and Science, Behind Nava India, Avanashi Road, Peelamedu, Coimbatore, Tamil Nadu 641 028, India

[2]Department of BCA, Hindusthan College of Arts and Science, Behind Nava India, Avanashi Road, Peelamedu, Coimbatore, Tamil Nadu 641 028, India

Abstract: *Identification of the best treatment methods for prediction of the anti drug becomes a core objective of correctness drug. Patient-specific examination facilitate finding of personality genomic characteristics designed for each patient, and thus be able to successfully forecast personality inherited hazard of infection and carry out adapted anti-cancer rehabilitation. Though all of the existing methods for patient-specific examination have been effectively applied and experimented for detection of tumor especially for anticancer drugs and it is performed based on the non-robust manners. To manage this problem a novel schema is introduced in this work to personalized cancer therapy to forecast and tumor to anticancer drugs. In the proposed predictive modeling method of tumor sensitivity to anti-cancer drugs has primarily focused on generating functions that map gene expressions and genetic mutation profiles to drug sensitivity. In the proposed PSO-HMM method based anti cancer drug sensitive prediction is performed based on the creation of the gene expression and gene profiles. In the proposed PSO-HMM method initially the tumor sympathy, gene expression point is primarily investigate all the way through calculation of distance function. The proposed work anti cancer drug sensitivity prediction is performed based on PSO-HMM with genomic characterizations. In order to solve the problem of C and D Particle Swarm Optimization with Hidden Markov model based prediction framework is trained on genomic and practical information with the intention of be able to appreciably progress the drug understanding forecast accurateness for collected samples from Cancer Cell Line Encyclopedia (CCLE) database . From the results it shows that the accuracy and prediction results of the proposed PSO-HMM method achieves higher tumor cancer detection results than the integrated approach ,elastic net and random forest techniques.*

Keywords: Drug sensitivity prediction, personalized cancer therapy, Particle Swarm Optimization (PSO), Hidden Markov Models (HMM).

1. Introduction

In recent work several number of investigation studies is proposed based on the personalized therapy and medicine based on highly developed biomedical technologies [1-2]. A vital problem designed for personal genome study is to expose the genomic features of a person patient with the intention of is applicable designed for behavior. The elastic net-type regularized degeneration (e.g., ridge, lasso [3], elastic net [4], etc.) have been extensively second-hand to discover biomarkers, and effectively carry out designed for recognize genomic features and calculating response unpredictable based on elevated dimensional gene expression dataset. The methods, though, be able to only offer outcome based on the usual genomic features of each and every one patients. In heart, it is not until now probable to make use of these methods to recognize genomic features designed for a person patient, thus it is inflexible to successful personalized cure and medication.

Wang et al. [5] measured the patient-specific way behavior support on a mixed representation, where the unchanging possessions representation the mean way of gene expressions profiles designed for patient groups and random effects explain patient difference beginning the cluster mean. Shimamura et al. [6] introduced a new method named as NetworkProfiler, meant for recognize patient-specific gene dictatorial networks relying on changing the coefficient values and kernel-based regression model. In the kernel based regression model, Gaussian kernel is introduced to successfully carry out patient precise examination based on neighborhood example approximately a patient.

However the majority of the elastic net-type regularization regression model successfully designed for patient precise examination, their accuracy and efficient obtain an unexpected turn designed for the bad in the existence of outliers, since the methods are builded through non-robust manners. In perform, the scientific and genomic alterations datasets frequently enclose outliers beginning a variety of sources such as experiment error, coding error, etc. and recent methods mightn't work well successfully discover patient-specific biomarkers and forecast anti-cancer drug sympathy. Even though the problem is dangerously significant, comparatively small concentration has been compensated to the toughness of patient-specific examination. The envision with the intention of the overall pipeline of the propose of personalized therapy determination be comparable to the workflow is illustrated in Fig. 1. The diverse steps in Fig. 1 are summarized as follows:

a) If the specific patient is detect through cancer.
b) A most important society of the tumorous province rely on the biopsy of the patient is recognized.
c) Hereditary and practical information is created intended for the tumor society. A number of them are,
   1) Mutation position of tumor suppressors and oncogenes

2) RNA microarray information designed for measuring of gene expression and protein phosphoarrays to measure protein expression

3) DNA methylation outlines information, methylation position.

d) A representation is creating rely on practical and obtainable genomic dimensions.

e) (E1-E2) Imperfect position of time series capacity knows how to be carried out to notice the dynamics of the still representation elucidated through the method in D.

f) Grouping therapy is considered make use of the personalized static representation in D and the self-motivated representation (if estimated).

g) A patient-derived mouse Xenograft model [7] know how to be second-hand to learning growth of opposition concurrently.

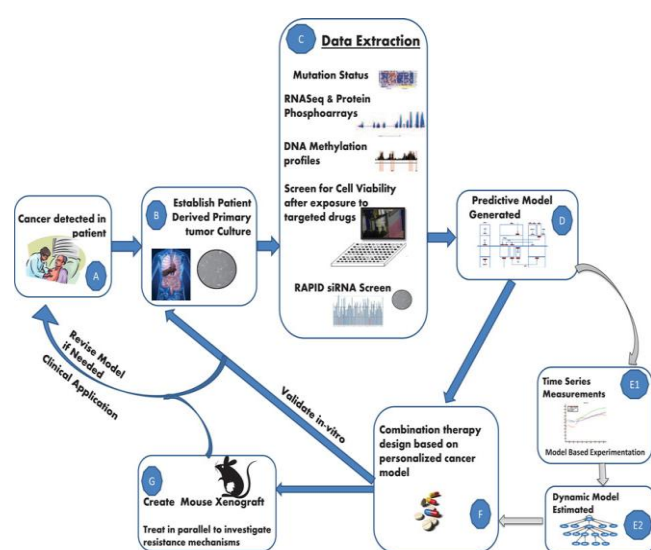(B, A revisited) In vitro-validation, representation reconsideration and patient therapy assortment

Figure 1: Combination therapy design workflow

A present method to creation of a drug sensitivity forecast representation frequently relies on genetic measurements by you. In the recent years a several number of methods have been proposed to detect the target drug sensitivity for personalized therapy [8-10]. In this work we present a novel therapy design model to solve the D problem and also perform the method based on the genomic measurements in step C. Since several number of studies in the recent work doesn't follow the above mentioned issues. In order to solve the problem of C and D Particle Swarm Optimization with Hidden Markov model based prediction framework is trained on genomic and practical information with the intention of be able to appreciably progress the drug understanding forecast accurateness for collected samples from Cancer Cell Line Encyclopedia (CCLE) database .This contains the information experiments on mouse embryonal rhabdomyosarcoma (eRMS) cell line designed for predictive models with different genomic categorization. Additional examination of the results is incorporated in the Conclusions part.

2. Background Knowledge

Several diverse methods be able to be implement toward

make drug sensitivity predictions as of pretreatment profiling data sets. Consistent through preceding DREAM challenge results [11], observed so as to no single technique category consistently outperformed the others. This suggest so as to the separation in performance is heavily based scheduled factors such as feature selection as well as method-specific implementations.

Screening for cytotoxic activity preserve be performed with natural product extracts, in annotate compound libraries or drugs isolated through normal drug design i.e. substances synthesized toward act against precise molecular targets [12]. The original stage of drug showing often involves a great amount of compound as well as is frequently approved out by means of tumour cell lines [13].

If a line of investigate is calculated toward objective a exacting cancer site otherwise genetically detailed subgroup, drug screens may possibly be approved out use a panel of tumour cell lines equivalent to so as to tumour type [14] To achieve additional specific in sequence since a screen as well as learning particular cells within detail, it is probable to execute image base program which have proven useful during the recognition of compound through proteosomal activity. Screening a large amount of compound is able to be very labor intensive.

Ongoing investigate in tumor sensitivity calculation is regularly base on genetic mutation, gene expression capacity otherwise an arrangement of genetic information [15]. The move toward of by means of inherited mutation designed for predicting the understanding is controlled by means of the presence of non-functional mutation as well as additional latent variables. Statistical tests encompass been use toward demonstrate that heritable mutations can be analytical of the drug sensitivity during non-small cell lung cancers [16] other than the organization rates used for the aberrant sample are still low.

In [17], gene expressions profiles are use toward predict the binarized efficiency of a drug in excess of a cell line through the accuracy of the planned classifiers range from 64 to 92 percent. In [18], a co-expression extrapolation (COXEN) move toward was used toward expect the drug sensitivity designed for sample outside the training set through an accurateness of around 82 as well as 75 percent in predicting the binarized sensitivity of bladder along with breast cancer cell lines correspondingly. The models were scheduled the gene expression as well as chemosensitivity reply information of the NCI-60 cell lines as well as tested taking place a panel of 40 human bladder cancer cell lines as well as 84 breast cancer patients.

Earlier studies [19] express to facilitate the prediction of drug understanding base on genomic characterization alone often suffers starting low accuracy as well as restrictions on the predictability of less general tumors. In this expose, explore the integration of the well-designed answer of the tumor culture toward a teaching set of drugs designed for increasing the calculation accurateness of the organization. The purpose is to combine the genomic characterization through the knowledge of the reply of the biological method (tumor culture) to detailed perturbations (training drugs) for

improved predictive modeling.

The planned approach is base on kernel-based elastic net-type regularization, as well as consequently be capable of execute patient-specific investigation all the way through neighborhood sample approximately a target patient. Moreover, method protect attain successfully designed for predicting anti-cancer drug sympathy and recognize drug response-specific biomarkers designed for every one patient still in the incidence of outliers, in view of the fact with the intention of the technique is support on a vigorous regularized regression through by means of a weight all the way through the Mahalanobis distance in principal component space [20].

3. Proposed Particle Swarm Optimization with Hidden Markov Model Methodology

In this work we present a novel therapy design model to solve the D problem and also perform the method based on the genomic measurements in step C. Since several number of studies in the recent work doesn't follow the above mentioned issues. In order to solve the problem of C and D Particle Swarm Optimization with Hidden Markov model based prediction framework is trained on genomic and practical information with the intention of be able to appreciably progress the drug understanding forecast accurateness for collected samples from Cancer Cell Line Encyclopedia (CCLE) database .This contains the information experiments on mouse embryonal rhabdomyosarcoma (eRMS) cell line designed for predictive models with different genomic categorization.

Let us consider an example C specifies set of all possible m primary tumor cells and S specifies the set of all n genes individuality which is determined for each and every primary tumor cells. Let G the genetic matrix of the each and every cell is denoted as m rows and N columns. Note that $N \geq n$ consists of numeric of genetic and epigenetic capability which is specified by α. Let $evalU_\lambda$ is specified as the sensitivity measure for drug sensitivity prediction for each and every one of the primary tumor cells is determined for each drug j, k. To measure the drug sensitivity of the each drug Baum-Welch algorithm is applied for HMM parameters which are automatically perform iteration .In general HMM the training samples results are not accurateness. In order to overcome these problems in this work the following constraints is described, $\lambda = (A, p, B)$ which maximizes $P(O \mid \lambda)$ subject to,

$$a_{ik} \geq 0 \quad (1)$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad (2)$$

$$P_j \geq 0 \quad (3)$$

$$U_i \geq 0 \quad (4)$$

Where $i = 1, ..K, k = 1, ..N$ and N is number of hidden states. However, the PSO was in the beginning developed designed for unrestrained optimization and must be modified to be present capable to be there built of A, p, m, U

parameters reconfigured to single hold through constraints. Each particle is considered as the gene cells and represented in the form of G the genetic matrix dimensional vector. Two sorts of arrangement of particles (encoding) be second-hand here (see the Table 1) - the *Type A* training is second-hand at what time the particle communicate to each and every one parameters λ, the type B training does not consist of preceding part of probabilistic parameters, for the reason with the purpose of it can be determined beginning stochastic constraints illustrate.

Table 1: Two types of representation of a particle

A(NXN)	p(N)	m(N)	U(N)
$a_{11,...}, a_{1N}, a_{N1,..}, a_{NN}$	$p_{1,...}p_n$	$m_{1,,,,}, m_N$	$U_{1,,,,}, U_N$
A(NXN-1)	p(N-1)	m(N)	U(N)
$a_{11,...}, a_{1N-1}, a_{N1,..}, a_{NN-1}$	$p_{1,...}p_{N-1}$	$m_{1,,,,}, m_N$	$U_{1,,,,}, U_N$

In this work the estimation results for gene matrix is determined based on following function

$$evalU_\lambda = -\ln P(O \mid \lambda) \quad (5)$$

Methods based on repair algorithm (PSO, PSO_{rf}) **.**The underestimated probability values of the cell in the gene matrix is determined based on moving the feature space of the genes in the matrix and determined based on the fitness evaluation only. These two following functions is majorly used in this work to measure the primary tumor drug results (PSO) is represented as the reference solution and (PSO_{rf}) is defined as the modified repaired version of the solution for gene matrix to drug sensitivity prediction . It is performed based on the transition matrix A which is determined from the gene matrix, which is converted into $a_{ij} / \sum_{j=1}^{N} a_{ij}$ from constraint (5). It is also similarly applied for initial probabilities p. Finally, a violation of positivity constraint (8) is measured based on the C_i (e.g. $c = 10^{-5}$) to c.

Method based on penalty function pso_p **:**The under determined feasible solution for gene matrix drug sensitivity prediction is represented as ,

$$evalU_\lambda = eval_F(\lambda) + penalty(\lambda) \quad (6)$$

The above function is replaced based on the fitness function K and it is altered as ,

$$evalU_\lambda = K + \xi\left[\left(\sum_{a_{ij}<0} a_{ij}^2\right) + \sum_{p_{ij}<0} p_{ij}^2 + \sum_{u_{ij}<0} (c - U_i)^2\right] \quad (7)$$

$\xi = 10$ for all experiments, thus, the final form of transition matrix A is represented based on the initial state probabilities is also computed using Equations 1 and 2. For instance, the p_N parameter is ,

$$P_N = 1 - \sum_{i=1}^{N-1} P_i \quad (8)$$

$evalU_\lambda$ is matched and evaluated based on the predefined threshold function ad_t, which is taken from C_{max} [22].

Some other similarity measures such correlation coefficient is also used to measure the target drug sensitivity results based on $evalU_\lambda$ At final stage of the work selected genomic characterization matrix G is analyzed to measure the drug and anti drug sensitivity results.

4. Results and Discussion

In order to measure the performance accuracy of the proposed and existing schema in this work the following publicly available datasets will be used such as CCLE, COSMIC, Cell Miner it consists of large number of dataset samples with genomic characterizations, which is implemented in Matlab environment. The experimentation results of the proposed and existing schema is measured based on the 10-fold cross validation with under different values λ_1 and λ_2 . MSE is also measured between the values for training and testing samples. The results in Table 2 illustrates that ISP, IRM and IRM-Elastic Net, PSO-HMM performs significantly than earlier methods , since the proposed PSO-HMM methods genomic characterizations based drug sensitivity prediction is done.

Table 2: Pearson's Correlation Coefficients between Observed and Predicted Sensitivities Values Using 10 Fold Cross-Validation of CCLE Data Set

	CCLE	RF	ISP	IRM	IRM-EN	PSO-HMM
Erotinib sensitivity correlation coefficients						
V=0.5	0.35	0.43	0.700	0.728	0.786	0.856
V=0.75	0.35	0.43	0.714	0.731	0.748	0.856
V=1	0.35	0.45	0.713	0.734	0.754	0.864
V=1.5	0.34	0.44	0.712	0.746	0.798	0.847
AZD0530 sensitivity correlation coefficients						
V=0.5	0.21	0.29	0.5197	0.587	0.621	0.789
V=0.75	0.21	0.29	0.5217	0.5898	0.6134	0.794
V=1	0.21	0.29	0.5234	0.5964	0.6147	0.799
V=1.5	0.21	0.29	0.4978	0.5941	0.6241	0.783

Sequential floating forward search (SFFS) is proposed in this work to perform the feature selection and predicative is done based on the selected features. The cost function is generated using PSO-HMM with number of iterations is specified to reach the target anti drug sensitivity results with 15 targets of Leave one out error analysis shown in Table 3.

Table 3: Leave one out error analysis for Mouse Embryonal Rhabdomyosarcoma

Drug name	Error	Predicated value	Actual value
Alisertib	0.5124	0.1356	1
Crizotinib	0.0123	0.8971	0.8479
GANT61	0.1568	0.1548	0.6589
BMS-754807	0.0045	0.5471	0.8471
BIX 01294	0.0056	0.4213	0.9874
Pelitinib	0	0	0
Obatoclax	0.05468	0.6154	0.894
TrichostatinA	0.0546	0.4568	0.7148
Cediranib	0.1546	0.2146	0.6589
Dasatinib	0.02356	0.7785	0.7894

Panobinostat	0.01457	0.6489	0.54689
Vorinostat	0.0145	0.5487	0.6987
SNS-032	0.0004	0.4568	0.8974
Carfilzomib	0.4568	0.2546	0.8906
AZD-8931	0	0	0
BEZ235	0.0045	0.6546	0.8478
cilengitide	0.01548	0.7154	0.8416

5. Conclusion and future work

The main aim and objective of this paper is to propose novel PSO-HMM methods for prediction of drug sensitivity especially for anti cancer prediction. The Particle Swarm Optimization (PSO) seems helpful and appropriate technique intended for Hidden Markov Models (HMM) training. The proposed PSO-HMM is applied for Anti-Cancer Drug Sensitivity based on the calculation of likelihood. The proposed PSO-HMM prediction makes use of well-designed and genomic information of training drugs. The proposed PSO-HMM is experimented to 60 drugs from embryonal rhabdomyosarcoma tumor cell and achieves higher prediction accuracy when compare to existing methods. The prediction results of the proposed PSO-HMM are evaluated using $evalU_\lambda$ among forecast and real investigational compassion. The present work will be extended to apply other dataset Cancer Genome Atlas (TCGA) which is collected from http://cancergenome.nih.gov/ and experimented using matlab environment. In the future; the other methods such as Linear PSO or pareto-ranking methods will be also used for anticancer drug sensitivity. Prospect investigate determination necessitate make use of the structure designed for creation of grouping treatment through diverse optimization goals such as circumvent struggle through targeting many self-governing pathways.

References

[1] L. Chin, J.M Andersen, P.A Futreal, "Cancer genomics: from discovery science to personalized medicine," Nature Medicine (17), pp. 297–303, 2011.

[2] S. Gua, G. Yinb, J.J Lee, "Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints," Contemp Clin Trials (36),pp.642–650,2013.

[3] R .Tibshirani , "Regression shrinkage and selection via the lasso," J R Stat Soc Series B (28), pp.267–288,1996.

[4] H. Zou, T. Hastie , "Regularization and variable selection via the elastic net," J R Stat Soc Series B (67),pp.301–320,2005.

[5] L. Wang, X. Chen, B. Zhang, "Statistical analysis of patient-specific pathway activities via mixed models, " J Biom Biostat (8), pp. 7313, 2013.

[6] T. Shimamura, S. Imoto, Y. Shimada, Y. Hosono, A. Niida, "A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition," PLoS ONE 6:e20804 , 2011.

[7] Richmond , Y. Su, "Mouse xenograft models vs gem models for human cancer therapeutics," Disease Models Mech (1), pp. 78–82, 2008.

[8] A. Hopkins, J. Mason, J. Overington, "Can we rationally design promiscuous drugs?," Current Opinion Struct. Biol (16), pp. 127–136, 2006.

[9] Z. A. Knight , K. M. Shokat, "Features of selective kinase inhibitors," Chem. Biol (12), pp. 621–637, 2005.

[10] B. J. Druker, "Translation of the Philadelphia chromosome into therapy for CML," Blood, 112, pp. 4808–4817, 2008.

[11] D. Marbach et al, "Wisdom of crowds for robust gene network inference," Nat. Methods 9, pp.796–804 , 2012.

[12] G. Damia, M. D'Incalci, "Contemporary pre-clinical development of anticancer agents--what are the optimal preclinical models?", Eur J Cancer 45, pp.2768-81,2009.

[13] R. H. Shoemaker , "The NCI60 human tumour cell line anticancer drug screen," Nature reviews (6), pp.813-23,2006.

[14] S. V. Sharma, D. A. Haber, J. Settleman, "Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents," Nature reviews (10), pp.241-53, 2010.

[15] J. Barretina, G. Caponigro, et.al , "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," Nature, (483), pp.603–607,2010.

[16] M. L. Sos, K. Michel et.al , "Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions," J. Clin. Investigation (119), pp. 1727–1740, 2009.

[17] J. E. Staunton, D. K. Slonim, H. A. Coller, P. Tamayo, M. J. Angelo, J. Park, U. Scherf, J. K. Lee, W. O. Reinhold, J. N. Weinstein, J. P. Mesirov, E. S. Lander, and T. R. Golub, "Chemosensitivity prediction by transcriptional profiling," in Proc. Nat. Acad. Sci. USA, pp. 10787–10792, 2001.

[18] J. K. Lee , D. M. Havaleshko, H. Cho, John N. Weinstein , E. P. Kaldjian, J. Karpovich, A. Grimshaw, and D. Theodorescu, "A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery," Proc. Nat. Acad. Sci. USA, pp. 13086–13091, 2007.

[19] M. Garnett, M. J. Garnett, E. J. Edelman et. Al, "Systematic identification of genomic markers of drug sensitivity in cancer cells," Nature (483), pp. 570–575, 2012.

[20] Park H, Konishi S, Robust solution path for high dimensional sparse regression modeling. Commun Stat Simul Comput, In press, 2014.

[21] X.Hu, R.C.Eberhart , " Solving constrained nonlinear optimization problems with particle swarm optimization," In: Proceedings of the Sixth World Multi conference on Systems, Cybernetics and Informatics, 2002.

[22] L. Tothfalusi ,L. Endrenyi, "Estimation of CMAX and TMAX in populations after single and multiple drug administrations," J Pharmacokinet Pharmacodyn (30), pp. 363–385, 2003.

1338